

ZASTOSOWANIE METOD ILOŚCIOWYCH W EKONOMII I FINANSACH

D E B I U T Y S T U D E N C K I E

2024

ZASTOSOWANIE METOD ILOŚCIOWYCH W EKONOMII I FINANSACH

pod redakcją
Alicji Grześkowiak
i Piotra Peterneka



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2024

Recenzja

dr hab. Joanna Krupowicz, prof. UEW

Redakcja wydawnicza

Aleksandra Śliwka

Korekta

Katarzyna Gwizda

Skład i łamanie

Adam Dębski

Projekt okładki

Beata Dębska

Na okładce wykorzystano zdjęcia z zasobów Adobe Stock

Praca opublikowana na licencji Creative Commons Uznanie autorstwa

Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0).

Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>



ISBN 978-83-67899-52-9 (wersja papierowa)

ISBN 978-83-67899-53-6 (wersja elektroniczna)

DOI: 10.15611/2024.53.6

Druk i oprawa: TOTEM

Spis treści

| | |
|--|-----|
| Wstęp | 7 |
| Patryk Augustyniok, Oktawia Golczyk: Aktywność turystyczna w wybranych krajach europejskich – analiza porównawcza | 11 |
| Aleksandra Ceglarska: Weryfikacja występowania anomalii kalendarzowych na rynku kryptowalut na przykładzie Bitcoina | 25 |
| Maja Ciechanowska: Analiza poziomu zaufania społecznego i jego uwarunkowań..... | 41 |
| Julia Dymek, Weronika Kolasa: Analiza stylu życia obywateli Unii Europejskiej w kontekście lat przeżytych w zdrowiu..... | 53 |
| Wiktoria Galarowicz: Predykcja cen mieszkań w dużych miastach w Polsce z wykorzystaniem uczenia maszynowego | 68 |
| Dominik Głowacki: Analiza wyceny piłkarzy za pomocą modelowania ekonometrycznego | 82 |
| Julia Matysiak: Sytuacja na rynku nieruchomości mieszkaniowych w Polsce na tle wybranych wskaźników społeczno-ekonomicznych | 99 |
| Łukasz Meller: Analiza atrakcyjności rynku lokali mieszkalnych w województwie mazowieckim..... | 115 |
| Julia Mikołajczak: Analiza jakości zarządzania budżetami i funduszami unijnymi w województwach w Polsce w 2020 roku | 129 |
| Karol Noganowicz: Związek infrastruktury transportowej z rozwojem regionalnym | 146 |
| Marta Sośnicka: Wpływ wybranych czynników społeczno-ekonomicznych na rynek książki w Polsce | 163 |
| Julia Wobalis: Analiza jakości systemów ochrony zdrowia w państwach Unii Europejskiej..... | 179 |

Wstęp

Podjęcie decyzji w warunkach gospodarki rynkowej wymaga dogłębnej znajomości otoczenia oraz uwarunkowań wewnętrznych organizacji. Kluczowym komponentem takich działań jest wykorzystanie metod ilościowych umożliwiających uzyskanie rzetelnych i precyzyjnych informacji oraz pozwalających na podjęcie optymalnych decyzji. Metody statystyczno-ekonometryczne aplikowane są w bardzo szerokiej klasie problemów o charakterze ekonomicznym i finansowym, a dodatkowo są w ciągłym rozwoju, dając nie tylko szerokie spektrum metodyk, ale i obejmując coraz większy zakres zagadnień.

Niniejsza monografia z serii „Debiuty Studenckie” pt. *Zastosowanie metod ilościowych w ekonomii i finansach* utwierdza w przekonaniu o niebagatelnych możliwościach metod statystyczno-ekonometrycznych, zawiera bowiem artykuły dotyczące różnych obszarów działalności gospodarczej człowieka oraz problemów społeczno-ekonomicznych, wykorzystując znaczący zestaw metodyk.

Autorami artykułów są studenci Uniwersytetu Ekonomicznego we Wrocławiu z kierunku Analityka gospodarcza. Publikacja składa się z 12 artykułów przedstawiających wyniki przeprowadzonych przez nich bardzo różnorodnych obszarowo badań, lecz mających wspólny element, którym jest wykorzystanie metod ilościowych. Użyte metodyki sięgają od klasycznych analiz obejmujących statystyki opisowe poprzez podejścia wielowymiarowe, takie jak analiza skupień czy modele ekonometryczne, aż po zastosowanie metod uczenia maszynowego. Autorzy zastosowali te metody w analizie różnych aspektów życia społecznego i gospodarczego – rozpatrywano aktywność turystyczną, zaufanie społeczne, styl życia, jakość służby zdrowia czy dostępność transportu. Trzy prace poświęcone zostały rynkowi nieruchomości, a dziewięć opracowań bardziej szczegółowym zagadnieniom, takim jak notowania kryptowalut, jakość zarządzania funduszami unijnymi, rynek książki w Polsce czy wycena piłkarzy.

Monografię rozpoczyna artykuł Patryka Augustynioka i Oktawii Golczyk, w którym to dokonano analizy porównawczej aktywności turystycznej w wybranych krajach europejskich. Autorzy zaproponowali odpowiednie wskaźniki, a następnie z wykorzystaniem hierarchicznych i niehierarchicznych metod grupowania zbadali podobieństwa między państwami Europy.

Drugie prezentowane w pracy opracowanie, autorstwa Aleksandry Ceglarskiej, dotyczy interesującego tematu kryptowalut. Autorka wykorzystując metody bootstrapowe, analizuje ryzyko inwestycji w Bitcoiny z odniesieniem do wybranych przedziałów czasowych. Zidentyfikowała tym sposobem występowanie zależności między rytmem dnia typowego inwestora a anomaliami godzinowymi notowań Bitcoina.

Kolejny artykuł poświęcony jest szeroko rozumianemu zaufaniu społecznemu. Maja Ciechanowska w swej pracy wyodrębnia czynniki mające największy wpływ na

poziom zaufania i wykazuje, że Europa nie jest jednorodna pod względem poziomu zaufania społecznego. Stwierdza także, że na poziom zaufania wpływ mogą mieć uwarunkowania geograficzne i kulturowe.

Następne opracowanie autorstwa Julii Dymek i Weroniki Kolasy poświęcone jest analizie stylu życia mieszkańców Unii Europejskiej, ale w odniesieniu do lat przeżytych w zdrowiu. Autorki wykorzystując narzędzia ekonometrii, identyfikują czynniki mające wpływ na liczbę lat przeżytych w zdrowiu, a następnie wykorzystują je do wykonania analizy porównawczej krajów Unii Europejskiej.

Artykuł Wiktorii Galarowicz poświęcony jest zastosowaniu metod uczenia maszynowego w przewidywaniu cen mieszkań w dużych miastach. Autorka bazując na danych historycznych, wyznacza zmienne mające wpływ na ceny, dokonuje ich gruntownego wstępnego przygotowania, a następnie stosuje wybrane algorytmy *machine learningu* do prognozowania cen nieruchomości.

Szósty artykuł, autorstwa Dominika Głowackiego, stosuje model ekonometryczny do wyceny piłkarzy. Autor poszukuje kluczowych czynników mających wpływ na wartość piłkarzy, a następnie próbuje wykorzystać procedurę modelowania ekonometrycznego do predykcji wysokości kwoty transferowej.

Następne opracowanie, którego autorką jest Julia Matysiak, dotyczy rynku nieruchomości. Autorka analizuje czynniki społeczno-ekonomiczne w powiatach w Polsce, odnosząc te warunki do sytuacji na rynku mieszkaniowym. Wyniki tego badania wskazują, że najkorzystniejsza sytuacja mieszkaniowa jest w dużych powiatach miejskich.

Podobną tematyką zajął się Łukasz Meller, który w kolejnym prezentowanym w monografii artykule zajął się badaniem atrakcyjności lokali mieszkalnych w województwie mazowieckim. Autor wskazuje czynniki mające wpływ na atrakcyjność takich nieruchomości, następnie dokonuje wyboru powiatów charakteryzujących się lepszą sytuacją na rynku lokali mieszkalnych.

Dziewiąty prezentowany w pracy artykuł, Julii Mikołajczak, dotyczy absorpcji funduszy unijnych przez województwa w Polsce. Autorka z wykorzystaniem metod analiz wielowymiarowych identyfikuje województwa, które najlepiej zarządzają swoim budżetem oraz te, które radzą sobie pod tym względem najgorzej.

Kolejne prezentowane opracowanie, autorstwa Karola Noganowicza, rozważa wpływ stanu infrastruktury transportowej na rozwój regionalny. Autor wykorzystując metody wielowymiarowej analizy porównawczej, stwierdza brak wyraźniej zależności pomiędzy jakością infrastruktury transportowej, a poziomem rozwoju regionalnego w roku 2021.

Jedenasty artykuł, autorstwa Marty Sośnickiej, rozpatruje zmiany na rynku książki. Autorka przeprowadziła ankietę wśród czytelników i zastosowała szereg metod formalnych do jej analizy. Badanie wykazało, że przyczyną regresu na rynku książki jest przede wszystkim zmiana trybu życia ludzi.

Monografię kończy artykuł Julii Wobalis, w którym porównane zostały systemy opieki zdrowotnej w krajach Unii Europejskiej. Autorka zidentyfikowała odpo-

wiednie wskaźniki, a następnie z wykorzystaniem metod wielowymiarowej analizy porównawczej pogrupowała i uszeregowała kraje pod względem jakości systemu ochrony zdrowia.

Reasumując, prezentowane artykuły poruszają wielorakie zagadnienia z obszaru ekonomii i finansów, a jednocześnie wykorzystują znaczną liczbę zróżnicowanych metod analitycznych. Świadczy to o wysokiej jakości prac Autorów, którzy debiutują w sferze badań naukowych.

*Alicja Grześkowiak
Piotr Peternek*

Patryk Augustyniok

e-mail: 181235@student.ue.wroc.pl

Oktawia Golczyk

e-mail: 181511@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Aktywność turystyczna w wybranych krajach europejskich – analiza porównawcza

DOI: 10.15611/2024.53.6.01

JEL Classification: C38, Z32

© 2024 Patryk Augustyniok, Oktawia Golczyk

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Augustyniok, P. i Golczyk, O. (2024). Aktywność turystyczna w wybranych krajach europejskich – analiza porównawcza. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 11-24). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Artykuł przedstawia analizę aktywności turystycznej w 24 krajach europejskich w 2022 roku. W pracy zdefiniowano pojęcie aktywności turystycznej oraz opisano czynniki wpływające na nią. Do analizy użyto dwóch metod grupowania: klasyfikacji hierarchicznej i k-średnich. Problemem badawczym zasadał się na znalezieniu podobieństw między państwami w Europie ze względu na zmienne opisujące partycypację w turystykę. Do wybrania odpowiedniej liczby skupień użyto wykresu łokcia bazującego na WSS (*within sum of square*). Analiza dowiodła, że niektóre kraje rozwinięte są podobne pod względem aktywności turystycznej, jednakże część z tych państw pogrupowana została z tymi o niższej jakości życia, np. Słowenia czy Węgry. Badanie nie wykazało wyraźnych różnic pomiędzy Europą Zachodnią a Wschodnią. Obie metody klasyfikacji utworzyły te same skupienia.

Słowa kluczowe: analiza skupień, klasyfikacja hierarchiczna, metoda k-średnich, turystyka, aktywność turystyczna

1. Wstęp

W artykule podjęto temat aktywności turystycznej w wybranych krajach europejskich. Problem badawczy dotyczył znalezienia podobieństw między państwami w Europie pod względem analizowanych zmiennych. W pracy wykorzystano dwie

metody grupowania: klasyfikację hierarchiczną i k-średnich, w celu porównania otrzymanych wyników. Postawiono następujące pytania badawcze:

- Czy kraje uznawane za wysoko rozwinięte wykazują podobieństwo w aspekcie aktywności turystycznej?
- Czy istnieją wyraźne różnice pomiędzy Europą Zachodnią a Wschodnią?
- Jakie różnice występują pomiędzy skupieniami w obu metodach badawczych?

Wybrany temat jest istotny ze względu na wiele czynników. Turystyka może stanowić ważny sektor gospodarki generujący znaczne przychody i tworzący wiele miejsc pracy (Kiryluk, 2014). Ma to bezpośredni wpływ na PKB kraju, przyczyniając się tym samym do jego rozwoju gospodarczego. W Europie znajdują się państwa, które przyciągają podróżnych z całego świata i są powszechnie uznawane za liderów w turystyce, np. Francja, Włochy czy Hiszpania. Niemniej jednak istnieją istotne bariery utrudniające możliwość podróżowania. Wielu ludzi nie może pozwolić sobie na aktywność turystyczną ze względu na ograniczenia finansowe, czasowe, zdrowotne czy też prawne. Wysokie koszty życia i skromny budżet domowy sprawiają, iż ludzie preferują oszczędzanie na inne cele, takie jak zdrowie, edukacja czy mieszkanie, jako że turystyka jest potrzebą wyższego rzędu (Kaczmarska, 2014). Dodatkowo, zmiany w kursach walut mogą wpłynąć na liczbę turystów (Ruane, 2014). Wzrost kosztów podróży wynikający z osłabienia waluty krajowej dotyczy zwłaszcza obywateli państw spoza strefy euro. Również ludność w krajach, które niedawno dołączyły do tej strefy, może borykać się z problemami finansowymi. Obowiązki rodzinne i zawodowe także są istotnym czynnikiem hamującym podróżowanie. Długie godziny pracy, brak elastyczności w harmonogramie lub ograniczony urlop mogą być przyczyną rezygnacji z aktywności turystycznej dla wielu osób.

2. Aktywność turystyczna

Turystyka to zjawisko społeczne, kulturowe i gospodarcze, które oznacza przemieszczanie się ludzi do krajów lub miejsc poza ich zwykłym otoczeniem w celach osobistych lub biznesowych (UN Tourism, b.d.a). Interdyscyplinarność tego zjawiska sprawia, że może być ono analizowane oraz interpretowane pod kątem wielu czynników (Więckowski, 2010). W niniejszym artykule postanowiono się skupić na aktywności turystycznej.

Aktywność turystyczna jest pojęciem często stosowanym przez naukowców, jednakże równie często jest ono w inny sposób interpretowane, m.in. ze względu na różne perspektywy badania tego zjawiska (np. społeczne, gospodarcze czy psychologiczne). Taki stan rzeczy utrudnia podejmowanie naukowych analiz i porównań. W. Alejsiak (2011) przytacza dwie definicje aktywności turystycznej. W szerokim znaczeniu jest to pojęcie oznaczające ogół zachowań ludzi związanych z ich uczestnictwem w różnych formach turystyki. Wszechstronne pojmowane aktywności turystycznej wykorzystywane jest głównie w rozważaniach teoretycznych, jednakże nieprecyzyjność tej definicji generuje trudności w badaniach empirycznych. Naj-

bardziej problematycznym elementem takiego rozumienia aktywności turystycznej jest określenie „ogół zachowań”, co w przypadku próby określenia stopnia zaawansowania tego zjawiska nie pozwala na ich wartościowanie. W węższym znaczeniu pojęcie aktywności turystycznej zinterpretowano jako uczestnictwo osób w danym okresie w różnych formach turystyki. Światowa Organizacja Turystyki definiuje ją jako ogół procesów i zachowań ludzi, traktowanych jako konsumentów, w trakcie przygotowania oraz w trakcie wyjazdu (UN Tourism, b.d.b).

Równie niejednoznacznie, jak w przypadku definicji aktywności turystycznej, opisywana jest jej klasyfikacja. Jednak wnioski omawiane w większości opracowań wynikają z podobnych założeń. Najczęściej spotykane jest ujęcie dwuskładnikowe (Kwiatkowski i in., 2018). Przykładem takiej typologii jest propozycja M. Jedlińskiej (2006, za: Kwiatkowski i in., 2018), która podzieliła czynniki wpływające na aktywność turystyczną na:

- czynniki wewnętrzne:
 - potrzeby,
 - postrzeganie,
 - postawy,
 - osobowość,
 - uczenie się,
 - ryzyko;
- czynniki zewnętrzne:
 - ekonomiczne, np. sytuacja gospodarcza kraju, dochody gospodarstw domowych, ceny;
 - społeczno-kulturalne – związane z wpływem społeczeństwa na otoczenie, np. sposób spędzania czasu wolnego czy otwartość na nowe kultury;
 - czynniki regionalne, np. odległość od miast, kurortów nadmorskich czy górskich, charakter regionu, infrastruktura.

W niniejszym opracowaniu skupiono się na czynnikach zewnętrznych, jako że to one mogą różnicować kraje czy regiony, zaś czynniki wewnętrzne rozpatrywać należy na poziomie jednostki.

3. Metodyka badań

3.1. Metody grupowania

W celu znalezienia odpowiedzi na pytania badawcze posłużono się metodami grupowania k-średnich i klasyfikacją hierarchiczną. Ideą analizy skupień jest znajdowanie homogenicznych podzbiorów w niejednorodnym zbiorze obiektów (Balicki, 2013). Opiera się ona na pomiarze stopnia zróżnicowania (lub podobieństwa) jednostek w wielowymiarowej przestrzeni. Do oceny stopnia zróżnicowania używane są różne miary. W poniższym artykule posłużono się odległością euklidesową. W efekcie zastosowania metod klasyfikacji otrzymuje się K rozłącznych i niepustych

klas takich, że obiekty należące do tych samych klas są jak najbardziej podobne, zaś obiekty należące do różnych skupień są jak najbardziej zróżnicowane.

Analizę rozpoczęto od utworzenia macierzy korelacji, jako że zależność pomiędzy zmiennymi jest jednym z zasadniczych elementów metod analizy wielowymiarowej (Balicki, 2013). W przypadku analiz skupień jak najmniej zmiennych powinno być ze sobą silnie skorelowanych, ponieważ może prowadzić to do powielenia informacji niesionej przez nie. Przed rozpoczęciem grupowania obiektów zmienne należy poddać standaryzacji, dzięki czemu ich wartości są niemianowane i o zbliżonych rzędach wielkości (Steinley i Brusco, 2008).

W klasyfikacji hierarchicznej każdy obiekt początkowo traktowany jest jako odrębna grupa. Następnie przeprowadzane jest sekwencyjne grupowanie na podstawie miar odległości, tak że w każdym kolejnym etapie najbliższe znajdujące się skupienia łączone są ze sobą. Przed każdym kolejnym etapem sekwencji należy określić odległość pomiędzy klastrami. Problem ten jest niejednoznaczny ze względu na wielorakość metod. W artykule posłużono się metodą Warda opierającą się na kryterium minimalnego wzrostu łącznej wewnątrzgrupowej wariancji, które rozumiane może być jako kryterium najmniejszej dodatkowej straty informacji (Balicki, 2013). Metoda k -średnich, inaczej nazywana metodą środków ciężkości, opiera się na minimalizacji sumy kwadratów odległości pomiędzy i -tym obiektem a wektorem miar położenia obliczonym na podstawie obiektów należących do k -tej klasy (Stanimir, 2006). W praktyce w metodzie środków ciężkości (k -średnich) stosuje się algorytm, w którym ustala się docelową liczbę klas (K) oraz początkową klasyfikację zbioru (dobieraną, np. losowo). Następnie iteracyjnie oblicza się odległości każdego obiektu od wektora średnich każdej klasy i wyznacza się nową klasyfikację poprzez przydzielenie obiektu do klasy, która jest dla niego najbliższa.

Niewątpliwie w obu metodach problemem jest dobór odpowiedniej liczby skupień. W artykule posłużono się metodą WSS (*within sum of squares*). Polega ona na wykreśleniu na osi Y sumy kwadratów odległości pomiędzy każdym punktem w klasie a środkiem klasy, zaś na osi X liczby skupień. Na wykresie poszukuje się punktu przegięcia, w którym spadek WSS staje się coraz mniej znaczący.

3.2. Charakterystyka zmiennych

Niniejszą analizę wykonano dla 24 krajów europejskich. Dane dotyczą 2022 roku i pobrane zostały z Eurostatu. Wybór państw podyktowany został ich dostępnością tych danych. W badaniu posłużono się następującymi zmiennymi:

- procent_zagranica – udział osób uczestniczących w zagranicznych wyjazdach turystycznych w celach prywatnych powyżej 1 nocy [% populacji];
- procent_ojczyzna – udział osób uczestniczących w krajowych wyjazdach turystycznych w celach prywatnych powyżej 1 nocy [% populacji];
- noce_zagranica – średnia liczba nocy spędzonych w trakcie zagranicznych wyjazdów [noce/osoba];

- `noce_ojczyzna` – średnia liczba nocy spędzonych w trakcie krajowych wyjazdów [`noce/osoba`];
- `powody_finansowe` – udział osób nieuczestniczących w turystyce z powodów finansowych [% osób nieuczestniczących w turystyce];
- `powody_zawodowe` – udział osób nieuczestniczących w turystyce z powodów zawodowych [% osób nieuczestniczących w turystyce].

Do badania wybrano dane przekrojowe, które są kompletne i wiarygodne. Analizy wykonano za pomocą języka R z wykorzystaniem bibliotek: `factoextra` oraz `phatmap`.

4. Wyniki analiz

4.1. Statystyki opisowe

W tej części pracy zostaną przedstawione statystyki opisowe dla analizowanych zmiennych (tab. 1).

Tabela 1. Statystyki opisowe

| Statystyki | <code>procent_zagranica</code> | <code>procent_ojczyzna</code> | <code>noce_zagranica</code> | <code>noce_ojczyzna</code> | <code>powody_finansowe</code> | <code>powody_zawodowe</code> |
|------------------------|--------------------------------|-------------------------------|-----------------------------|----------------------------|-------------------------------|------------------------------|
| Średnia | 14,71 | 26,46 | 6,83 | 5,84 | 36,49 | 16,32 |
| Mediana | 9,70 | 27,47 | 4,33 | 4,89 | 39,51 | 16,89 |
| Odchylenie standardowe | 15,05 | 12,73 | 6,74 | 4,51 | 14,04 | 7,99 |
| Wariancja próbki | 226,48 | 161,96 | 45,49 | 20,31 | 197,25 | 63,82 |
| Kurtoza | 7,26 | 0,19 | 7,86 | 2,95 | 0,42 | 0,04 |
| Skośność | 2,34 | 0,23 | 2,33 | 1,83 | -0,74 | 0,24 |
| Minimum | 2,16 | 1,05 | 0,52 | 1,23 | 0,87 | 2,09 |
| Maksimum | 70,10 | 55,70 | 32,04 | 18,95 | 56,09 | 33,50 |

Źródło: opracowanie własne na podstawie (Ec.europa.eu, 2022).

Średni odsetek populacji, który uczestniczy w wyjazdach zagranicznych, wynosi 14,71%. W połowie badanych krajów udział osób uczestniczących w wyjazdach zagranicznych nie przekracza 9,7%. Minimalna wartość dla tej zmiennej wynosi 2,16% (Rumunia), a maksimum 70,1% (Luksemburg), co świadczy o znacznym zróżnicowaniu między krajami europejskimi.

Udział osób korzystających z wyjazdów krajowych wynosi 26,46%, co oznacza, że większy odsetek badanej populacji wybiera podróże w obrębie własnego kraju. Najmniejszy udział osób podróżujących po kraju odnotowano dla Luksemburgu (1,05%), co prawdopodobnie jest wynikiem niewielkiej powierzchni tego państwa. Z kolei największy odsetek ludności korzystającej z wyjazdów krajowych odnosi się do Francji (55,7%).

Średnia liczba nocy spędzanych podczas wyjazdów zagranicznych wynosi 6,83, co jest prawie o jedną noc więcej niż w przypadku wyjazdów w obrębie kraju. Mediana wynosi 4,33, co oznacza, że połowa populacji spędza co najmniej 4,33 nocy za

granicą. Zmienna przyjmuje rozkład prawostronny, co świadczy na niekorzyść sektora turystycznego, ponieważ większość państw europejskich ma krótsze pobyty zagraniczne niż średnia.

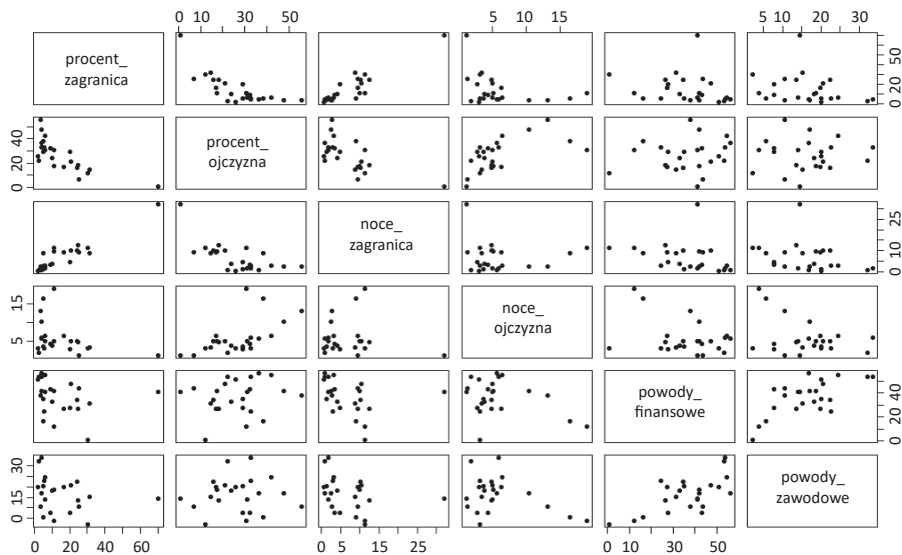
Najwięcej nocy spędzonych w kraju podczas wyjazdów turystycznych spędzają Szwedzi (18,95). Odchylenie standardowe jest równe 4,51, co oznacza, że przeciętna liczba nocy spędzonych w trakcie wyjazdów odchyła się od średniej o 4,51 nocy.

Średnio więcej osób nie uczestniczy w turystyce z powodów finansowych (36,49%) niż z powodów zawodowych (16,32%). Spośród badanych krajów największy procent osób rezygnujących z podróży z powodów finansowych odnotowano w Grecji (56,09%), natomiast najmniejszy w Szwajcarii (0,87%). Skośność wskazuje na rozkład lewostronny, co oznacza, że w badanej próbie przeważają kraje, które przyjmują większą wartość zmiennej niż średnia. Może to świadczyć o niekorzystnej sytuacji finansowej mieszkańców Europy.

Największy odsetek osób nieuczestniczących w wyjazdach turystycznych z powodów zawodowych spośród analizowanych krajów odnotowano w Portugalii, najniższy zaś w Szwajcarii.

4.2. Korelacje

W następnym kroku analizy zbadano zależności pomiędzy badanymi zmiennymi za pomocą korelacji liniowej Pearsona. Rysunek 1 przedstawia rozrzut zmiennych, a rysunek 2 macierz korelacji.



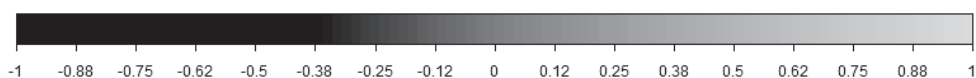
Rys. 1. Wykres rozrzutu

Źródło: opracowanie własne.

Na podstawie wykresu rozrzutu nie można znaleźć zależności liniowej pomiędzy parami zmiennych. Żadne obserwacje nie układają się liniowo.

Korelacja

| | | | | | |
|-------------------|----------------|------------------|-----------------|------------------|---------------|
| procent_zagranica | 0.92 | -0.23 | -0.26 | -0.75 | -0.35 |
| 0.92 | noce_zagranica | -0.31 | -0.31 | -0.66 | -0.06 |
| -0.23 | -0.31 | powody_finansowe | 0.66 | 0.17 | -0.34 |
| -0.26 | -0.31 | 0.66 | powody_zawodowe | 0.02 | -0.36 |
| -0.75 | -0.66 | 0.17 | 0.02 | procent_ojczyzna | 0.59 |
| -0.35 | -0.06 | -0.34 | -0.36 | 0.59 | noce_ojczyzna |



Rys. 2. Korelacje między zmiennymi

Źródło: opracowanie własne na podstawie (Ec.europa.eu, 2022).

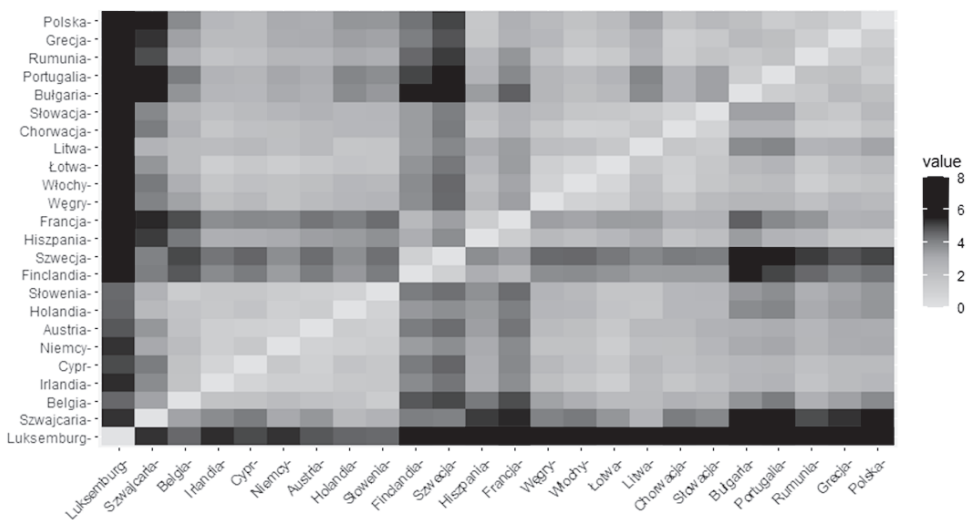
Z rysunku 2 wynika, że tylko dwie zmienne są ze sobą istotnie skorelowane – noce_zagranica i procent_zagranica. Ich korelacja wynosi 0,92, co oznacza, że wraz ze wzrostem wartości jednej zmiennej wzrasta również wartość drugiej cechy. Nale-

ży jednak zauważyć, że ten rysunek nie wskazuje na tak silną zależność. Rozbieżność ta może być spowodowana występowaniem obserwacji odstającej. W znacznym stopniu skorelowany ze sobą jest procent_zagranica i procent_ojczyzna ($-0,75$). Wskazuje to na prawidłowość, że jeśli udział osób uczestniczących w wyjazdach zagranicznych wzrośnie, to udział osób podróżujących w obrębie kraju zmaleje. Najmniejsza korelacja między zmiennymi występuje dla powody_zawodowe i procent_ojczyzna ($0,02$) oraz noce_zagranica i noce_ojczyzna ($-0,06$). Test na istotność współczynnika korelacji wykazał, że tylko 5 par zmiennych charakteryzuje się istotną statystycznie zależnością: procent_zagranica – procent_ojczyzna, procent_zagranica – noce_zagranica, procent_ojczyzna – noce_zagranica, procent_ojczyzna – noce_ojczyzna oraz powody_finance – powody_zawodowe.

4.3. Analiza skupień

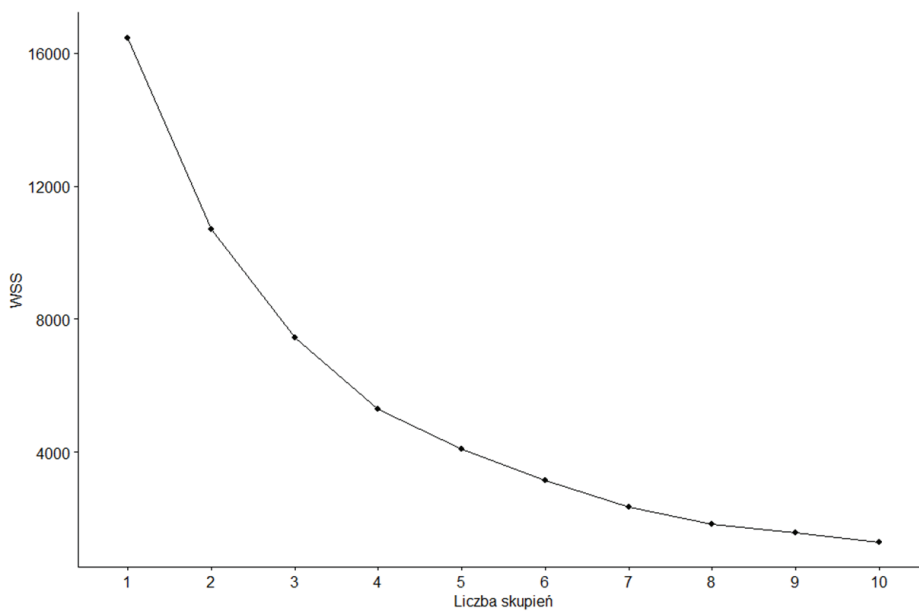
Do odpowiedniego przeprowadzenia analizy skupień niezbędne było zestandaryzowanie zmiennych, aby były one porównywalne. W tym podrozdziale porównane zostaną ze sobą wyniki klasyfikacji hierarchicznej oraz metody k-średnich. Na rysunku 3 przedstawiono macierz odległości pomiędzy zmiennymi.

Na podstawie macierzy odległości można zauważyć, że Luksemburg odbiega znacząco od większości państw. Mimo to zdecydowano się go pozostawić w analizie, jako że niesie to informację o znaczącym zróżnicowaniu aktywności turystycznej w Europie. Widać również skupienie państw: Francji, Hiszpanii, Szwecji i Finlandii, które oddalone są znacząco od reszty krajów, zaś występują w swoim pobliżu.



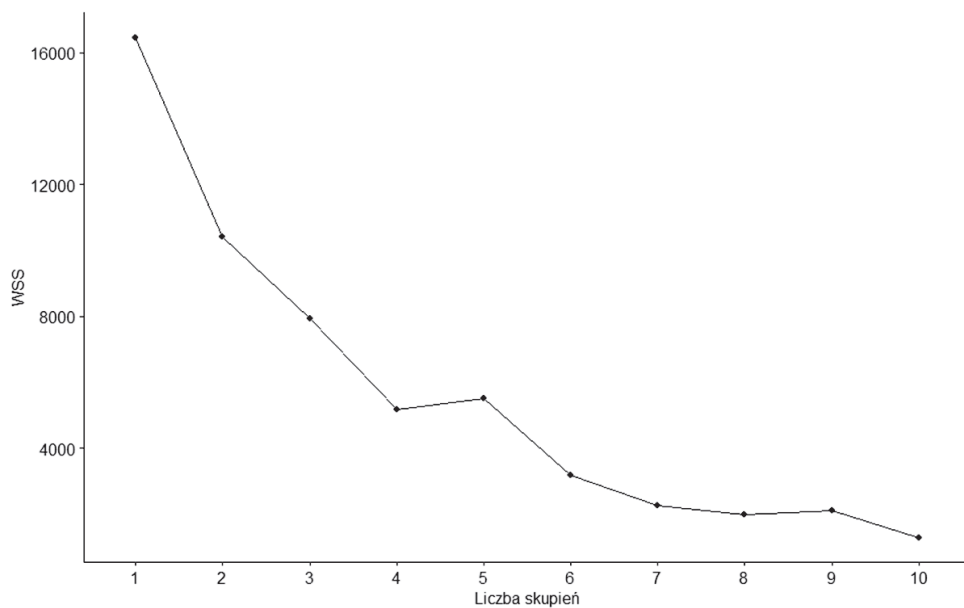
Rys. 3. Macierz odległości

Źródło: opracowanie własne na podstawie (Ec.europa.eu, 2022).



Rys. 4. Wykres łokcia – klasyfikacja hierarchiczna

Źródło: opracowanie własne.



Rys. 5. Wykres łokcia – k-średnich

Źródło: opracowanie własne.

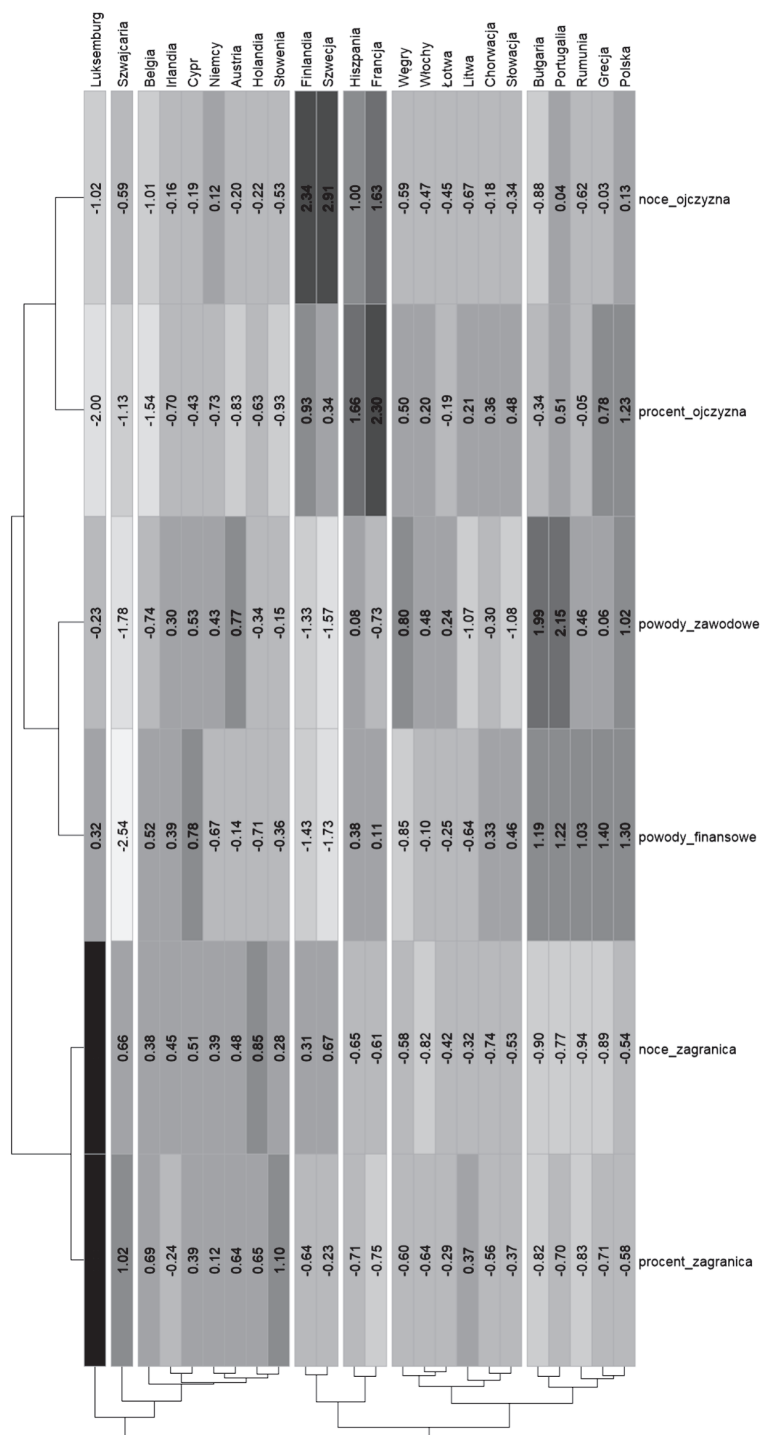
W kolejnym kroku wyznaczono optymalną liczbę skupień dla obu metod, używając do tego metody WSS. Wykresy zostały przedstawione odpowiednio dla klasyfikacji hierarchicznej na rysunku 4, a dla metody k -średnich na rysunku 5.

Z powyższych rysunków wywnioskować można, że dla klasyfikacji hierarchicznej nie występuje wyraźny punkt przegięcia sugerujący liczbę skupień. W przypadku k -średnich zdecydowano się wybrać 7 skupień, jako że suma kwadratów odległości podczas zwiększania liczby skupień powyżej 7 nie maleje znacząco. W celu porównania wyników dla klasyfikacji hierarchicznej zdecydowano się przyjąć taką samą liczbę klastrów. Wybór ten wspiera również coraz mniej znaczący spadek odległości sumy kwadratów odległości.

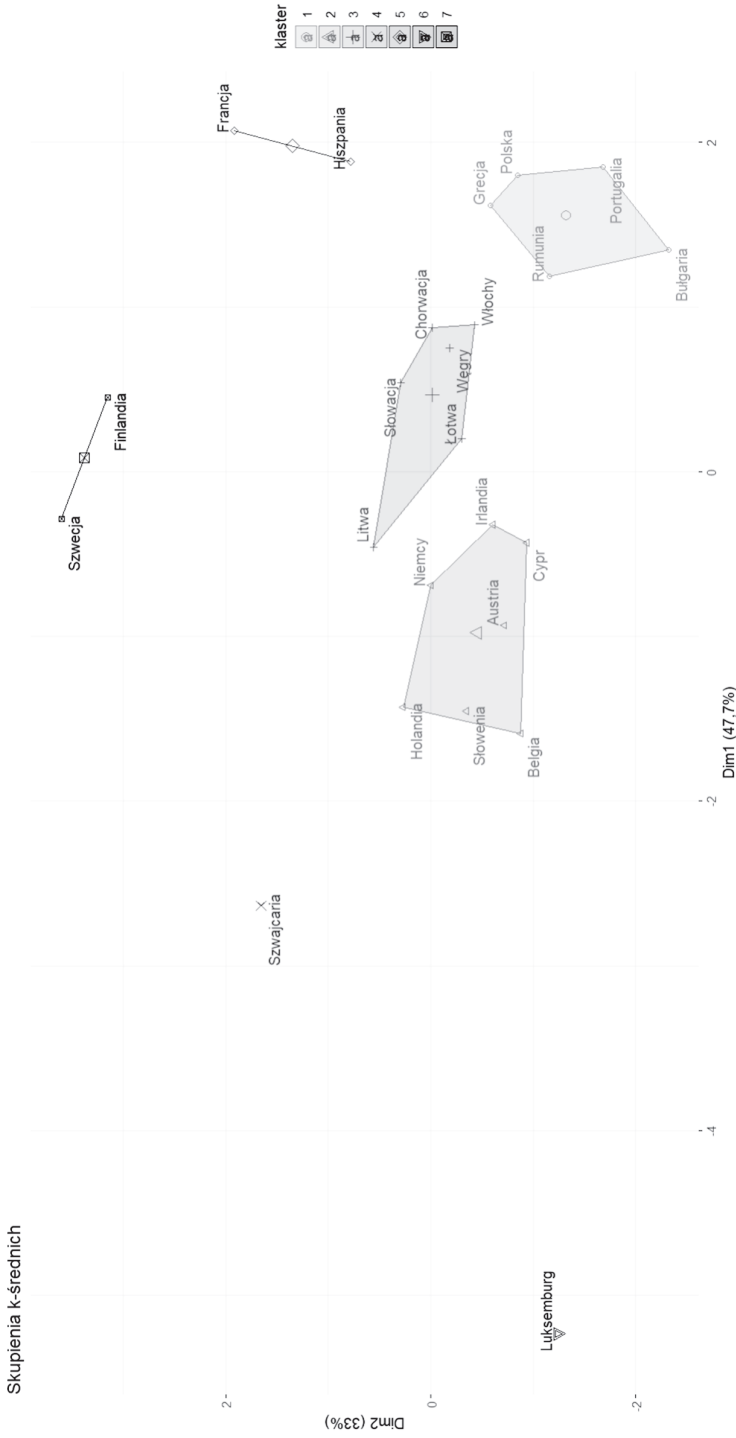
4.3.1. Klasyfikacja hierarchiczna i metoda k -średnich

W kolejnej części analizy wyznaczono skupienia dla obu metod. Na rysunku 6 przedstawiono podział na skupienia dla klasyfikacji hierarchicznej, a na rysunku 7 dla k -średnich. Porównując oba rysunki, można wywnioskować, iż obie metody utworzyły identyczne grupy.

Pierwsze dwa skupienia są jednoelementowe. Tworzone są przez Luksemburg i Szwajcarię, a więc wysoko rozwinięte państwa europejskie. Oba skupienia odznaczają się wysokimi wartościami zmiennych $\text{procent_zagranica} - \text{noce_zagranica}$ oraz niskimi wartościami zmiennych $\text{procent_ojczyzna} - \text{noce_ojczyzna}$. Świadczy to o wysokiej partycypacji obywateli tych państw w zagranicznej turystyce, a także o ich niskim udziale w turystyce krajowej. Kolejne skupienie składa się z Belgii, Irlandii, Cypru, Niemiec, Austrii, Holandii i Słowenii. Skupienie to odznacza się przeciętnymi wartościami dla każdej ze zmiennych. Wynikać to może z faktu, że każde z tych państw jest atrakcyjne turystycznie, jednakże obywatele tych krajów stać również na wyjazdy zagraniczne. Kolejne dwa skupienia składają się z dwóch państw, kolejno: Finlandii i Szwecji oraz Hiszpanii i Francji. Klastry te odznaczają się wysokimi wartościami dla zmiennych $\text{procent_ojczyzna} - \text{noce_ojczyzna}$. Świadczy to o znacznym udziale obywateli w turystyce krajowej. Wynikać to może z rozbudowanej infrastruktury turystycznej w tych państwach. W przypadku Hiszpanii i Francji turystyce krajowej sprzyjają warunki atmosferyczne, a także dostęp do Morza Śródziemnego i terenów górskich. Finlandia i Szwecja to kraje skandynawskie, turystycznie różnorodne, z terenami górskimi i podbiegunowymi oraz licznymi jeziorami, co może skłaniać do podróżowania krajowego. Co ciekawe, państwa te uznawane są za kraje o wysokim poziomie życia, co mogłoby prowadzić do wniosków, że partycypacja w turystyce zagranicznej w tych państwach będzie na wysokim poziomie, jednakże rysunek 6 wskazuje, że poziom ten jest przeciętny. Świadczyć to może o odpowiednim wykorzystaniu potencjału turystycznego państw oraz odpowiednim promowaniu atrakcji w kraju. Kolejne skupienie tworzone jest przez Węgry, Włochy, Łotwę, Litwę, Chorwację i Słowację. Państwa te również odznaczają się przeciętnymi wartościami dla każdej ze zmiennych. Ciekawym jednak jest, że Włochy oraz Chorwacja, a więc państwa uznawane za jedno z głównych kierunków turystycznych, pogrupo-



Rys. 6. Podział na skupienia – klasyfikacja hierarchiczna
 Źródło: opracowanie własne na podstawie (Ec.europa.eu, 2022).



Rys. 7. Podział na skupienia – k-średnich

Źródło: opracowanie własne na podstawie (Ec.europa.eu, 2022).

wane zostały z Łotwą, Litwą, Węgrami i Słowacją, które są mniej popularne wśród turystów. Wynikać to może z wysokiego zatrudnienia obywateli w turystyce w przypadku Chorwacji oraz Włoch, co skutkuje przeciętnym poziomem aktywności turystycznej, porównywalnym do reszty państw w skupieniu. Ostatnie ze skupień tworzone jest przez Bułgarię, Rumunię, Grecję, Portugalię i Polskę. Państwa te odznaczają się niskimi wartościami dla zmiennych *procent_zagranica* – *noce_zagranica* oraz wysokimi wartościami dla zmiennych opisujących powody nieuczestniczenia w turystyce. Wynikać to może z niskich dochodów obywateli tych państw, co uniemożliwia im wyjazdy zagraniczne.

5. Zakończenie

Po przeprowadzeniu analiz udało się odpowiedzieć na postawione pytania badawcze. Obie metody badawcze, czyli klasyfikacja hierarchiczna i metoda k-średnich pogrupowały wybrane kraje w te same klastry. Kraje wysoko rozwinięte gospodarczo utworzyły osobne skupienia, czyli Luksemburg, Szwajcaria, Szwecja i Finlandia. Jednakże niektóre z państw o wysokim poziomie życia, jak Niemcy, Belgia i Holandia, pogrupowane zostały z Cyprzem czy Słowenią. Dwa pierwsze skupienia charakteryzują się wysokim udziałem turystyki zagranicznej, natomiast skupienie z krajami skandynawskimi odznacza się wysokim udziałem turystyki krajowej. Analiza klastrów nie wskazała wyraźnego podziału krajów na Europę Zachodnią i Wschodnią. Jednak kraje, takie jak Polska, Bułgaria, Rumunia, Portugalia i Grecja utworzyły jedno skupienie. Są to państwa z wysokim udziałem osób, które rezygnują z turystyki z powodów finansowych. Oznacza to, że wysokie koszty życia skłaniają do oszczędzania i wydawania pieniędzy na inne cele niż podróźnicze. Osobne skupienie utworzyły Hiszpania i Francja, które wyróżniają się wysokim udziałem turystyki krajowej i liczbą nocy spędzonych na wyjazdach w obrębie własnego kraju. Może to wynikać z faktu, iż kraje te mają bogatą ofertę turystyczną, atrakcyjne zaplecze historyczne i zróżnicowane położenie geograficzne. Ponadto przyjazne warunki atmosferyczne również skłaniają do podróży po tych krajach. Co ciekawe, Włochy nie należą do tego skupienia, a tworzą skupienie razem z Węgrami, Słowacją, Litwą, Łotwą i Chorwacją.

Literatura

- Alejski, W. (2011) Aktywność turystyczna: międzynarodowe i krajowe zróżnicowanie oraz kwestia wykluczenia społecznego. *Turyzm*, 21(1-2), 7-16.
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Ec.europa.eu. (2022). *Tourism*. Pobrano z <https://ec.europa.eu/eurostat/web/main/data/database>
- Kaczmarek, A. (2014). Wybrane czynniki rozwoju turystyki. *Studia Ekonomiczne*, (176), 202-215.

- Kiryłuk, H. (2014). Turystyka jako narzędzie rozwoju gospodarczego Unii Europejskiej. *Przedsiębiorczość i Zarządzanie*, 15(8), 137-154.
- Kwiatkowski, C., Harasim, E. i Hawlena, J. (2018). *Aktywność turystyczna – wybrane aspekty*. Instytut Naukowo-Wydawniczy „Spatium”.
- Ruane, M. C. M. (2014). Exchange Rate and Tourism: Evidence from the Island of Guam. *Journal of Economics and Economic Education Research*, 15(2), 165-185.
- Stanimir, A. (red.). (2006). *Analiza danych marketingowych. Problemy, metody, przykłady*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Steinley, D., i Brusco, M. J. (2008). A New Variable Weighting and Selection Procedure for K-means Cluster Analysis. *Multivariate Behavioral Research*, 43(1), 77-108, <https://doi.org/10.1080/00273170701836695>
- UN Tourism (b.d.a). *Glossary of tourism terms*. Pobrano 11 marca 2024 z <https://www.unwto.org/glossary-tourism-terms>
- UN Tourism (b.d.b.). *International Recommendations for Tourism Statistics*. Pobrano z <https://www.unwto.org/tourism-statistics/on-basic-tourism-statistics-irts-2008>
- Więckowski, M. (2010). *Turystyka na obszarach przygranicznych Polski*. Polska Akademia Nauk. Instytut Geografii i Przestrzennego Zagospodarowania.

Tourism Activity in Selected European Countries – Comparative Analysis

Abstract: The article presents an analysis of tourism activity in 24 European countries in 2022. The concept of tourism activity was defined, then the factors influencing it were described in the study. Two clustering methods were used for the analysis: hierarchical classification and *k*-means. The research problem is to find similarities between countries in Europe based on variables describing participation in tourism. The elbow method, based on WSS (within sum of square), was used to select the appropriate number of clusters. The analysis showed that some developed countries exhibit similarity in terms of tourist activity, however, some of these countries were grouped with those of lower quality of life, such as Slovenia or Hungary. The study did not show clear differences between Western and Eastern Europe. Both classification methods created the same clusters.

Keywords: cluster analysis, hierarchical clustering, *k*-means clustering, tourism, tourism activity

Aleksandra Ceglarska

e-mail: 181345@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Weryfikacja występowania anomalii kalendarzowych na rynku kryptowalut na przykładzie Bitcoina

DOI: 10.15611/2024.53.6.02

JEL Classification: G12, G14

© 2024 Aleksandra Ceglarska

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Ceglarska, A. (2024). Weryfikacja występowania anomalii kalendarzowych na rynku kryptowalut na przykładzie Bitcoina W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 25-40). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Ogromny postęp technologiczny oraz przeniesienie płatności do Internetu otworzyło zupełnie nowe możliwości w zarządzaniu finansami oraz dokonywaniu płatności. W 2008 roku nastąpił przełom, gdyż powstała nowa koncepcja płatności elektronicznych – Bitcoin, który szybko stał się najpopularniejszą kryptowalutą na świecie. Obecnie tematy dotyczące Bitcoina i innych kryptowalut są bardzo aktualne, zwłaszcza pytanie, czy Bitcoin może funkcjonować jako środek płatniczy i pieniądź. Rynek finansowy ściśle związany jest z pojęciem ryzyka, dlatego dokonano analizy Bitcoina w tym kontekście, określając ryzyko jako odchylenie standardowe stóp zwrotu notowań Bitcoina. Przeprowadzony test bootstrap pozwolił ustalić, że ryzyko inwestycji w Bitcoin różni się w poszczególnych podokresach czasowych. Zidentyfikowano zależności między rytmem dnia typowego inwestora a anomaliami godzinowymi Bitcoina, uznano więc, że tematyka ta jest wciąż aktualna i warto przeprowadzić dalsze badania w tym zakresie.

Słowa kluczowe: Bitcoin, kryptowaluty, efekty kalendarzowe, ryzyko inwestycyjne, test bootstrap

1. Wstęp

Pojęcie transakcji, czyli po prostu wymiany towaru lub usługi, towarzyszyło ludzkości od zarania dziejów. Koncepcja środka pieniężnego ewaluowała od zboża, soli, koralików czy muszli poprzez pierwsze wybite w zachodniej Turcji monety i pieniądź papierowy aż do płatności internetowych, kart płatniczych oraz tak abstrakcyjnych jeszcze 20 lat temu płatności telefonem lub zegarkiem. Przeniesienie wielu codziennych działań do Internetu (w tym właśnie płatności) wynika z ogromnego postępu informatyzacji w społeczeństwie (Dopierała i Borodo, 2014).

W 2008 roku powstała jednak zupełnie nowa koncepcja płatności elektronicznej, która od tamtego czasu rozwinęła się do niesamowitych rozmiarów. Mowa tutaj oczywiście o Bitcoinie umożliwiającym internetowy transfer pieniędzy bez pośrednictwa żadnej instytucji finansowej, którego twórcą jest Satoshi Nakamoto. Tematyka kryptowalut, a szczególnie Bitcoina, ostatnimi czasy jest bardzo aktualna. Coraz więcej osób zadaje sobie pytanie: Czy Bitcoin może być środkiem płatniczym i pełnić funkcję pieniądza? Istnieją kraje, które rozważają wprowadzenie Bitcoina jako oficjalnego środka płatniczego, a prekursorem w tej dziedzinie jest rząd Salwadoru, który zdecydował się na ten krok 6 września 2021 roku. Od tamtego czasu w Salwadorze Bitcoinem można zapłacić za mieszkanie, zakupy, usługi, a nawet kolację w restauracji. Pomimo negatywnych skutków ekonomicznych i zadłużenia kraju na skutek wprowadzenia kryptowaluty zwolenników takiej idei nie brakuje. Nic zresztą dziwnego, skoro do końca 2021 roku około 34 miliony obywateli Stanów Zjednoczonych (ponad 10% całego społeczeństwa) posiadało w swoim portfelu jakąś kryptowalutę (Bińkowski, 2022). W 73% przypadkach był nią właśnie Bitcoin. Dzięki tak dużemu zainteresowaniu tematem kryptowalut naukowcy i analitycy coraz częściej przeprowadzają badania, aby sprawdzić, co wpływa na zmienność Bitcoina, oraz czy istnieją jakieś trendy lub prawidłowości wynikających z anomalii kalendarzowych opisujących występowanie pewnych regularnych trendów i wzorców związanych z określonymi dniami tygodnia („efekt poniedziałku”), miesiącami („efekt stycznia”) lub okresami w roku („efekt końca kwartału”). Takie anomalie mogą wpływać na kursy różnych instrumentów finansowych, zaprzeczając tym samym teorii rynku efektywnego.

Celem pracy jest określenie, w jaki sposób wartości ryzyka Bitcoina kształtują się w ciągu dnia, oraz stwierdzenie występowania ewentualnych prawidłowości związanych z rytmem dnia inwestorów. Postanowiono przeprowadzić weryfikację występowania anomalii kalendarzowych na rynku kryptowalut na przykładzie Bitcoina w zakresie ryzyka mierzonego jako odchylenie standardowe stóp zwrotu. Postawiono następujące pytania badawcze:

1. Czy istnieją godziny, w których ryzyko Bitcoina regularnie wzrasta lub maleje?
2. Czy ryzyko Bitcoina wzrasta, gdy statystycznie wszyscy są w pracy, czy wtedy gdy raczej wszyscy śpią?
3. W których godzinach w ciągu doby średni poziom ryzyka Bitcoina jest największy, a w których najmniejszy? Z czego może to wynikać?

Aby móc udzielić odpowiedzi na zadane pytania badawcze, podjęto szereg działań mających na celu jak najlepsze zrozumienie działania Bitcoina, jego trendów i związanych z nim zależności, a także przeprowadzono analizy, zastosowano metody obliczeniowe, zinterpretowano odpowiednie wykresy oraz przeprowadzono testy statystyczne. Głównym etapem badania jest wykonanie testu bootstrap i ostateczne potwierdzenie występowania anomalii godzinowych ryzyka inwestycji w Bitcoin.

2. Ryzyko inwestycji w kontekście instrumentów finansowych

2.1. Funkcjonowanie rynku finansowego w kontekście ryzyka

Rynek finansowy stwarza przestrzeń dla transakcji kupna i sprzedaży dokonywanych za pomocą różnych instrumentów finansowych. Dzięki podaży utworzonej przez sprzedających oraz popytowi zależnemu od kupujących lub inwestorów, ceny aktywów finansowych, takich jak akcje lub obligacje, często się zmieniają, podobnie jak ogólne warunki rynkowe. Rynek ten skupia inwestorów, którzy posiadają wolne środki pieniężne i w momencie kupna danego instrumentu finansowego zamieniają wolny pieniądź w kapitał inwestycyjny, którego posiadanie wiąże się z prawdopodobieństwem uzyskania bardzo dużych korzyści, ale również z dużym ryzykiem (Business Insider, 2021). Na każdym rynku finansowym: pieniężnym, kapitałowym, walutowym, instrumentów pochodnych lub depozytowo-walutowym dokonywane transakcje obciążone są w pewnym stopniu ryzykiem. Terminem ryzyka określa się pewne niebezpieczeństwo dotyczące możliwości nieuzyskania oczekiwanych sukcesów lub nawet możliwego poniesienia nieplanowanych strat (Bijańska i Wodarski, 2014). Istnieje kilka rodzajów ryzyka: ryzyko stopy procentowej, ryzyko kursu walutowego, ryzyko kredytowe, ryzyko polityczne i ryzyko płynności (Porcenałuk, 2013). Każdy inwestor dąży do zmniejszenia tego ryzyka na różne sposoby. Możliwa jest dywersyfikacja portfela (rozbicie kapitału na odmienne inwestycje) lub zakup instrumentów finansowych o krótkim terminie wykupu i dużej płynności (Jajuga, 2006).

2.2. Waluta tradycyjna i waluta kryptograficzna a poziom ryzyka

Poziom ryzyka zależy od wielu czynników: warunków rynkowych, czasu trwania inwestycji i przeprowadzonej analizy rynku oraz rozeznania się na nim. Jednak głównym czynnikiem, który na takie ryzyko wpływa, jest niewątpliwie rodzaj instrumentu finansowego, w który się inwestuje. Różnice w kontekście ryzyka zaobserwowano w przypadku porównania waluty tradycyjnej z walutą kryptograficzną (tab. 1).

Tabela 1. Porównanie waluty tradycyjnej i kryptograficznej

| Waluta tradycyjna | Waluta kryptograficzna |
|---|---|
| Handel odbywa się tylko w określonych godzinach. | Handel odbywa się całą dobę. |
| Długi czas transakcji. | Bardzo krótki czas transakcji. |
| Tworzona przez rządy państwowe. | Tworzona przez specjalne oprogramowanie. |
| Kontrolowana przez instytucje finansowe, banki i rządy. | Nie jest kontrolowana przez scentralizowane instytucje. |
| Materialna forma: banknoty oraz monety. | Brak materialnej formy, jest to fragment kodu. |
| Raczej nie są spekulacyjne. | Waluta spekulacyjna. |
| Utrzymuje optymalną wartość, jest dosyć stabilna. | Bardzo wysokie amplitudy wartości, niestabilność. |

Źródło: opracowanie własne.

Wszystkie wymienione w tabeli 1 aspekty walut w istotny sposób wpływają na różnice w ryzyku pomiędzy nimi.

Bitcoin jest w dużej mierze walutą spekulacyjną, o dużych amplitudach wartości i dosyć małej stabilności, przez co ryzyko inwestycji w niego jest dosyć wysokie. Porównując ryzyko inwestycji na działających w Polsce giełdach walut tradycyjnych i giełdach kryptowalut, dochodzi się do wniosku, że ryzyko przy inwestowaniu w waluty tradycyjne jest duże niższe niż w sytuacji inwestowania, np. w Bitcoin (Kądziołka, 2015).

3. Metodyka badań

3.1. Narzędzia zastosowane do charakterystyki rozkładów

Rynek finansowy jest dynamiczny i podlega ciągłym zmianom. Na jego różnorodność wpływa wiele czynników: różne typy inwestorów mających różne cele i strategie inwestycyjne, występowanie asymetrii informacji, podatność rynku na zmiany polityczne, katastrofy naturalne lub wydarzenia gospodarcze czy regulacje prawne. Tak więc również szeregi finansowe są nieprzewidywalne i może w nich występować ryzyko pojawienia się obserwacji odstających, np. w wyniku błędnych pomiarów, manipulacji i oszustw, losowych fluktuacji lub gwałtownych zmian w danych finansowych. Eliminacja takich obserwacji jest szczególnie ważna przy mało licznych zbiorach danych (Budka i in., 2013) lub przy analizie danych o wysokiej częstotliwości.

Podstawowymi parametrami pozwalającymi na wychwycenie obserwacji odstających jest kurtoza, która w przeciwieństwie do mediany lub kwartylii jest czuła na obserwacje nietypowe. Wartość średniej także może być zniekształcona nawet z powodu pojedynczej obserwacji odstającej. Dopiero 50% dużych błędów może zaburzyć wartość mediany, w przypadku średniej próg wynosi 0% (Trzpiot, 2013). Zwracać uwagę powinna szczególnie dodatnia wartość kurtozy, wskazująca na rozkład leptokurtyczny, w którym jest większe prawdopodobieństwo wystąpienia obserwacji odstających niż w przypadku rozkładu normalnego. Aby zlokalizować obserwacje nietypowe można posłużyć się również wykresem punktowym oraz wykresem pudełkowym, który przedstawia poziom rozproszenia lub koncentracji danych. Na jego podstawie da się określić zarówno obserwacje centralne, jak i obserwacje odstające (Kosiorowski i in., 2018). Nie jest on jednak metodą doskonałą. W przypadku rozkładów skośnych przedstawienie na nim obserwacji odstających może być wątpliwe, dlatego wciąż powstają nowe modyfikacje tego wykresu (Walker i in., 2018).

Nadrzędne znaczenie ma scharakteryzowanie skośności potencjalnie występującej w rozkładach stóp zwrotu (Piontek, 2007). Najbardziej popularnym sposobem określenia asymetrii rozkładu jest wyznaczenie współczynnika asymetrii (współczynnika skośności). Miara ta jednak nie jest jednoznaczna, gdy przyjmuje wartość zero, które powinno wskazywać na występującą symetrię rozkładu. Istnieją takie

rozkłady, dla których pomimo występującej asymetrii, współczynnik również przyjmuje wartość zero. Współczynnik SK , przyjmujący wartości przybliżone do zera, jednoznacznie wskazuje na symetrię rozkładu (Piontek, 2007). Jest on wyrażony następującym wzorem:

$$SK = \frac{m_3}{m_2^{3/2}}, \quad (1)$$

gdzie: m_i to i -ty moment centralny.

Należy przeprowadzić test Jarque-Bera, który w klasycznej wersji badając normalność rozkładu, uwzględnia zarówno skośność, jak i kurtozę. Standardowa wartość statystyki Jarque-Bera jest odpowiednia tylko dla danych, których rozkład jest zbliżony do rozkładu normalnego. Aby móc ją wykorzystać w przypadku analizowanych danych, mianownik należy skorygować o grubość ogonów. Statystyka, jak w przypadku tradycyjnej wersji, jest przybliżona rozkładem χ^2 , jednak o jednym, a nie dwóch stopniach swobody, i przyjmuje następującą postać (Godfrey i Orme, 1991):

$$RS = n \frac{(\sqrt{b_1})^2}{[9 + m_6 m_2^{-3} - 6m_4 m_2^{-2}]}, \quad (2)$$

gdzie: m_i to odpowiednie i -te momenty zwykłe rozkładu, $\sqrt{b_1}$ to wartość miary skośności (SK), a n to wielkość próby.

Dzięki skorygowaniu mianownika zmniejsza się statystyka testowa, więc w mniejszej liczbie przypadków odrzuca się hipotezę zerową świadczącą o symetrii rozkładu.

Jeśli statystyka RS jest większa od wartości krytycznej z rozkładu χ^2 , istnieją podstawy do odrzucenia hipotezy zerowej o symetrii i normalności rozkładu. Jeśli statystyka RS jest mniejsza od wartości krytycznej, to nie istnieją przesłanki do odrzucenia hipotezy zerowej, rozkład jest symetryczny i normalny.

Powyższe analizy skośności i kurtozy mają na celu udowodnienie, że klasyczne testy zakładające normalność rozkładu nie mogą zostać użyte w tej analizie. Skorzystano więc z testu bootstrap, który nie wymaga spełnienia założeń o normalności rozkładów.

3.2. Zastosowanie metod bootstrapowych

Metody bootstrapowe znajdują zastosowanie przede wszystkim w analizie i prognozowaniu szeregów czasowych. Ich główne założenie dotyczy traktowania próby bootstrapowej składającej się z losowo wygenerowanych elementów, tak samo jak wylosowaną próbkę dla całej populacji (Gulgowski i Włosik, 2013). Metody bootstrapowe należą do metod symulacyjnych (Pekasiewicz i in., 2014) i są stosowane do oszacowania nieznanymi wartościami powiązanych z modelami statystycznymi (Boos, 2003). Bootstrap po raz pierwszy został wykorzystany przez B. Efrona w 1977 r.

jako pomocne narzędzie, zwłaszcza w bardziej zaawansowanych modelach. Metody bootstrapowe pozwalają na szerszą analizę wyników w porównaniu do stosowania metod klasycznych (Foszcz, 2006). Wyższość tej metody ponad metodami klasycznymi jest potwierdzana w kilku punktach (Hesterberg i in., 2003):

1. Wymaga mniej założeń. Metoda bootstrap nie wymaga potwierdzenia występowania rozkładu normalnego, dużej lub małej próby badawczej.
2. Pozwala na większą dokładność wyników.
3. Łatwość w zrozumieniu mechanizmów. Metody bootstrapowe są intuicyjne i budują pewną analogię, dzięki czemu są zrozumiałe dla badaczy i odbiorców.

Biorąc pod uwagę wszystkie zalety i możliwości testu bootstrap, zdecydowano się go użyć w przeprowadzanej analizie. Zapewni on rzetelność wyników niezależnie od rozkładu obserwacji bądź ich dużej liczby. Metoda bootstrap polega na wylosowaniu kolejnych próbek, wzorując się na próbce wyjściowej o takiej samej liczbie obserwacji jak ta próbka. Jest to proces losowania ze zwracaniem, co oznacza, że możliwe jest wielokrotne wylosowanie tego samego obiektu. Wartości średniej z wybranych próbek większe od statystyki testowej wyrażają p -wartość, która z kolei na założonym poziomie istotności pozwala odrzucić lub przyjąć hipotezę zerową i określić, czy porównywane próbki różnią się od siebie, czy nie. Definiuje się hipotezę zerową:

$$H_0 : \text{Próbki istotnie nie różnią się od siebie,}$$

wobec hipotezy alternatywnej:

$$H_1 : \text{Próbki istotnie się od siebie różnią.}$$

Są to ogólne hipotezy opisujące test bootstrap, natomiast, aby móc dokładniej przeprowadzić badania, sformułowano dodatkowe hipotezy:

1. Dla porównania średnich próbek:

$$H_0 : \hat{X}_1 = \hat{X}_2,$$

wobec hipotezy alternatywnej:

$$H_1 : \hat{X}_1^1 \hat{X}_2,$$

2. Dla porównania median próbek:

$$H_0 : \hat{M}_{e_1} = \hat{M}_{e_2},$$

wobec hipotezy alternatywnej:

$$H_1 : \hat{M}_{e_1}^1 \hat{M}_{e_2},$$

Dla zaprezentowanych hipotez określono osobne statystyki testowe:

Test 1 – porównujący średnie:

$$\left| \hat{X}_1 - \hat{X}_2 \right|$$

Test 2 – porównujący mediany:

$$\left| \hat{M}_{e_1}^1 - \hat{M}_{e_2} \right|$$

Test bootstrap zakłada, że dwie wybrane próby nie różnią się od siebie. Postawiono połączyć ze sobą dwie badane próbki i na podstawie całego zbioru określić rozkład statystyki testowej. Następnie wartość tę trzeba porównać z wartością statystyki uzyskanej dla wygenerowanych próbek bootstrap. Potem w próbie badawczej należy odnaleźć obiekty, których wartości są większe od statystyki testowej. Z odnalezionych próbek oblicza się średnią, stanowiącą p -wartość. Jeśli p -wartość jest mniejsza od przyjętego poziomu istotności, to istnieją podstawy do odrzucenia hipotezy zerowej – próbki istotnie różnią się od siebie. Jeśli p -wartość jest większa od przyjętego poziomu istotności, to nie ma podstaw do odrzucenia hipotezy zerowej, czyli próbki istotnie się od siebie nie różnią.

4. Wyniki badań

4.1. Znaczenie podstawowych parametrów rozkładu ryzyka stóp zwrotu Bitcoina

Aby przeprowadzić analizę wykorzystano pięciominutowe notowania Bitcoina w okresie od 8.11.2022 (10:55) do 8.02.2023 (10:55), dzięki czemu uzyskano ponad 26 tysięcy obserwacji¹. Dane pozyskano z bazy internetowej Eikon, która jest platformą finansową firmy Refinitiv, zawierającą szeroki zakres danych rynkowych dostarczanych z giełd na całym świecie. Godziny dotyczą czasu GMT+1².

W celu uporządkowania pięciominutowych notowań Bitcoina postanowiono podzielić wszystkie obserwacje na sześć okresów, każdy zawierający dane z czterech godzin.

Każdy przedział zawiera 92 obserwacje – tyle, ile dni liczy okres badawczy:

1. 8:00-11:55
2. 12:00-15:55
3. 16:00-19:55
4. 20:00-23:55
5. 24:00-3:55
6. 4:00-7:55

¹ dokładnie 26 498

² GMT – *Greenwich Mean Time*, czyli czasu w Greenwich w Londynie

Dla każdego z tych przedziałów obliczono odchylenie standardowe stóp zwrotu:

$$\frac{x_{t+1} - x_t}{x_t}, \quad (3)$$

gdzie x_t to wartość notowania Bitcoina w dniu x .

Podział doby w ten konkretny sposób motywowany jest najbardziej powszechnym rytmem dnia typowego przedstawiciela społeczeństwa.

Okresy czasu dzielą dobę na okresy, kiedy ludzie zazwyczaj są w pracy (od 8:00 do 11:55 oraz od 12:00 do 15:55), kiedy po pracy i odpoczywają, oddają się pasjom lub aktywnościom fizycznym (od 16:00 do 19:55), okres wieczorny przeznaczony na relaks oraz sen (od 20:00 do 7:55). Wydawać by się mogło, że okresami o największym ryzyku Bitcoina będą trzeci oraz czwarty przedział czasowy, kiedy inwestor – typowy przedstawiciel społeczeństwa – może usiąść do komputera, telefonu oraz zająć się swoimi inwestycjami, w tym Bitcoinem. Przewiduje się więc, że to okres 3 i 4 będą okresami o największym ryzyku Bitcoina, gdyż duża aktywność inwestorów w sieci internetowej wiąże się z dużym rozproszeniem podejmowanych decyzji kupna-sprzedaży. W podokresach 1 i 2 analogicznie ryzyko powinno być najmniejsze (ze względu na wykonywaną w tym czasie pracę i obowiązki). Aby móc potwierdzić te przypuszczenia, zdecydowano się na przeprowadzenie szeregu analiz umożliwiających potwierdzenie lub odrzucenie supozycji badacza.

Kluczowym elementem tego badania było obliczenie ryzyka inwestowania mierzonego jako odchylenie standardowych stóp zwrotu w poszczególnych podokresach, aby móc stwierdzić, czy występują w nich prawidłowości rozkładu analizowanej zmiennej. Do obliczeń użyto programu R i RStudio, który wygenerował tabelę z odchyleniami w każdym podokresie dla poszczególnych dni. Podstawowe statystyki opisowe przedstawiono w tabeli 1.

Tabela 2. Podstawowe statystyki opisowe ryzyka stóp zwrotu Bitcoina

| Statystyki | Numer przedziału czasowego | | | | | |
|----------------|----------------------------|------------------|-------------|-------------|------------|-----------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Przedział godzinowy | | | | | |
| | 8:00-11:55 | 12:00-15:55 | 16:00-19:55 | 20:00-23:55 | 24:00-3:55 | 4:00-7:55 |
| Średnia | 0,0012788 | 0,0016133 | 0,0012549 | 0,0012762 | 0,0012976 | 0,0008549 |
| Odchyl. stand. | 0,0030311 | 0,0031726 | 0,0009960 | 0,0010435 | 0,0011498 | 0,0007745 |
| Mediana | 0,0006891 | 0,0009375 | 0,0009946 | 0,0009000 | 0,0009186 | 0,0006665 |
| Kurtoza | 76,726567 | 61,762482 | 7,1909804 | 8,6896287 | 10,876394 | 15,143793 |
| Skośność | 8,4570679 | 7,3691391 | 2,4582007 | 2,5039807 | 2,7641817 | 3,6172936 |

Źródło: opracowanie własne.

Analizując tabelę 2, można stwierdzić, że największe średnie wartości występują w podokresie 2. Niewiele mniejsze można zaobserwować w podokresach 1, 3, 4 i 5.

Najmniejsze przeciętne wartości Bitcoina widoczne są dla podokresu ostatniego, zawierającego obserwacje od godziny 4:00 do 7:55. Prawdopodobnie niskie wartości wynikają w tym przypadku z najmniejszej aktywności społeczeństwa w tym przedziale godzinowym. Zdecydowana większość osób, które żyją w analizowanej strefie czasowej, nie korzysta wtedy z Internetu, nie kupuje kryptowalut, a cały rynek po prostu jest nieaktywny. Analogicznie można stwierdzić, że podokres 2 w godzinach od 12:00 do 15:55 jest z kolei okresem największej aktywności społeczeństwa. To wtedy zdecydowana większość ludzi pracuje, uczy się, korzysta z komputera i Internetu, śledzi giełdę, czyta wiadomości oraz przeprowadza operacje finansowe, w tym na rynku kryptowalut. Już na podstawie tak prostych parametrów jak średnia można zacząć potwierdzać hipotezę o wpływie rytmu dnia społeczeństwa na wartości ryzyka Bitcoina. Uwagę przyciągają większe wartości odchylenia standardowego dla dwóch pierwszych podokresów. Oznacza to większe zróżnicowanie wartości względem średniej i możliwość występowania obserwacji odstających. Kolejnym aspektem, który powinien zwrócić uwagę, jest wysoka wartość kurtozy w dwóch pierwszych podokresach. Jest to wyraźna przesłanka o występowaniu wartości ekstremalnej, zwłaszcza biorąc pod uwagę fakt, iż jest to rozkład leptokurtyczny (dodatnia wartość kurtozy), charakteryzujący się większym ryzykiem występowania ogonów rozkładu (i obserwacji odstających) niż rozkład normalny.

Przeanalizowano wartość skośności obliczonej zgodnie ze wzorem 4:

$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3, \quad (4)$$

gdzie: n to liczba obserwacji, x_i to i -ta wartość w zbiorze danych, s to odchylenie standardowe, a \bar{x} to średnia arytmetyczna zbioru danych.

Stwierdzono występowanie asymetrii prawostronnej we wszystkich przedziałach, co oznacza, że wydłużony jest prawy ogon rozkładu. Wartości średniej oraz kurtozy dla podokresu 1 i 2 wyraźnie wskazują na występowanie obserwacji odstających, które należy znaleźć w dalszej analizie. W związku z powyższym sformułowano następujące wnioski:

1. Bitcoin przyjmował średnio największe wartości w godzinach 12:00-16:00, w środku dnia, kiedy prawdopodobnie największa liczba osób korzysta z Internetu, śledzi rynki akcji oraz obserwuje zachowania kryptowalut, w tym Bitcoina.
2. Najmniejsze średnie wartości Bitcoina zauważalne są w godzinach 4:00-8:00. Jest to przedział czasowy, w którym zdecydowana większość społeczeństwa śpi bądź szykuje się do rozpoczęcia nowego dnia, a więc nie ma czasu wykonywać transakcji i wpływać na rynek kryptowalut, w tym Bitcoina.
3. Konieczna jest szczegółowa analiza dwóch pierwszych przedziałów pod kątem występowania w nich obserwacji odstających.

Poddano analizie wykresy pudełkowe oraz punktowe dla dwóch pierwszych przedziałów i zlokalizowano obserwacje odstające. Obserwacje takie powinno się

usunąć ze zbioru danych, jednak w tym przypadku ich brak znacząco zaburzyłby stworzone przedziały i mógłby negatywnie wpłynąć na wiarygodność wyników. Aby zachować ciągłość obserwacji, postanowiono za nietypowe obserwacje podstawić wartości uśrednione (średnia z pozostałych obserwacji dla konkretnego podokresu). Ponownie poddano analizie oba typy wykresów i nie stwierdzono więcej komplikacji mogących utrudnić całe badanie.

4.2. Test Jarque-Bera

Postanowiono zbadać symetrię oraz normalność rozkładu stóp zwrotu Bitcoina. Wykorzystano współczynnik SK (1) oraz zmodyfikowany wzór Jarque-Bera (2). Otrzymało następujące wyniki:

Tabela 3. Wyniki testu Jarque-Bera

| Przedziały godzinowe | SK | RS |
|----------------------|----------|----------|
| 8:00-11:55 | 2,344967 | 8,024767 |
| 12:00-15:55 | 3,284107 | 3,694340 |
| 16:00-19:55 | 2,458201 | 5,887322 |
| 20:00-23:55 | 2,503981 | 2,987919 |
| 24:00-3:55 | 2,764182 | 2,479858 |
| 4:00-7:55 | 3,617294 | 3,650702 |

Źródło: opracowanie własne.

Współczynnik SK w każdym przypadku jest większy od 0, co oznacza, że rozkład odchyleń zwrotu Bitcoina nie jest symetryczny. Statystyka testowa RS dla podokresu 1 i 3 jest większa od wartości krytycznej CV, która wyniosła 3,841459, więc dla tych przedziałów nie występuje symetria rozkładu. W przypadku pozostałych podokresów wartość RS jest mniejsza od wartości krytycznej, nie ma więc podstaw do odrzucenia hipotezy zerowej o symetrycznym rozkładzie ryzyka stóp zwrotu Bitcoina.

4.3. Wykrywanie różnic za pomocą testu bootstrap

Ze względu na wyniki powyższych testów nie można skorzystać z metod klasycznych. Wykorzystano więc test bootstrap, który będzie przeprowadzany kolejno dla wszystkich przedziałów następujących po sobie parami. Do obliczeń wykorzystano program R i RStudio. Pierwszym etapem było obliczenie średniej z obu badanych próbek, a następnie obliczenie statystyki testowej będącej wartością bezwzględną różnicy średniej próbek. Następnie należało obliczyć medianę i również wyznaczyć statystykę testową (wartość bezwzględna różnicy median dwóch próbek).

Następnie stworzono wektor składający się z połączonych wartości przedziału 1 i 2, aby przetestować, czy jego pierwsza połówka (przedział 1) różni się od drugiej połówki (przedział 2), czyli czy średnia wartość ryzyka jest inna w podokresie 1 niż w podokresie 2.

Kolejnym krokiem jest określenie, jakie jest prawdopodobieństwo otrzymania pierwszej i drugiej statystyki testowej większej lub równej obserwowanej statystyce testowej przy założeniu, że hipoteza zerowa jest prawdziwa i statystyka testowa jest zbliżona do 0.

W tym celu sprawdzono, ile z 10 tysięcy próbek bootstrap przyjmuje wartości niemniejsze od pierwszej statystyki testowej (dla średniej). Następnie obliczono średnią z wszystkich próbek przyjmujących równe lub większe wartości od obserwowanej statystyki testowej i w ten sposób otrzymano p-wartość. Te same działania powtórzono również dla mediany. Wyniki dla wszystkich podokresów przedstawiono w tabeli 4.

Tabela 4. Wyniki testu bootstrap

| Para przedziałów czasowych | Test średniej | | Test mediany | |
|----------------------------|--------------------|-----------|--------------------|-----------|
| | statystyka testowa | p-wartość | statystyka testowa | p-wartość |
| 1 i 2 | 0,00033690 | 0,0454 | 0,0002484 | 0,0137 |
| 2 i 3 | 0,00005412 | 0,7440 | 0,0000571 | 0,5430 |
| 3 i 4 | 0,00002132 | 0,9151 | 0,0000945 | 0,4812 |
| 4 i 5 | 0,00002100 | 0,8928 | 0,0000190 | 0,8445 |
| 5 i 6 | 0,00044270 | 0,0029 | 0,0002521 | 0,0019 |
| 6 i 1 | 0,00011722 | 0,4595 | 0,0001172 | 0,8329 |

Źródło: opracowanie własne.

Analizując powyższe wartości na przykładzie podokresów 8:00-11:55 i 12:00-15:55, można stwierdzić, że:

- w przypadku testu średnich p-wartość wyniosła 4,54%, więc zakładając poziom istotności 5%, istnieją podstawy do odrzucenia hipotezy zerowej, ryzyko inwestycji w Bitcoin różni się w badanych okresach;
- w przypadku testu median p-wartość wyniosła 1,37%, co oznacza, że z 10 tysięcy analizowanych próbek bootstrap 137 przyjmuje wartości większe od statystyki testowej. 1,73% jest mniejszy od przyjętego poziomu istotności, więc istnieją podstawy do odrzucenia hipotezy zerowej – próbki różnią się od siebie.

Patrząc analogicznie na pozostałe przykłady, można zauważyć, że jedynie podokres 1 i 2 oraz 5 i 6 istotnie się od siebie różnią pod względem wielkości ryzyka Bitcoina. Pomiędzy pozostałymi podokresami badanymi parami taka zależność nie występuje.

4.4. Poziom ryzyka w podokresach a poziom ryzyka w całym okresie

Test bootstrap pozwala jedynie na określenie, czy podokresy różnią się od siebie, w tym przypadku w kontekście ryzyka mierzonego jako odchylenia standardowe stóp zwrotu. Na tym etapie nie wiadomo jednak, w których przedziałach ryzyko to maleje lub rośnie. Aby odpowiedzieć na to pytanie, należy wykorzystać podstawowy parametr rozkładu – średnią. Średnie obliczone dla poszczególnych przedziałów należy porównać ze średnią obliczoną dla wszystkich obserwacji razem, aby móc stwierdzić, w których przedziałach ryzyko Bitcoina, mierzone jako odchylenie standardowe stóp zwrotu, jest większe, a w których mniejsze od średniego poziomu ryzyka w całym okresie. Średnie przedstawiono w tabeli 5. Jej wartości dla podokresu 8:00-11:55 i 12:00-15:55 różnią się od tych zawartych w tabeli 2 ze względu na występujące w tych podokresach obserwacje odstające. Dwie obserwacje odstające zostały zidentyfikowane i skorygowane poprzez zastąpienie ich wartości średnią z wartości poprzedzającej i następującej po nich, dzięki czemu wyniki badań oraz wartości parametrów są bardziej precyzyjne, wiarygodne i rzetelne.

Tabela 5. Poziom ryzyka w poszczególnych przedziałach czasowych

| Przedział czasowy | Średnia |
|-------------------|---------|
| Cały okres | 0,1161 |
| 8:00-11:55 | 0,0972 |
| 12:00-15:55 | 0,1309 |
| 16:00-19:55 | 0,1255 |
| 20:00-23:55 | 0,1276 |
| 24:00-3:55 | 0,1298 |
| 4:00-7:55 | 0,0855 |

Źródło: opracowanie własne.

Analizując tabelę 4, można stwierdzić, że podokresy pierwszy i drugi istotnie się od siebie różnią. Średnia od godziny 8:00 do 11:55 jest mniejsza od średniej w całym okresie, zatem przeciętne ryzyko w tym przedziale jest niższe niż w całym badanym okresie. Od godziny 12:00 ryzyko zdecydowanie rośnie i utrzymuje wartość wyższą od średniej do godziny 3:55. Pomiędzy przedziałami zawierającymi godziny od 12:00 do 3:55 nie występują istotne różnice – w każdym z nich średnia podokresu jest wyższa od średniej całkowitej, więc przeciętne ryzyko jest wyższe niż w całym badanym okresie. Potwierdzić także można różnicę zidentyfikowaną w teście bootstrap pomiędzy przedziałem 5 i 6. Dopiero od godziny 4:00 następuje spadek ryzyka poniżej ogólnej średniej. Podsumowując, najmniejsze ryzyko inwestycji w Bitcoin mierzone jako odchylenie standardowe stóp zwrotu zauważalne jest od godziny 4:00 do 11:55. Jest to okres w ciągu doby, kiedy zdecydowana większość społeczeństwa śpi (4:00-7:55) lub pracuje (8:00-11:55), więc nie ma czasu korzystać z Internetu i ewentualnie interesować się kryptowalutami, w tym Bitcoinem. Okres przedpołudniowy, czyli właśnie do godziny 12:00, jest przez naukowców uważany

za najbardziej produktywny, kiedy warto podjąć się trudniejszych zadań, odbywać spotkania, wtedy nasz mózg pracuje najlepiej. Według badań zaraz po przebudzeniu u większości ludzi można odnotować dosyć wysoki poziom hormonu stresu, co powoduje, że mózg jest bardziej chłonny (Brueck, 2020; More, 2004). W tym czasie więc społeczeństwo nie interesuje się Bitcoinem.

W przedziałach czasowych zawierających godziny od 12:00 do 3:55 ryzyko inwestycji w Bitcoin wyraźnie wzrasta. Najlogicznym założeniem byłoby stwierdzenie, że do godziny 16:00 ryzyko przyjmuje dosyć niskie wartości ze względu na standardowe godziny pracy. Okazuje się jednak, że jest to podokres najbardziej aktywny. Można to wytłumaczyć teorią, że po godzinie 12:00 mózg staje się mniej chłonny, ludzie potrzebują przerwy i często sięgają po urządzenia mobilne, aby odpocząć po dużym intelektualnym wysiłku. Wtedy również następuje coraz popularniejsza przerwa na lunch, podczas której również społeczeństwo może wpływać na Bitcoin. Około godziny 16:00 wartość kryptowaluty spada w porównaniu z podokresem 2, jednak potem stopniowo się podnosi. Zaskakujące jest to, że większe zainteresowanie Bitcoinem (tłumaczone wzrostem ryzyka Bitcoina) występuje w przedziale godzinowym 24:00-3:55 niż od godziny 16:00 do 23:55. Prawdopodobnie wynika to z faktu, że Bitcoinem zainteresowani są w znacznej mierze ludzie młodzi, których rytm życia może różnić się od tego założonego przez badacza. Warto przeanalizować to zagadnienie pod kątem różnych godzin otwarcia giełd w różnych istotnych krajach, co może wpływać na analizę oraz interpretację wyników.

5. Zakończenie

Pomimo istnienia ponad 4 tysięcy kryptowalut, które codziennie można kupować, wymieniać i sprzedawać, jedno jest pewne – rynek ten nie osiągnął jeszcze swojego maksymalnego potencjału. Wciąż będzie się rozwijał, ewoluował i modernizował przez wiele kolejnych lat i zapewne powstanie jeszcze dużo więcej kryptowalut. Warto podkreślić, że mimo istnienia licznych prac badawczych (np. Fauzi i Othman, 2020; Choithani, 2024) koncentrujących się na analizie i badaniu rynku kryptowalut nadal jest potrzeba jego dalszej eksploracji. Badanie kryptowalut w nowych kontekstach i aspektach jest konieczne, gdyż nie ulega wątpliwości, że ich wpływ na rynek finansowy oraz gospodarkę pieniężną całego świata będzie stale wzrastał.

Przeprowadzenie wszystkich powyższych analiz i testów pozwoliło z sukcesem odpowiedzieć na zadane pytania badawcze. Stwierdzono, że występują różnice pomiędzy czterogodzinnymi podokresami czasu w ciągu doby. Oznacza to, że ryzyko mierzone jako odchylenie standardowe stóp zwrotu różni się od siebie w kolejnych godzinach doby. Przeprowadzenie testu bootstrap pozwoliło stwierdzić, że podokres 2 (12:00-15:55) istotnie różni się od okresu go poprzedzającego (8:00-11:55), ale jest podobny do okresu po nim następującego (16:00-19:55). Podokres 5 (24:00-3:55) istotnie różni się od okresu po nim następującego (4:00-7:55) i jest podobny do okresu go poprzedzającego (20:00-23:55).

Test bootstrap pozwala jedynie na stwierdzenie występowania różnic, nie określa jednak konkretnych zmian. Aby odpowiedzieć na wszystkie pytania badawcze należało porównać średnią wartość odchylenia standardowego stopy zwrotu z poszczególnych okresów z całkowitą średnią odchylen standardowych stóp zwrotu wyznaczoną ze wszystkich obserwacji. Dzięki temu zabiegowi możliwe było ustalenie, w którym podokresie ryzyko Bitcoina średnio wzrasta lub maleje. W dwóch podokresach 8:00-11:55 oraz 4:00-7:55 ryzyko maleje w stosunku do średniej całkowitej. Tak niskie wartości są spowodowane tym, że w tych godzinach większość społeczeństwa śpi lub intensywnie pracuje i nie ma czasu ingerować w rynek Bitcoina. Ryzyko wzrasta od godziny 12:00 i pozostaje na poziomie powyżej średniej aż do godziny 3:55. Są to godziny mniej produktywne, kiedy społeczeństwo relaksuje się, odpoczywa, kończy pracę i ma czas, aby korzystać z Internetu i śledzić giełdę.

Podsumowując, udało się odpowiedzieć na pytania badawcze. Istnieją godziny, w których ryzyko Bitcoina regularnie wzrasta i maleje. Poziom ryzyka jest większy w godzinach 12:00-3:55 niż średnia w całym okresie, a więc w dość szerokim przedziale czasowym, kiedy ludzie częściowo są w pracy, a częściowo już odpoczywają. Statystycznie w porannych godzinach pracy ryzyko Bitcoina maleje, a od godziny 12:00 wzrasta.

Zdecydowanie warto pogłębić te badania i spróbować przeprowadzić tę samą analizę na krótszych przedziałach czasowych lub w inny sposób, niż powyżej podzielić dobę na odcinki. Warto także przeanalizować rynek kryptograficzny na przykładzie innych kryptowalut i stwierdzić, czy anomalie kalendarzowe występują tylko na przykładzie Bitcoina, czy jest to zjawisko globalne dla wszystkich kryptowalut.

Ciekawym zagadnieniem wydaje się to, czy za kilka lat udowodnione trendy pozostaną stabilne, czy też ulegną zmianie. Wszystkie te refleksje zachęcają naukowców i analityków z całego świata do stałego zgłębiania wiedzy dotyczącej kryptowalut oraz do przeprowadzania nowych badań i analiz na rynku kryptograficznym.

Literatura

- Bijańska, J. i Wodarski, K. (2014). Ryzyko w decyzjach inwestycyjnych przedsiębiorstw. *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie*, (70).
- Bińkowski, P. (2022, 23 kwietnia). *Ponad 10% Amerykanów będzie posiadać kryptowaluty do końca roku – sugeruje nowe badanie*. Bitcoin.pl. Pobrano 22 marca 2023 z <https://bitcoin.pl/badanie-adopcja>
- Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science*, 18(2), 168-174.
- Brueck, H. (2020, 22 września). *Naukowcy podpowiadają, jak najlepiej zaplanować swój dzień pracy*. Business Insider. Pobrano 10 czerwca 2024 z <https://businessinsider.com.pl/rozwoj-osobisty/kiedy-jestesmy-najbardziej-produktywni-w-pracy/ze83g9y>
- Budka, A., Kayzer, D., Pietruczuk, K. i Szoszkiewicz, K. (2013). Zastosowanie wybranych procedur do wykrywania obserwacji nietypowych w ocenie jakości rzek. *Infrastruktura i Ekologia Terenów Wiejskich*, 3(II).

- Business Insider (2021, 9 marca). *Czym jest rynek finansowy?* Pobrano 10 czerwca 2024 z <https://businessinsider.com.pl/gielda/co-to-jest-rynek-finansowy-i-co-nalezy-o-nim-wiedziec/3pl6nxq>
- Choithani, T., Chowdhury, A., Patel, S., Patel, P., Patel, D. i Shah, M. (2024). A Comprehensive Study of Artificial Intelligence and Cybersecurity on Bitcoin, Crypto Currency and Banking System. *Annals of Data Science*, 11(1), 103-135. <https://doi.org/10.1007/s40745-022-00433-5>
- Doman, R. (2004). Wykorzystanie danych wysokiej częstotliwości w prognozowaniu zmienności polskich indeksów giełdowych za pomocą modeli zmienności stochastycznej. *Acta Universitatis Lodzianis. Folia Oeconomica*, (177), 311-328.
- Dopierała, Ł., & Borodo, A. (2014). Znaczenie waluty kryptograficznej Bitcoin jako środka wymiany. *Współczesna Gospodarka*, 5(2).
- Fauzi, M. A., Paiman, N. i Othman, Z. (2020). Bitcoin i kryptowaluta: wyzwania, możliwości i przyszłe prace. *The Journal of Asian Finance, Economics and Business*, 7(8), 695-704.
- Foszcz, D. (2003). Zastosowanie metod bootstrapowych do bilansowania produkcji na przykładzie O/ZWR KGHM „Polska Miedź” SA. *Inżynieria Mineralna*, (3), 64-69.
- Godfrey L., i Orme C. (1991). Testing for Skewness of Regression Distributions. *Economic Letters*, (37), 31-34.
- Gulgowski, J. i Wolnik, B. (2013). *Metody bootstrapowe w statystyce*. Instytut Matematyki, Uniwersytet Gdański.
- Hesterberg, T. (2003). Bootstrap and Permutation Tests—Companion chapter 18 to the „Practice of Business Statistics”. *The Practice of Business Statistics*.
- Jajuga, K. (2006). Ogólne zasady inwestowania. *Anatomia Sukcesu. Instytucje i zasady funkcjonowania rynku kapitałowego*. Fundacja Edukacji Rynku Kapitałowego/ Komisja Nadzoru Finansowego.
- Kądziołka, K. (2015). Ocena ryzyka inwestycji w kryptowalutę Bitcoin. *Współczesna Gospodarka*, 6(3), 1-8.
- Kosiorowski, D., Rydlewski, J. P. i Zawadzki, Z. (2018). Wykrywanie funkcjonalnych obserwacji odstających na przykładzie monitorowania jakości powietrza. *Przeгляд Statystyczny*, 65(1), 82-98.
- More, S. T. B. (2004). Productive. *Environmental Justice Politics*, (163), 111-121.
- Pekasiewicz, D., Domański, C., Baszczyńska, A. i Witaszczyk, A. (2014). *Testy statystyczne w procesie podejmowania decyzji*. Wydawnictwo Uniwersytetu Łódzkiego.
- Piontek, K. (2007). Pomiar i testowanie skośności rozkładów stóp zwrotu instrumentów finansowych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu. Taksonomia*, (14), 122-130.
- Porcenałuk, P. (2013). Analiza wybranych miar ryzyka płynności dla akcji notowanych na GPW w Warszawie w latach 2001-2011. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (323), 289-297.
- Trzpiot, G. (2013). Wybrane statystyki odporne. *Studia Ekonomiczne*, (152), 163-173.
- Walker, M. L., Dovoedo, Y. H., Chakraborti, S., i Hilton, C. W. (2018). An Improved Boxplot for Univariate Data. *The American Statistician*, 72(4), 348-353.

Verification of the Occurrence of Calendar Anomalies in the Cryptocurrency Market Based on the Example of Bitcoin

Abstract: Huge technological progress and the transfer of payments to the Internet have opened up completely new possibilities in financial management and making payments. In 2008, there was a breakthrough when a new concept of electronic payments was created – Bitcoin, which quickly became the most popular cryptocurrency in the world. Currently, topics regarding Bitcoin and other cryptocurrencies are very topical, especially the question of whether Bitcoin can function as a means of payment and money. The financial market is closely related to the concept of risk, therefore Bitcoin was analysed in this context, defining risk as the standard deviation of Bitcoin rates of return. The bootstrap test conducted allowed us to determine that the risk of investing in Bitcoin differs in individual time subperiods. Relationships between social customs and Bitcoin hourly anomalies were identified, so it was concluded that this topic is still relevant and it is worth conducting further research in this area.

Keywords: Bitcoin, cryptocurrencies, calendar effects, investment risk, bootstrap test

Maja Ciechanowska

e-mail: 181372@student.ue.wroc.pl

ORCID: 0009-0006-0903-1315

Uniwersytet Ekonomiczny we Wrocławiu

Analiza poziomu zaufania społecznego i jego uwarunkowań

DOI: 10.15611/2024.53.6.03

JEL Classification: C38, O15

© 2024 Maja Ciechanowska

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Ciechanowska, M. (2024). Analiza poziomu zaufania społecznego i jego uwarunkowań. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 41-52). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Zaufanie społeczne jest istotną wartością mającą wpływ na funkcjonowanie każdego społeczeństwa. W ostatnich latach, w obliczu pandemii koronawirusa i wojny w Ukrainie, z powrotem zainteresowano się jego pozytywnymi skutkami. Celem pracy jest analiza poziomu zaufania społecznego i wybranych determinant tego zjawiska w krajach europejskich. W pracy zastosowano analizę korelacji i analizę skupień metodą k -średnich. Przeprowadzone analizy pozwoliły wyodrębnić czynniki mające największy wpływ na poziom zaufania. Są to, m.in. PKB per capita, stopień wykształcenia, korupcja i wielkość wydatków socjalnych w państwie. Wykazały one, że Europa nie jest jednorodna pod względem poziomu zaufania społecznego. Zauważono, że wpływ na nie mogą mieć uwarunkowania geograficzne i kulturowe.

Słowa kluczowe: zaufanie społeczne, analiza skupień, k -średnich, rozwój gospodarczy

1. Wstęp

Przedmiotem niniejszego artykułu jest zjawisko zaufania społecznego. W dzisiejszych czasach, w dobie Internetu i globalizacji, tradycyjne postrzeganie zaufania ulega zmianie. Mimo to jego rola we współczesnym świecie nie zmniejsza się. Zaufanie społeczne nadal stanowi podstawę funkcjonowania społeczności i jest uważane za niezbędny element systemów prawnych i struktur demokratycznych. Ponadto ma istotny wpływ na wiele czynników ekonomicznych.

Głównym zamierzeniem pracy jest zbadanie poziomu zaufania społecznego w wybranych krajach europejskich oraz podział państw na grupy podobne pod

względem poziomu zaufania społecznego i wybranych wskaźników rozwoju ekonomicznego. Określono następujące pytania badawcze:

- Czy istnieje związek pomiędzy poziomem zaufania a wybranymi czynnikami społeczno-ekonomicznymi?
- Czy państwa europejskie są jednolite pod względem poziomu zaufania społecznego?

Wykorzystanie metod wielowymiarowej analizy porównawczej pozwoli na sklasyfikowanie państw według odpowiednich grup oraz scharakteryzowanie ich pod względem poziomu zaufania i wybranych czynników rozwoju gospodarczego.

W artykule wykorzystano dane pochodzące z Europejskiego Sondażu Społecznego i Eurostatu. Zaufanie jest zjawiskiem o nieuchwytnym charakterze, dlatego trudno jest je precyzyjnie zmierzyć. Do jego oszacowania najczęściej wykonuje się badania ankietowe. Ma to jednak pewne ograniczenia – nie ma pewności, czy deklarowany poziom zaufania ma odzwierciedlenie w rzeczywistych sytuacjach życiowych (Stiglitz i in., 2019). Dlatego też powstają innowacyjne projekty mierzące zaufanie w alternatywny sposób, takie jak testy utajonych skojarzeń i gry eksperymentalne. Nie mają one jednak zastosowania w pomiarach prowadzonych na szerszą skalę i służą jedynie za pomoc i potwierdzenie wyników badań ankietowych.

2. Zaufanie społeczne w kontekście rozwoju gospodarczego i jakości życia

Zaufanie jest pojęciem szeroko analizowanym w ramach wielu dyscyplin naukowych. Według socjologa P. Sztompki (2007) zaufanie jest podstawowym sposobem radzenia sobie z niepewnością i brakiem możliwości, aby móc kontrolować przyszłość. Współczesny świat cechuje się dużym stopniem ryzyka, a działania innych są nieprzewidywalne. Stąd przy podejmowaniu decyzji niezbędna jest nadzieja, przekonanie bądź właśnie zaufanie, że sprawy potoczą się zgodnie z założeniami.

W literaturze wyróżnia się najczęściej trzy odmiany zaufania – interpersonalne, społeczne i instytucjonalne. Z punktu widzenia ekonomicznego najbardziej istotne jest zaufanie społeczne. Opiera się ono na codziennym doświadczaniu życia w społeczeństwie (Nowakowski, 2008). Jego adresatem są grupy podobnych do siebie jednostek pod względem konkretnej cechy. W analizach badacze najczęściej posługują się formą zaufania społecznego jako synonimu zaufania w ogóle. Jego poziom mierzy się zazwyczaj, pytając respondentów, czy zgadzają się ze stwierdzeniem, że „większości ludzi można ufać”.

Z zagadnieniem zaufania ściśle jest związane pojęcie kapitału społecznego będącego decydującym zasobem rozwoju społeczności i społeczeństwa, a zaufanie jest jednym z jego ważniejszych elementów. Growiec (2011) udowadnia, że kapitał społeczny ma niezależny od PKB per capita wpływ na poczucie szczęścia. Społeczeństwa szczęśliwsze to takie, w których zgeneralizowane zaufanie do innych ludzi jest powszechne (Growiec, 2011). Inne badanie (Mikucka i in., 2017) również dowodzi,

że w długim czasie zmiany zaufania społecznego dodatnio i istotnie korelują z satysfakcją z życia.

Z kolei negatywny wpływ na zaufanie mają nierówności w dochodach. Osłabiają one więzi międzyludzkie oraz poczucie solidarności i współpracy, co prowadzi do erozji kapitału społecznego. Kraje o najwyższych poziomach równości ekonomicznej (kraje skandynawskie, Holandia i Kanada) notują zarazem najwyższe poziomy zaufania społecznego (Stiglitz i in., 2019). Interesujący jest fakt, że rozwój gospodarczy wpływa na relację nierówności i zaufania. W krajach bogatszych negatywne oddziaływanie nierówności w dochodach na poziom zaufania jest większe. Natomiast w słabiej rozwiniętych krajach Europy Wschodniej wpływ ten nie jest obserwowany (Kanitsar, 2022).

J. Doyle (2023) wskazuje, że zaufanie społeczne wiąże się również ze stopniem ochrony środowiska. Zauważyć trzeba, że aby działania proekologiczne okazały się skuteczne, działania pojedynczych jednostek mogą być niewystarczające. Sukces działań proekologicznych zależy od współpracy tysięcy osób. W takich sytuacjach istotną rolę odgrywa zaufanie społeczne, które ma potencjał do łagodzenia obaw, na przykład przed nieodpowiedzialnymi zachowaniami. Zachęca także ludzi do współpracy. Zaufanie społeczne jest powiązane z gotowością jednostek do podejmowania osobistych poświęceń na rzecz ochrony środowiska (Doyle, 2023).

Spółeczeństwa z wysokim poziomem zaufania mają szczęśliwszych obywateli, silniejszą demokrację, bardziej wydajny rząd, mniej korupcji i wyższe wskaźniki wzrostu gospodarczego. Chociaż niektóre z tych konsekwencji są prawdopodobnie wyolbrzymione lub mogą być jedynie przyczyną zaufania społecznego, a nie jego skutkiem, to nie ulega wątpliwości, że zaufanie społeczne jest ważnym zjawiskiem (Serritzlew i in., 2014). Jak pisze P. Sztompka (2007), nie jest przypadkiem, że ranking krajów, od wysokiego do niskiego zaufania, „jest uderzająco zbieżny z rankingami wzrostu gospodarczego, dochodu narodowego per capita, zdrowotności, średniej długości życia, stopnia skolaryzacji i swobód politycznych”.

Sprawdzianem dla poziomu zaufania społecznego okazała się pandemia koronawirusa COVID-19. Obnażyła w wielu krajach słabości obowiązujących tam modeli społecznych. Kryzys zdrowotny dotknął o wiele silniej społeczeństwa, w których panuje niski poziom zaufania społecznego i duża nieufność wobec instytucji publicznych. Nieufność wobec obcych sprzyjała dystansowaniu społecznemu, ale jednocześnie brak zaufania wobec decydentów zmniejszył stosowanie się do zaleceń ekspertów zdrowia publicznego (Marczewski, 2020).

3. Metodyka badań

Na podstawie analizowanych raportów i literatury przedmiotu wyodrębniono zbiór 12 zmiennych, który posłuży do wykonania obliczeń. Dane pochodzą z 2020 roku i prezentują się następująco:

- *gdp* – PKB *per capita*, wyrażone w euro;
- *gini* – współczynnik Giniego, rozumiany jako miara nierówności rozkładu dochodów, mierzona na skali od 0 do 100;
- *geg* – luka w zatrudnieniu kobiet i mężczyzn (ang. *gender employment gap*), definiowana jako różnica między wskaźnikami zatrudnienia mężczyzn i kobiet w wieku 20-64 lata;
- *edu* – odsetek osób posiadających wykształcenie wyższe;
- *crime* – odsetek osób dotkniętych przez przestępczość, przemoc lub wandalizm;
- *protection* – całkowite wydatki na ochronę socjalną wyrażone w euro na mieszkańca;
- *corruption* – Indeks Percepcji Korupcji (ang. *Corruption Perceptions Index*, CPI) – jest to globalny wskaźnik korupcji w sektorze publicznym, oparty na 13 ocenach ekspertów i ankietach przeprowadzanych wśród przedsiębiorców, gdzie wynik 0 oznacza bardzo wysoki poziom korupcji, a wynik 100 oznacza bardzo „czysty” (nieskorumpowany) kraj;
- *religion* – stopień religijności mierzony na skali od 0 do 10, gdzie 0 oznacza najniższy poziom cechy, a 10 najwyższy (podobnie jak dla kolejnych 4 zmiennych);
- *trust* – stopień zaufania społecznego, mierzony na skali od 0 do 10; wyrażony stwierdzeniem „większości ludzi można ufać i nie trzeba być zbyt ostrożnym”;
- *polit* – stopień zaufania do polityków, mierzony na skali od 0 do 10;
- *party* – stopień zaufania do partii politycznych, mierzony na skali od 0 do 10;
- *parl* – stopień zaufania do parlamentu, mierzony na skali od 0 do 10.

W analizie uwzględniono 25 krajów europejskich spośród 39 ze względu na brak danych dla pozostałych. Pierwszych 7 zmiennych zostało pobranych z Eurostatu, z kolei zmienne od 8 do 12 zostały przygotowane na podstawie danych z Europejskiego Sondażu Społecznego (*European Social Survey*), który co dwa lata przeprowadza międzynarodowe badanie postaw i zachowań. Jego celem jest obserwacja zmian społecznych zachodzących w Europie. Próby, na których jest przeprowadzane badanie w każdym kraju, są reprezentatywne dla osób w wieku 15 lat i starszych. Analizę rozpoczęto od sprawdzenia, czy w zbiorze występują obserwacje odstające. W tym celu posłużono się wykresami pudełkowymi. Bazują one na jednowymiarowym kryterium kwartylowym, które zakłada obserwację za odstającą, jeśli znajduje się ona poza przedziałem:

$$\left\langle Q_1 - \frac{3(Q_3 - Q_1)}{2}, Q_3 + \frac{3(Q_3 - Q_1)}{2} \right\rangle, \quad (1)$$

gdzie: Q_1 – pierwszy kwartył, Q_3 – trzeci kwartył.

Jeżeli odległość obserwacji od bliższego kwartyła przekracza wartość 1,5 wielkości odstepu międzykwartylowego, to uznaje się ją za odstającą, natomiast jeśli

odległość zostanie przekroczona trzykrotnie, określa się ją mianem ekstremalnej (Aczel i Sounderpandian, 2018).

W celu podziału państw na podobne grupy pod względem badanych zmiennych zastosowano analizę skupień metodą k -średnich. Jej ideą jest podział obiektów z jednego zbioru na jednorodne grupy. Elementy w grupie powinny być do siebie jak najbardziej podobne, natomiast pomiędzy grupami podobieństwo powinno być jak najmniejsze. Na początku analizy skupienia nie są znane, tak jak ich skład, liczba i charakterystyki. Pozostaje to określane dopiero w trakcie analizy (Balicki, 2013).

Do wykonania analizy skupień konieczna jest normalizacja zmiennych. Ujednoczeniu podlegają rzędy wielkości i jednostki miary zmiennych. W pracy zmienne zostały poddane standaryzacji określonej wzorem: (Stanimir, 2006):

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (2)$$

gdzie: x_{ij} jest to wartość j -tej zmiennej dla i -tej obserwacji, \bar{x}_j to średnia arytmetyczna, zaś s_j odchylenie standardowe.

Metoda k -średnich opiera się na algorytmie, który klasyfikuje obiekty do odpowiednich klas. Pierwszym krokiem jest ustalenie liczby klas, tzn. k . Można ją określić na podstawie wykresu osypiska lub indeksu silhouette. Algorytm ma charakter iteracyjny. W każdej iteracji oblicza się środki ciężkości dla klas zgodnie ze wzorem:

$$v_{kj} = \frac{1}{n_k} \sum_{i \in C_k} z_{ij}, \quad (3)$$

gdzie: n_k jest to liczba obiektów w k -tej klasie, a z_{ij} to standaryzowana wartość zmiennej X_j dla i -tego obiektu.

Następnie wyznacza się odległości obiektów od wektora średnich każdej klasy. Ustalana jest nowa klasyfikacja, każdy obiekt przydzielany jest do klasy, dla której odnotowano najmniejszą odległość od wektora średnich. Algorytm jest powtarzany aż do momentu, w którym otrzymane zostają takie same klasyfikacje w dwóch kolejnych iteracjach (Stanimir, 2006).

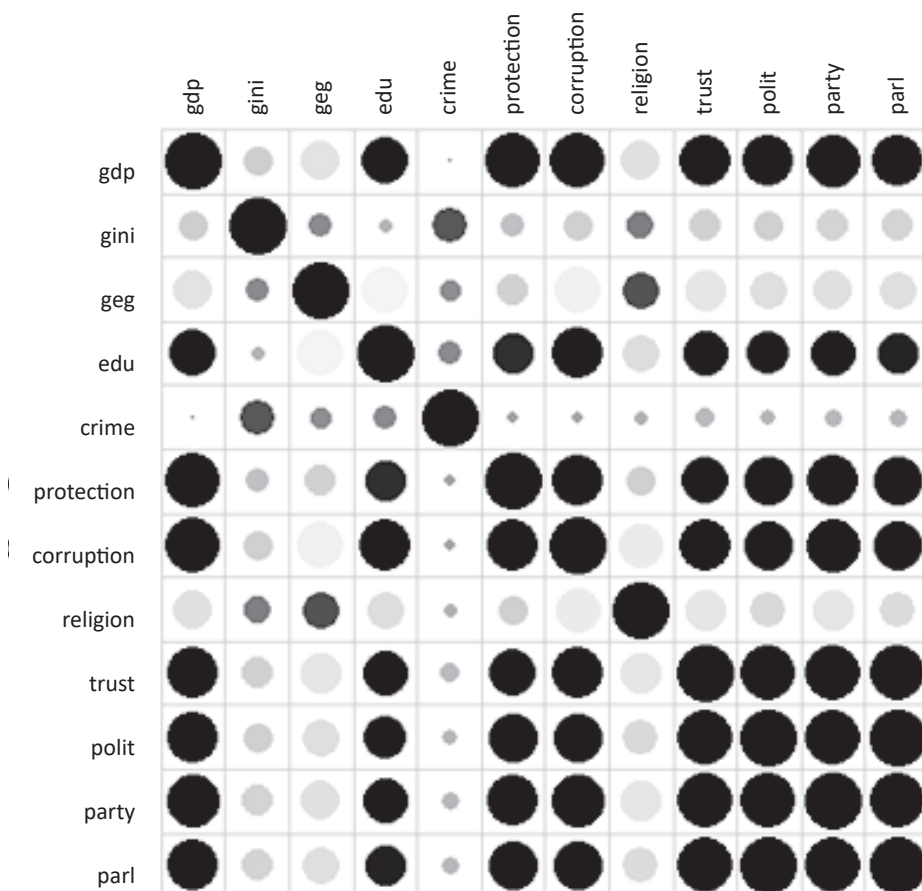
4. Wyniki badań

W pierwszym kroku zbadano występowanie obserwacji odstających w zbiorze danych. Odnotowano 5 obserwacji odstających. Dla zmiennej *crime* są to Bułgaria, Grecja, Francja i Holandia. Dla zmiennej *gini* wartość odstająca występuje dla Bułgarii. Pozostawiono do dalszej analizy wszystkie kraje, aby w jak najbardziej pełny sposób zrealizować cel niniejszego badania, który dotyczył przedstawienia sytuacji społeczno-ekonomicznej Europy. Ze względu na braki danych wielkość próby została znacznie okrojona na początkowym etapie analizy, a dalsze eliminowanie krajów

miałoby negatywny wpływ na wartość informacyjną badania. Należy podkreślić, że wartości odstające występowały jedynie dla dwóch zmiennych.

Następnie zbadano współczynniki zmienności dla analizowanych cech. Większość zmiennych charakteryzuje się średnim zróżnicowaniem. Bardzo duże zróżnicowanie (powyżej 60%) można zaobserwować dla zmiennych wyrażonych w euro, czyli dla PKB (zmienna *gdp*) i wydatków socjalnych (zmienna *protection*).

Pomiędzy zmiennymi odnotowane zostało występowanie wielu zależności. Na rysunku 1 przedstawiono wizualizację macierzy korelacji. Im punkt jest większy, tym korelacja jest silniejsza. Dodatnie korelacje zostały zaznaczone ciemnym kolorem. Ponadto sprawdzono istotność współczynników korelacji na poziomie 5%.



Rys. 1. Macierz korelacji pomiędzy zmiennymi

Źródło: opracowanie własne.

W szczególności warto się przyjrzeć zmiennym opisującym różne rodzaje zaufania. Zgodnie z przypuszczeniami zmienne te są silnie skorelowane – współczynniki korelacji wynoszą od 0,88 aż do 0,99 w przypadku zaufania do polityków i parlamentu. Wysokie dodatnie korelacje występują także pomiędzy zaufaniem społecznym (zmienna *trust*) a wskaźnikiem korupcji (a właściwie braku korupcji – im wyższa wartość, tym kraj jest mniej obciążony korupcją) oraz PKB *per capita*. Ponadto istotne dodatnie korelacje występują pomiędzy zaufaniem a wydatkami na opiekę socjalną i poziomem wykształcenia, natomiast ujemne zależności odnotowano między zaufaniem a luką w zatrudnieniu i religijnością. Interesujące jest, że poziom zaufania nie jest istotnie zależny od przestępczości w kraju i nierówności dochodowych.

Na rysunku 2 przedstawiono kształtowanie się poziomu zaufania w Europie. Mapę wykonano dla zaufania społecznego (zmienna *trust*), jednak można ją także odnieść do jego pozostałych rodzajów. Mapy w znacznej mierze pokrywałyby się ze względu na wcześniej wykryte silne korelacje pomiędzy rodzajami zaufania.



Rys. 2. Poziom zaufania w wybranych krajach Europy w 2020 roku

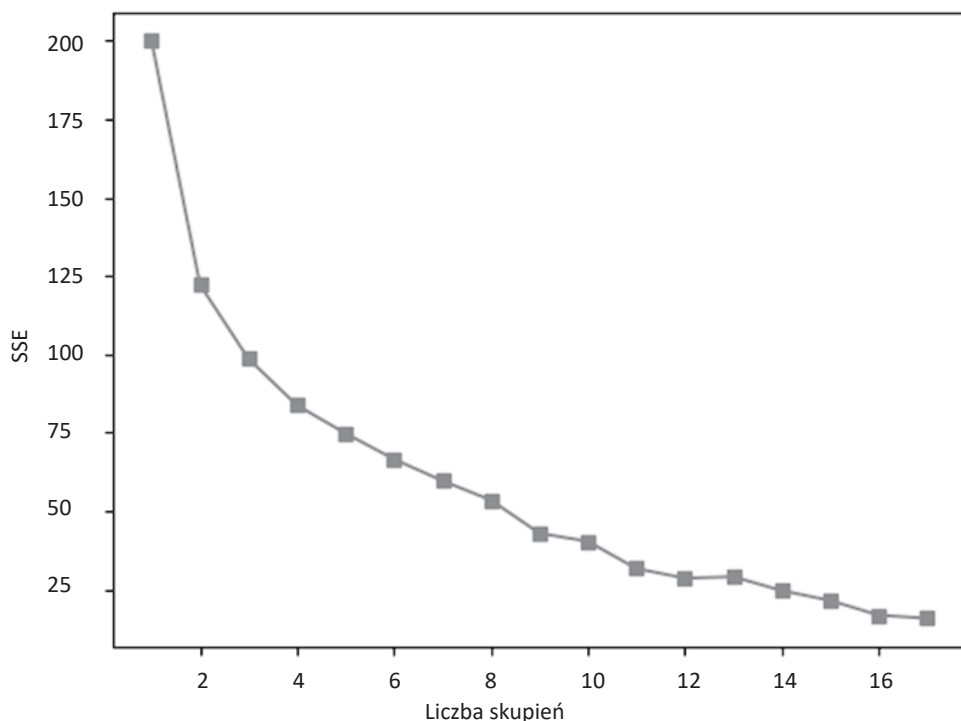
Źródło: opracowanie własne na podstawie GeoNames.

Najwyższym poziomem zaufania cechują się Holandia, Szwajcaria i niektóre kraje nordyckie – Norwegia, Finlandia i Islandia. Na zachodzie i południu Europy zaufanie przyjmuje przeciętne poziomy. Najniższe obserwuje się w Europie Środkowej, w tym w Polsce, co jest związane przede wszystkim z czynnikami historycznymi, takimi jak lata komunistycznych rządów. Różnica pomiędzy państwem o najwyższym

a najniższym poziomie zaufania wynosi 4,01 (zmienna jest mierzona na skali od 0 do 10).

Przed przystąpieniem do analizy skupień postanowiono wykluczyć zmienne, które były ze sobą bardzo silnie skorelowane. Za maksymalną wartość współczynnika korelacji przyjęto 0,8. Ze zbioru wykluczono zmienne *gdp*, *polit*, *party* i *parl*.

W celu podziału państw na grupy zastosowano metodę *k*-średnich. Analizę rozpoczęto od wyboru liczby skupień, które zapewnią optymalny podział obserwacji. Na rysunku 3 przedstawiono wykres osypiska. Podziału należy dokonać w miejscu, gdzie sumy wariancji wewnątrzgrupowej, czyli SSE, przestają gwałtownie maleć. Wykres jednak nie daje jednoznacznej informacji, w którym miejscu należałoby dokonać podziału. Rozważano wybór 3 lub 4 skupień. Aby sprawdzić słuszność wyboru, obliczono i porównano wartości indeksu silhouette. Wyższe wartości indeksu, a więc lepszą jakość klasyfikacji, odnotowano dla 3 skupień, zatem do dalszej analizy przyjęto $k = 3$.



Rys. 3. Wykres osypiska

Źródło: opracowanie własne.

W tabeli 1 przedstawiono przynależność krajów do utworzonych skupień, natomiast w tabeli 2 średnie poziomy badanych zmiennych dla poszczególnych skupień.

Tabela 1. Przynależność państw do skupień

| Numer skupienia | Państwa przynależące do danego skupienia |
|-----------------|---|
| 1 | Czechy, Chorwacja, Węgry, Polska, Portugalia, Słowenia, Słowacja |
| 2 | Niemcy, Estonia, Hiszpania, Francja, Litwa, Holandia, Austria, Finlandia, Szwecja, Islandia, Norwegia, Szwajcaria |
| 3 | Bułgaria, Grecja, Włochy, Czarnogóra, Macedonia Północna, Serbia |

Źródło: opracowanie własne.

Tabela 2. Średnie poziomy zmiennych dla poszczególnych skupień

| Skupienie | gini | geg | crime | protection | religion | trust | edu | corruption |
|-----------|-------|-------|-------|------------|----------|-------|-------|------------|
| 1 | 26,19 | 10,23 | 5,87 | 3978,31 | 4,76 | 4,30 | 25,34 | 53,00 |
| 2 | 29,38 | 6,15 | 9,28 | 12 965,78 | 4,02 | 5,66 | 36,11 | 76,50 |
| 3 | 33,58 | 15,80 | 11,61 | 5929,97 | 5,70 | 3,94 | 24,16 | 44,17 |

Źródło: opracowanie własne.

Do 1 skupienia należy 7 krajów, w tym Polska. Cechują się one średnim poziomem zaufania społecznego, najniższymi wydatkami na cele socjalne, stosunkowo niskim poziomem wykształcenia i średnimi poziomami religijności, i korupcji. Są również najbezpieczniejszymi obszarami (najniższa wartość zmiennej *crime*) o najmniejszych nierównościach w dochodach (najniższa wartość zmiennej *gini*).

**Rys. 4.** Przynależność państw do skupień

Źródło: opracowanie własne na podstawie GeoNames.

Druga grupa składa się z 12 państw o zdecydowanie najwyższych poziomach zaufania. Cechują się one także najwyższym poziomem wykształcenia, największymi wydatkami na cele socjalne, najniższym poziomem religijności i najmniejszą luką

w zatrudnieniu. Jednocześnie charakteryzują się stosunkowo wysokim poziomem przestępczości przy niskiej korupcji. Do tej grupy przynależą kraje nordyckie i większość krajów Europy Zachodniej, więc państwa o największym poziomie rozwoju gospodarczego.

Ostatnia grupa 3 składa się z 6 krajów Europy Południowej i dawnej Jugosławii. Obywatele zamieszkujący te tereny najmniej sobie ufają. Są jednocześnie najbardziej religijni i słabo wykształceni. Państwa zmagają się z największym poziomem korupcji, przestępczości oraz z nierównościami w zatrudnieniu w kobiet.

Na podstawie mapy przedstawionej na rysunku 4 można zauważyć, że utworzone skupienia cechują się podobnym położeniem geograficznym. Europa została podzielona na trzy części: kraje zachodnioeuropejskie, kraje środkowo- i wschodnioeuropejskie wraz z Portugalią oraz kraje południowoeuropejskie.

5. Zakończenie

Celem niniejszego artykułu było zbadanie zjawiska zaufania społecznego w wybranych krajach europejskich oraz wyodrębnienie potencjalnych czynników wpływających na jego poziom.

Zaobserwowano istnienie prawidłowości pomiędzy zaufaniem społecznym a rozwojem gospodarczym kraju. Odnotowano silną dodatnią korelację pomiędzy PKB i stopniem zaufania. Zauważono, że zaufanie wiąże się pozytywnie z poziomem wykształcenia, a negatywnie z luką w zatrudnieniu i poziomem korupcji w kraju. Ponadto zaufanie społeczne jest bardzo silnie dodatnio skorelowane z zaufaniem do polityków, partii politycznych i parlamentu.

Zastosowanie metody *k*-średnich umożliwiło podział Europy na trzy grupy. Podział ten jest spójny geograficznie. Najwyższym zaufaniem charakteryzują się państwa Europy Zachodniej i Północnej. Cechują się one największym poziomem rozwoju gospodarczego – największymi wydatkami na cele socjalne, najwyższym poziomem wykształcenia, najmniejszą korupcją i równym traktowaniem kobiet.

Do drugiej grupy zaliczono kraje wschodnio- i środkowoeuropejskie (wraz z Portugalią). Panuje tam stosunkowo niski poziom zaufania przy jednocześnie wysokim bezpieczeństwie i małych nierównościach w dochodach.

Najmniejszym zaufaniem obdarzają się mieszkańcy krajów bałkańskich i południowoeuropejskich. Region ten charakteryzuje się najmniejszym dobrobytem, wysoką korupcją i przestępczością.

Mimo wyróżnienia potencjalnych czynników, które wyjaśniają istnienie zaufania, nie ma podstaw do określenia ciągu przyczynowo-skutkowego. Prawdopodobnie istnieje zależność obustronna. Warto również podkreślić, że na zaufanie mają wpływ czynniki bezpośrednio niemierzalne, takie jak przekazy kulturowe, tradycja i różnorodność etniczna. Poziom zaufania jest zazwyczaj wyższy w dojrzałych systemach demokratycznych. Państwa postkomunistyczne, jak Polska lub dotknię-

te kryzysami i konfliktami zbrojnymi w końcu XX wieku (przykładem są państwa bałkańskie czy Grecja) cechują się niższym PKB i zaufaniem społecznym. Budowanie i umacnianie zaufania w społeczeństwie jest długotrwałym procesem. Wpływ na niego mają wydarzenia polityczne i historyczne, takie jak pandemia COVID-19, a także wojna w Ukrainie, której konsekwencje dotyczące zaufania nie są jeszcze znane. Jest to zatem czynnik podatny na zmiany, zależny od sytuacji ogólnoswiatowej.

Temat zaufania mógłby zostać rozwinięty na wielu jeszcze innych płaszczyznach i stanowić przedmiot dalszych badań. Z uwagi na charakter i ramy tego artykułu jego celem było w sposób możliwie spójny przybliżyć kwestię zaufania społecznego jako istotnego czynnika wpływającego na kształtowanie sytuacji ekonomicznej państwa.

Literatura

- Aczel, A. D. i Sounderpandian, J. (2018). *Statystyka w zarządzaniu*. Wydawnictwo Naukowe PWN.
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Doyle, J. (2023). Social Trust, Cultural Trust, and the Will to Sacrifice for Environmental Protections. *Social Science Research*, (109). <https://doi.org/10.1016/j.ssresearch.2022.102779>
- Growiec, K. (2011). Kapitał społeczny: geneza i społeczne konsekwencje. *Krakowskie Studia Małopolskie*, (16). <http://dx.doi.org/10.15804/ksm201154>
- Kanitsar, G. (2022). The Inequality-Trust Nexus Revisited: At What Level of Aggregation Does Income Inequality Matter for Social Trust? *Social Indicators Research*, 163(1), 171-195. <https://doi.org/10.1007/s11205-022-02894-w>
- Marczewski, P. (2020). *Epidemia nieufności. Zaufanie społeczne w czasie kryzysu zdrowotnego*. Forum Idei Fundacji im. Stefana Batorego.
- Mikucka, M., Sarracino, F., i Dubrow, J. K. (2017). When Does Economic Growth Improve Life Satisfaction? Multilevel Analysis of the Roles of Social Trust and Income Inequality in 46 Countries, 1981-2012. *World Development*, 93(C), 447-459. <https://doi.org/10.1016/j.worlddev.2017.01.002>.
- Nowakowski, K. (2008). Wymiary zaufania i problem zaufania negatywnego w Polsce. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, LXX(1), 213-233.
- Serritzlew, S., Sønderskov, K. M. i Svendsen, G. T. (2014). Do Corruption and Social Trust Affect Economic Growth? A Review. *Journal of Comparative Policy Analysis: Research and Practice*, 16(2), 121-139. <https://doi.org/10.1080/13876988.2012.741442>
- Stiglitz, J. E., Fitoussi, J.-P., i Durand, M. (2019). Poza PKB. Mierzmy to, co ma znaczenie dla rozwoju społeczno-gospodarczego. W: XXXII Olimpiada wiedzy ekonomicznej. Gospodarka światowa. Wolny handel czy protekcjonizm? *Biuletyn PTE*, 3(86), 56.
- Stanimir, A. (red.). (2006). *Analiza danych marketingowych: problemy, metody, przykłady*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
- Sztompka, P. (2007). *Zaufanie*. Wydawnictwo Znak.

Determinants of Social Trust in European Countries

Abstract: Social trust is an important element of every society which contributes economic development and well-being. The aim of the article was to analyse the level of social trust and selected determinants of this phenomenon in European countries. In the thesis *k*-means clustering was used to group countries. The number of clusters was chosen using the elbow method and the Silhouette score. The variables which have a significant impact on the level of trust were distinguished. The conducted research proved that determinants like GDP per capita, education, corruption and social protection have significant impact on social trust. It was found that Europe is divided into three different groups. It has been noticed that geographical location and cultural heritage affect the level of trust.

Keywords: social trust, cluster analysis, *k*-means, human capital, socioeconomic indicators

Julia Dymek

e-mail: 181441@student.ue.wroc.pl

ORCID: 0009-0008-0600-8552

Weronika Kolasa

e-mail: 181728@student.ue.wroc.pl

ORCID: 0009-0002-1034-0028

Uniwersytet Ekonomiczny we Wrocławiu

Analiza stylu życia obywateli Unii Europejskiej w kontekście lat przeżytych w zdrowiu

DOI: 10.15611/2024.53.6.04

JEL Classification: I12, I14

© 2024 Julia Dymek, Weronika Kolasa

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Dymek, J. i Kolasa, W. (2024). Analiza stylu życia obywateli Unii Europejskiej w kontekście lat przeżytych w zdrowiu. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 53-67). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Przedmiotem artykułu jest analiza stylu życia obywateli krajów Unii Europejskiej oraz czynników wpływających na lata przeżyte w zdrowiu. Celem analizy jest wyodrębnienie tych czynników i porównanie państw europejskich pod względem ich występowania. Analizę danych przeprowadzono za pomocą statystyk opisowych i modelu ekonometrycznego, aby zbadać wpływ zmiennych objaśniających na zmienną objaśnianą, którą są lata przeżyte w zdrowiu. Przeprowadzono także analizę skupień z użyciem zmiennych dotyczących stylu życia. Wykazano, iż Polska jest podobna pod względem wybranych czynników do innych krajów położonych w Europie Środkowo-Wschodniej. Ponadto ich konfiguracja jest zbliżona do siebie dla krajów Europy Zachodniej i porównywalna z krajami skandynawskimi.

Słowa kluczowe: lata przeżyte w zdrowiu, styl życia, współczesny Europejczyk, model ekonometryczny, analiza skupień

1. Wstęp

Z roku na rok wzrasta odsetek osób starszych w populacji krajów europejskich. Nie ulega wątpliwości, że jedną z przyczyn jest rozwój technologii i medycyny przekładający się na dłuższe życie. Warto natomiast przyjrzeć się temu, jak wyglądają lata życia w zdrowiu, czyli *healthy life years* (HLY). Na długość życia oraz lata przeżyte

w zdrowiu z pewnością mają wpływ, między innymi: warunki środowiskowe, styl życia, dieta, poziom stresu oraz czynniki genetyczne (Achremowicz, 2012). Warto zauważyć, że życie w dobrym zdrowiu jest zagadnieniem złożonym, gdyż zależne jest od wielu obszarów życia człowieka. Przede wszystkim charakterystyka współczesnego Europejczyka, a w szczególności jego stylu życia, jest powiązana z latami przeżyтыми w zdrowiu. Przedmiotem artykułu jest analiza stylu życia obywateli krajów Unii Europejskiej [UE] oraz czynników wpływających na lata przeżyte w zdrowiu. Celem analizy jest zidentyfikowanie tych czynników i porównanie państw europejskich pod względem ich występowania.

Do przeprowadzenia analizy zastosowano statystyki opisowe, które umożliwiły wskazanie podstawowych informacji, takich jak mediana, średnia, wartości maksymalne i odchylenie standardowe. Następnie utworzono model ekonometryczny, umożliwiający zbadanie wpływu zmiennych objaśniających na zmienną objaśnianą, którą są lata przeżyte w zdrowiu. Wykorzystano również klasyfikację hierarchiczną z grupy wielowymiarowej analizy porównawczej. Niniejsza analiza została przygotowana na podstawie danych pobranych z internetowej bazy danych Urzędu Statystycznego UE – Eurostatu (<https://ec.europa.eu/eurostat/web/main/data/database>) oraz The Organisation for Economic Co-operation and Development (<https://www.oecd.org/>). Baza zawiera wyniki badań statystycznych przeprowadzonych na terenie UE i strefy euro z wykorzystaniem danych zgromadzonych przez krajowe organy statystyczne. Do analizy zostały pobrane dane przekrojowe, które są rzetelne i wiarygodne. Pytania badawcze w niniejszej pracy brzmią następująco:

- Czy zmienna objaśniana – lata życia w zdrowiu, będzie opisywana przez czynniki dotyczące stylu życia w modelu ekonometrycznym?
- Czy Polska jest podobna pod względem czynników opisujących styl życia obywateli do innych krajów położonych w Europie Środkowo-Wschodniej?
- Czy konfiguracja czynników dotyczących stylu życia dla krajów Europy Zachodniej jest zbliżona do siebie i dodatkowo porównywalna z sytuacją krajów skandynawskich?

2. Lata życia w zdrowiu oraz styl życia – definicje i ich aspekty

Oczekiwane lata życia w zdrowiu, zwane również oczekiwaną długością życia bez niepełnosprawności – *Disability Free Life Expectancy* (DFLE), definiuje się jako liczbę lat, przez którą dana osoba ma nadal żyć w dobrym stanie. Ten wskaźnik statystyczny jest zestawiany osobno dla mężczyzn i kobiet, w chwili urodzenia oraz w wieku 50 i 65 lat (Ec.europa.eu, 2015).

M. Sauerberg (2021) analizował, czy struktura wykształcenia społeczności ma wpływ na lata życia w zdrowiu. Ponadto twierdził on, że zarówno *life expectancy*, czyli oczekiwana długość życia, jak i *healthy life years* (HLY), czyli lata życia w zdrowiu, stanowią podstawowe miary oceny sytuacji zdrowotnej społeczeństw. W odróżnieniu od wskaźnika oczekiwanej długości życia wskaźnik lat życia w zdrowiu

zawiera informacje zarówno o śmiertelności, jak i te dotyczące zdrowia (Mathers, 2002, za: Sauerberg, 2021). Warto jednak wspomnieć, iż badany wskaźnik może stanowić problem w momencie porównywania go między krajami. Wpływ na to mają przede wszystkim różnice wynikające ze sposobów jego szacowania. Mianowicie używane są rozmaite dane dotyczące zdrowia. W dodatku samo słowo „zdrowie” posiada wiele definicji (Sauerberg, 2021).

Długie pozostawanie w zdrowiu ma wpływ na wzrost ekonomiczny. Przekłada się przede wszystkim na większą produktywność w pracy. Wynika to z faktu, iż mniej osób korzysta ze zwolnień zdrowotnych, czy odchodzi na wcześniejszą emeryturę (Sojka, 2020).

W 2018 roku P. Bogaert i in. opublikowali raport dotyczący wskaźnika opisującego lata życia w zdrowiu. Wynika z niego, że głównym celem UE jest zmniejszenie obciążeń społecznych i ekonomicznych, a także wydłużenie średniej długości życia. Co więcej, autorzy twierdzą, iż wskaźnik HLY służy nie tylko do opisanego lat życia w zdrowiu, ale także do oceny zrównoważonego rozwoju. Co ciekawe, analiza wskaźnika HLY doprowadziła do zmiany polityki w Estonii, głównie ze względu na słabą ocenę Narodowego Planu Zdrowia. Z kolei na Litwie wskaźnik lat przeżytych w zdrowiu służy jako główne kryterium oceny programu rządu (Bogaert i in., 2018).

Na stan zdrowia człowieka z pewnością wpływa wiele czynników. M. Lalonde, pełniący funkcję ministra zdrowia Kanady w 1974 roku, przedstawił czynniki mające wpływ na stan zdrowia ludzi. Determinanty wpływające na zdrowie to styl życia, środowisko fizyczne, czynniki genetyczne i opieka zdrowotna (Lalonde, 1974).

3. Metodyka badań

3.1. Charakterystyka wykorzystanych zmiennych

Analizowane dane pochodzą z 2019 roku. Badanie przeprowadzono dla dwudziestu pięciu obserwacji, którymi są państwa należące do UE. Z powodu braku danych należało wykluczyć kilka państw. Do analizy wybrano osiem zmiennych mających wpływ na zdrowie człowieka. Zmienne te dotyczą stylu życia i są to:

1. Zmienna X_1 – narażenie na zanieczyszczenie powietrza pyłem zawieszonym PM_{2,5} ($\mu\text{g}/\text{m}^3$), dotyczy stężenia drobnego pyłu zawieszonego (PM_{2,5}) w powietrzu i opisuje zanieczyszczenie powietrza, które stanowi największe zagrożenie dla zdrowia na całym świecie, dotykając więcej ludzi niż jakiegokolwiek inne zanieczyszczenie. Przewlekła ekspozycja na pył PM_{2,5} znacznie zwiększa ryzyko wystąpienia chorób układu oddechowego i sercowo-naczyniowego. Dane odnoszą się do narażenia populacji na więcej niż $10 \mu\text{g}/\text{m}^3$ i są wyrażone jako średnie roczne.
2. Zmienna X_2 , czyli wskaźnik lat przeżytych w zdrowiu (lata), w chwili urodzenia (ang. HLY) określa oczekiwaną długość życia w zdrowiu oraz bez niepełnosprawności. Wskaźnik HLY dostarcza informacji na temat względnego stanu zdrowia populacji w krajach UE.

3. Zmienna X_3 – opisuje odsetek populacji poszczególnych krajów UE, który spożywa od 1 do 4 porcji warzyw i owoców dziennie (%).
4. Zmienna X_4 określa odsetek osób otyłych na podstawie wskaźnika BMI (%) – *Body Mass Index*. Wskaźnik BMI definiuje się jako iloraz masy ciała wyrażonej w kilogramach i kwadratu wzrostu wyrażonego w metrach (kg/m^2). Prawidłowy stosunek wagi do wzrostu wynosi od 18,5 do 25. Natomiast wskaźnik przyjmujący wartości równe lub większe 30 wskazuje na otyłość.
5. Zmienna X_5 , określa, jaki odsetek społeczeństwa danego państwa deklaruje podejmowanie aktywności fizycznej przynajmniej raz w tygodniu (%). Do aktywności fizycznej zalicza się ćwiczenia aerobowe i trening wzmacniający mięśnie.
6. Zmienna X_6 określa występowanie symptomów depresyjnych (%). Dane wskazują, jaki odsetek społeczeństwa danego państwa cierpi na zaburzenia nastroju. Stan zdrowia został oceniony subiektywnie przez respondentów.
7. Zmienna X_7 opisuje wydatki na zdrowie (% PKB), które mierzą ostateczną konsumpcję towarów i usług opieki zdrowotnej. Do bieżących wydatków na zdrowie zalicza się, między innymi indywidualną opiekę zdrowotną i usługi zbiorowe. Opieka zdrowotna jest finansowana przede wszystkim za pomocą wydatków rządowych i obowiązkowych ubezpieczeń zdrowotnych. Wskaźnik jest mierzony jako udział w PKB.
8. Zmienną X_8 , czyli konsumpcję alkoholu (litry/osobę), definiuje się jako roczną sprzedaż czystego alkoholu w litrach na osobę w wieku 15 lat i starszą. Spożywanie alkoholu może doprowadzić do wielu szkodliwych konsekwencji zarówno zdrowotnych i społecznych. Może zwiększać ryzyko szeregu nowotworów, udaru mózgu i marskości wątroby. Spożywanie alkoholu przyczynia się również do śmierci i niepełnosprawności poprzez wypadki i urazy, napaści, przemoc, zabójstwa i samobójstwa.

3.2. Metody analityczne zastosowane w badaniu

Wpływ wybranych zmiennych na lata przeżyte w zdrowiu zbadano, stosując model ekonometryczny. W celu stworzenia takiego modelu utworzono zbiór potencjalnych zmiennych objaśniających, wybrano analityczną postać zależności, przeprowadzono estymację modelu weryfikację oraz przedstawiono wnioski (Dziechciarz, 2012). Estymacja modelu polega na wyznaczeniu przybliżonych wartości parametrów strukturalnych. Model po oszacowaniu przyjmuje postać:

$$\hat{y} = a_0 + a_1x_1 + \dots + a_jx_j + \dots + a_mx_m. \quad (1)$$

Następną czynnością po oszacowaniu parametrów modelu jest jego weryfikacja, czyli sprawdzenie poprawności. Można to zrobić za pomocą testów, między innymi: testu *t*-Studenta i testu Fishera, które pozwalają na analizę istotności parametrów strukturalnych, testu Breuscha-Pagana, badającego homoskedastyczność,

testu Shapiro-Wilka, badającego normalność rozkładu reszt, oraz testów dotyczących innych własności składnika resztowego.

Kolejną metodą zastosowaną w badaniu jest analiza skupień (Dziechciarz, 2012). W celu wyodrębnienia państw podobnych do siebie pod względem stylu życia współczesnego Europejczyka zastosowano klasyfikację hierarchiczną. Procedury hierarchiczne zalicza się do metod klasyfikacji. W rezultacie metody klasyfikacji mają doprowadzić do tego, że badane obiekty będą podzielone na skupienia. Skupienie zawiera obiekty, które ze względu na badane zjawisko są homogeniczne, czyli są w małej odległości od siebie. Natomiast obiekty, które należą do różnych klas, dzieli duża odległość, co oznacza, iż nie są one do siebie podobne. Wówczas mowa jest o zjawisku heterogeniczności. Pierwszym etapem klasyfikacji hierarchicznej jest standaryzacja danych oraz utworzenie macierzy odległości. Do najczęściej wykorzystywanych funkcji odległości zalicza się: Czebyszewa, euklidesową, miejską, potęgową i euklidesową do kwadratu (Statsoft.pl, 2024). Początkowo wszystkie klasy są uznawane za klasy jednoelementowe. Następnie dla każdej klasy jednoelementowej znajduje się najbardziej podobne inne klasy i łączy się je ze sobą. Na każdym etapie tego łączenia wylicza się macierz odległości. Jeśli chodzi o macierz odległości między skupieniami, to można wyróżnić następujące metody obliczania odległości między skupieniami, są to metody, między innymi pojedynczego wiązania, pełnego wiązania, średnich połączeń, środków ciężkości, mediany czy Warda (Statsoft.pl, 2024). Korzystając z metody Warda, dąży się do tego, aby różnice wewnątrzgrupowe pomiędzy obiektami były jak najmniejsze. Ostatecznie otrzymuje się jedną, spójną klasę, którą należy kolejno podzielić na skupienia (Balicki, 2013).

4. Wyniki badań

4.1. Statystyki opisowe

Pierwszym etapem analizy było wyliczenie podstawowych statystyk opisowych, które zostały przedstawione w tabeli 1.

Tabela 1. Statystyki opisowe badanych zmiennych

| Zmienne | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 |
|------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Statystyki opisowe | | | | | | | | |
| Średnia | 12,79 | 62,03 | 51,83 | 17,99 | 14,25 | 6,63 | 8,43 | 10,07 |
| Mediana | 12,13 | 61,85 | 54,20 | 17,40 | 10,70 | 6,05 | 8,35 | 10,65 |
| Minimum | 5,64 | 53,10 | 28,70 | 11,70 | 5,60 | 2,60 | 5,50 | 6,30 |
| Maksimum | 22,77 | 73,30 | 67,50 | 24,50 | 32,40 | 10,80 | 11,70 | 11,90 |
| Odchylenie standardowe | 4,67 | 5,16 | 9,91 | 3,20 | 8,00 | 2,30 | 1,83 | 1,53 |

Źródło: opracowanie własne.

Zmienna X_1 , dotycząca narażenia na zanieczyszczenie powietrza, przyjęła średnią wartość $12,79 \mu\text{g}/\text{m}^3$, a mediana ukształtowała się na poziomie $12,13 \mu\text{g}/\text{m}^3$. Wartość minimalna dla tej zmiennej wyniosła $5,64 \mu\text{g}/\text{m}^3$ dla Finlandii, co wskazuje, że w tym kraju w powietrzu znajduje się najmniej szkodliwych substancji w porównaniu do innych krajów UE. Natomiast maksymalna wartość jest związana z Polską i wyniosła $22,77 \mu\text{g}/\text{m}^3$. Statystyki dla zmiennej X_2 wskazują, że obywatele Szwecji cieszą się największą liczbą lat przeżytych w zdrowiu. Natomiast dla obywateli Łotwy wartość tej zmiennej wynosi około 53 lata, co stanowi minimum. Analizując zmienną X_3 , można stwierdzić, iż średnio 51,83% obywateli 25 krajów UE spożywa od 1 do 4 porcji warzyw i owoców dziennie. Minimalna wartość to 28,7% i została odnotowana dla Holandii, a maksymalna 67,50% dla Belgii. Jeśli chodzi o zmienną X_4 , dotyczącą odsetka osób otyłych na podstawie wskaźnika BMI, to średnio 17,99% społeczeństwa badanych krajów UE jest otyłych. Maksymalną wartość zaobserwowano dla Węgier – ponad 24,5% społeczeństwa jest otyła. Natomiast najmniejszą wartość wyróżniono dla Włoch. W przypadku zmiennej X_5 średnio około 14% obywateli UE podejmuje aktywność fizyczną przynajmniej raz w tygodniu. Dla zmiennej X_6 – objawów zaburzeń depresyjnych – średnia ukształtowała się na poziomie 6,63%. Minimalna wartość tej zmiennej została odnotowana dla Grecji i wyniosła 2,60%, natomiast maksymalna 10,80% dla Francji, X_7 określającej wydatki na system ochrony zdrowia wyniosła 8,43. X_8 można stwierdzić, iż średnio 10,07% obywateli UE spożywa codziennie alkohol. Co więcej, krajem, w którym obywatele spożywają najmniej alkoholu, jest Grecja. Natomiast najwięcej alkoholu spożywa się w Czechach. Na podstawie wyznaczonego odchylenia standardowego można wyciągnąć wniosek, iż lata przeżyte w zdrowiu różnią się przeciętnie o 5,16 lat od średniej wyznaczonej dla badanych krajów. Procent populacji poszczególnych krajów europejskich spożywających od 1 do 4 porcji warzyw i owoców dziennie różni się przeciętnie o 9,91% od wyznaczonej dla tej zmiennej średniej.

4.2. Model ekonometryczny

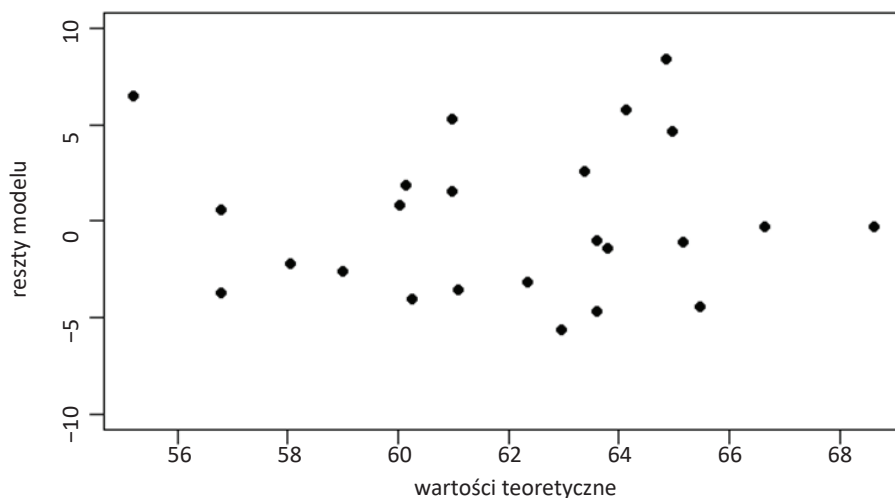
Pierwszym etapem było wytypowanie zmiennej objaśnianej i zmiennych objaśniających spośród analizowanych zmiennych. Zmienną objaśnianą jest zmienna X_2 – lata przeżyte w zdrowiu. Natomiast zmiennymi objaśniającymi są pozostałe zmienne. Następnie oszacowano parametry modelu, w którym było siedem zmiennych objaśniających. Przeprowadzono test t -Studenta, który umożliwia testowanie istotności poszczególnych parametrów. Z wyłączeniem zmiennej X_4 wszystkie parametry strukturalne zmiennych objaśniających miały p -value przekraczające przyjęty poziom istotności 0,05. W związku z tym nie było podstaw do odrzucenia hipotezy zerowej i parametry przy zmiennych nie były statystycznie istotne. Zatem zmienne: X_1 , X_3 , X_5 , X_6 , X_7 , X_8 nie miały istotnego wpływu na zmienną objaśnianą. W związku z tym postanowiono utworzyć drugi model, zawierający już tylko jedną zmienną objaśniającą – odsetek osób otyłych na podstawie wskaźnika BMI. Ponadto p -value dla testu t -Stu-

denta dla współczynnika wyrazu wolnego i zmiennej X_4 są mniejsze od przyjętego poziomu istotności 0,05. Zatem odrzucono hipotezę zerową i stwierdzono, iż zmienna X_4 ma istotny wpływ na zmienną objaśnianą. Ostateczna postać modelu jest następująca:

$$\hat{Y} = -1,047X_4 + 80,861. \quad (2)$$

Błędy szacunku parametrów wynoszą: $S(a_4) = 0,261$ i $S(a_0) = 4,767$. Jeśli chodzi o kierunek wpływu zmiennej objaśniającej, to w badanych krajach UE wzrost odsetka osób otyłych sprzyja zmniejszaniu się liczby lat przeżytych w zdrowiu. W modelu zostało wyjaśnione 36% zmienności zmiennej objaśnianej. Standardowy błąd modelu S_e jest równy 4,01, co wskazuje, iż przeciętne różnice między wartościami teoretycznymi, a rzeczywistymi zmiennej objaśnianej wynoszą około 4 lata.

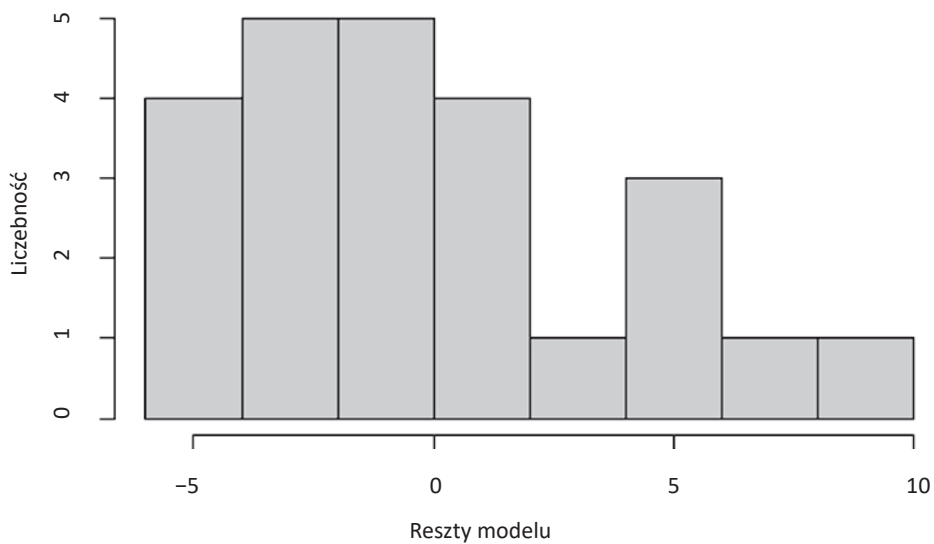
Następnie wykonano diagnostykę własności reszt. Wygenerowany został wykres reszt względem wartości teoretycznych, widoczny na rysunku 1.



Rys. 1. Wykres reszt względem wartości teoretycznych

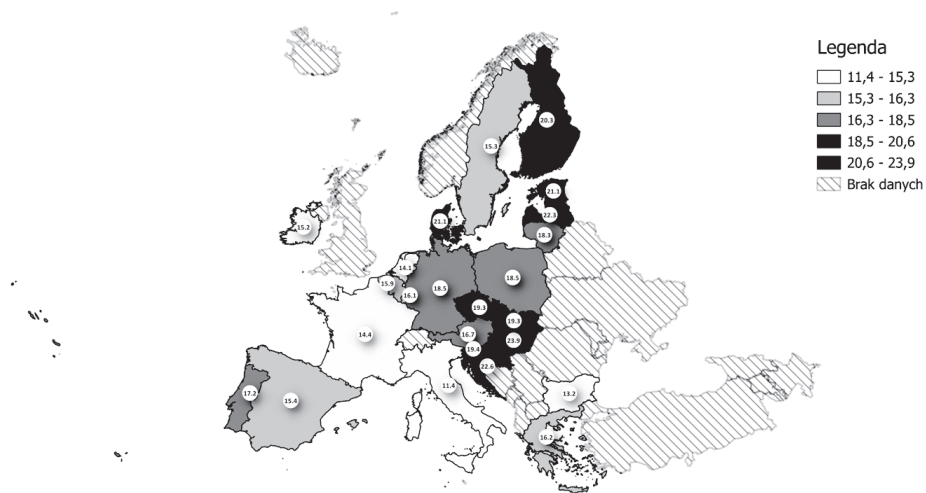
Źródło: opracowanie na podstawie badań własnych.

Na rysunku 1 nie można zauważyć żadnego wzorca sugerującego zależność funkcyjną, na przykład liniową czy w postaci wielomianu. Zatem nie występuje nieliniowość. Nie ma też żadnych skupień, co wskazuje na homoskedastyczność. Przeprowadzenie testu Breusch'a-Pagana również potwierdziło te wnioski. P-value ukształtowało się na poziomie 0,9774 oraz przyjęto poziom istotności 0,05. Ustalono zatem hipotezę zerową mówiącą o tym, iż występuje homoskedastyczność. Następnie utworzono histogram reszt modelu, który znajduje się na rysunku 2.



Rys. 2. Histogram rozkładu reszt modelu

Źródło: opracowanie na podstawie badań własnych.



Rys. 3. Odsetek osób otyłych na podstawie wskaźnika BMI w 25 krajach UE w 2019 roku

Źródło: opracowanie na podstawie badań własnych.

Na podstawie rysunku 2 można podejrzewać, iż reszty mają rozkład normalny. Normalność rozkładu zweryfikowano testem Shapiro-Wilka. Hipoteza zerowa tego testu mówi o tym, że reszty mają rozkład normalny, natomiast alternatywna przeciwnie – reszty nie mają rozkładu normalnego. Statystyka testowa wyniosła 0,94671. P-value ukształtowało się na poziomie 0,2297. W związku z tym nie ma podstaw do odrzucenia hipotezy zerowej. Można uznać, że reszty mają rozkład normalny.

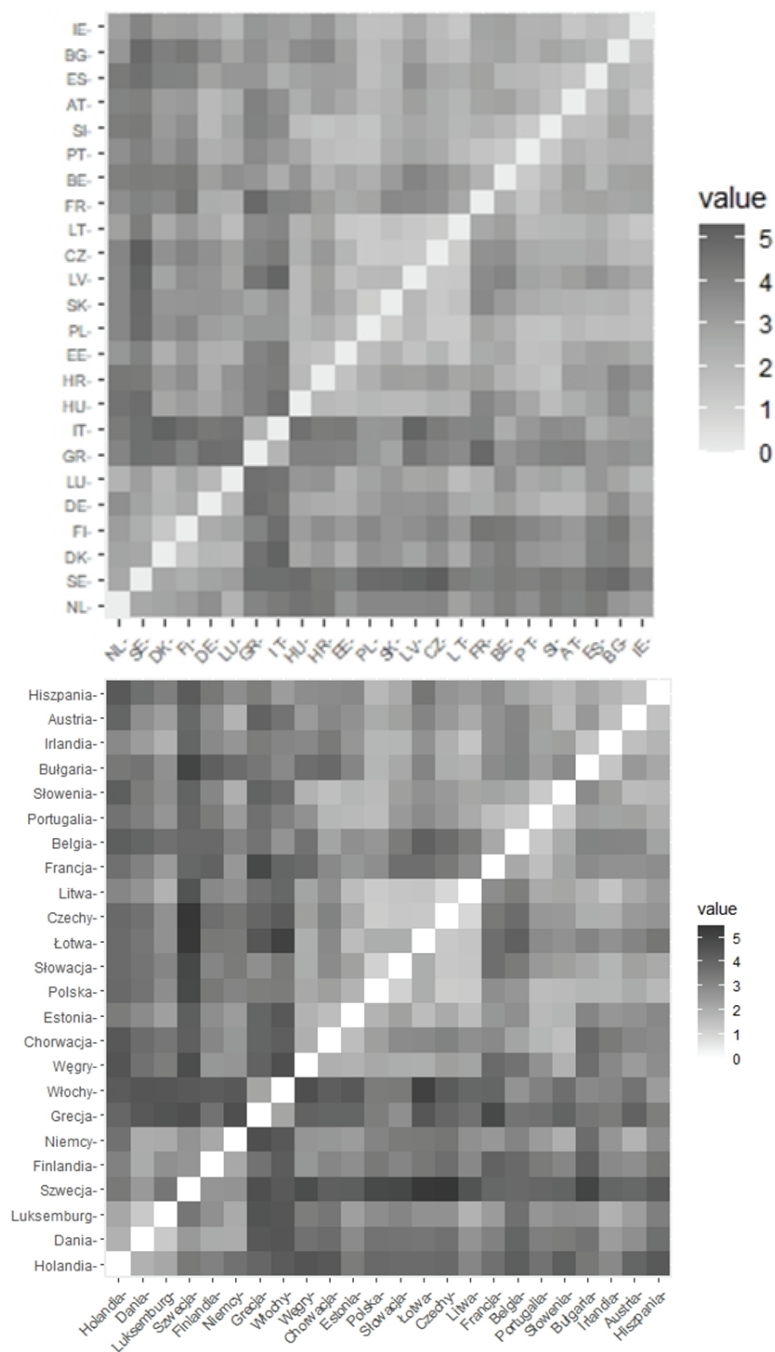
W związku z istotnym wpływem zmiennej X_4 na zmienną objaśnianą, zdecydowano o głębszej analizie tej zmiennej. Na rysunku 3 przedstawiono wartości zmiennej X_4 dla poszczególnych krajów UE. Na jego podstawie można stwierdzić, iż najniższy odsetek osób otyłych na podstawie wskaźnika BMI odnotowano w następujących krajach: Francji, Włoszech, Bułgarii, Irlandii oraz Holandii. Natomiast największe wartości dla tej zmiennej zostały zaobserwowane w Chorwacji, Danii, Estonii, na Węgrzech oraz na Łotwie. Ponadto zmienna lata przeżyte w zdrowiu przyjmuje niskie wartości dla tych krajów. W przypadku Irlandii warto również dodać, iż plasuje się ona na trzecim miejscu pod względem liczby lat przeżytych w zdrowiu. Natomiast na pierwszym miejscu i tym samym największą liczbę lat przeżytych w zdrowiu odnotowuje się w Szwecji. Należy także zauważyć, że w Szwecji odsetek osób otyłych na podstawie wskaźnika BMI nie jest duży w porównaniu do pozostałych badanych krajów. Zgodnie z rysunkiem 3 najlepiej wypadają Włochy. W przypadku lat przeżytych w zdrowiu zajmują one czwartą pozycję.

4.3. Klasyfikacja hierarchiczna

Zanim dokonano analizy skupień, wybrano spośród zmiennych takie, które posłużyły do scharakteryzowania współczesnego Europejczyka ze względu na prowadzony styl życia. Są to zmienne:

- X_3 – odsetek populacji spożywającej od 1 do 4 porcji warzyw i owoców dziennie (%),
- X_4 – odsetek osób otyłych na podstawie wskaźnika BMI (%),
- X_5 – odsetek społeczeństwa danego państwa deklarujący podejmowanie aktywności fizycznej przynajmniej raz w tygodniu (%),
- X_6 – odsetek społeczeństwa danego państwa cierpiącego na zaburzenia depresyjne (%),
- X_8 – spożycie alkoholu (litry/osobę).

Pierwszym etapem było przygotowanie danych do dalszej analizy. W tym celu wykonano standaryzację zmiennych. Aby określić, które obserwacje powinny być połączone, konieczny jest wybór miary podobieństwa między obiektami. Zatem wyznaczono odległości euklidesowe między badanymi państwami. Warto wspomnieć, iż na początku obiekty tworzą skupienia jednoelementowe. Macierz odległości umożliwia zidentyfikowanie i połączenie pary obiektów znajdujących się najbliżej siebie. To skutkuje redukcją wymiaru macierzy odległości. Następnie łączy się kolejne obiekty i skupienia ze sobą, aby ostatecznie uzyskać jedno duże skupienie przedstawione w formie dendrogramu. Celem połączenia skupień i obiektów ze sobą należy wyzna-



Rys. 4. Macierz odległości

Źródło: opracowanie na podstawie badań własnych.

czyć macierz odległości między skupieniami, korzystając z wybranej metody grupowania. Do analizy hierarchicznej skupień wykorzystano jedną z metod wyznaczania macierzy odległości między skupieniami – metodę Warda. To jedna z wielu metod obliczania odległości już między skupieniami, pozwalająca zachować warunek homogeniczności.

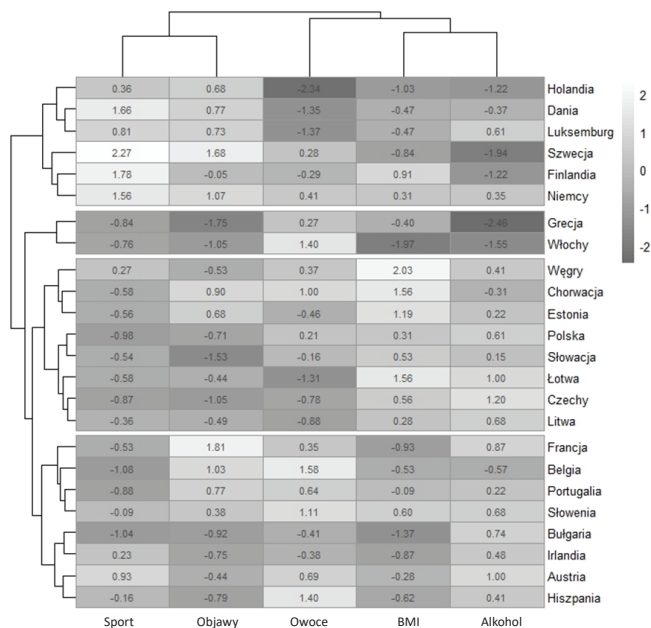
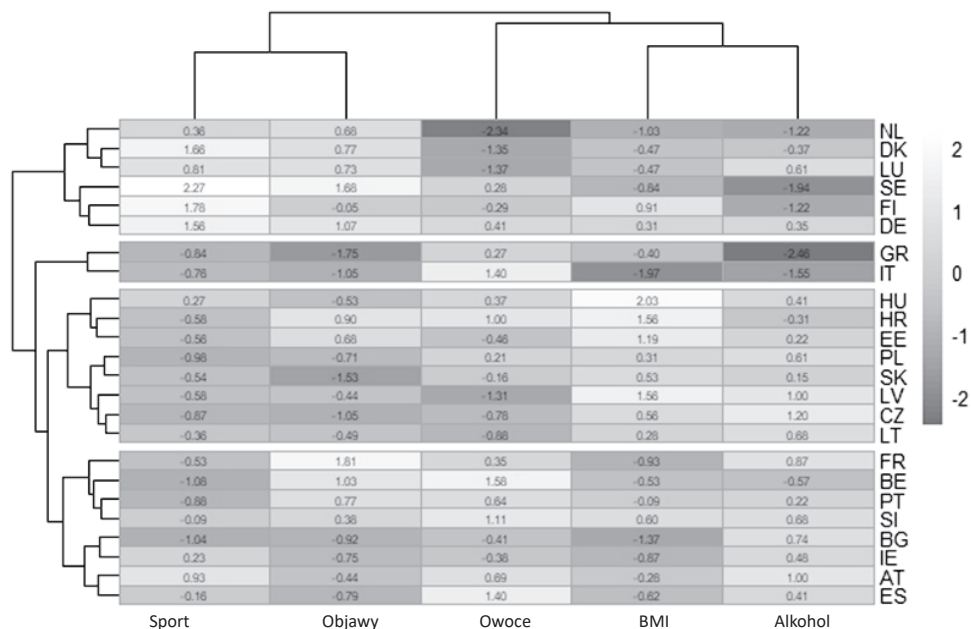
Na rysunku 4 przedstawiona została wcześniej wspomniana macierz odległości pomiędzy parami obiektów. Im wyższe wartości odległości, tym ciemniejsze kwadraty. Na podstawie wizualizacji macierzy odległości przedstawionej na rysunku 4 szuka się pary obiektów najbardziej podobnych do siebie. Są one podobne w momencie, gdy odległość między nimi jest najmniejsza.

Na rysunku 5 znajduje się wizualizacja wartości standaryzowanych oraz podział obiektów badanych na grupy.

Jeśli dla danego kraju otrzymano ujemną wartość standaryzowaną, oznacza to, że badane zjawisko plasuje się poniżej średniej. Przykładowo dla Holandii – należącej do trzeciego skupienia – wartość standaryzowana wyniosła $-1,22$ dla zmiennej – dzienna ilość spożytego alkoholu. Oznacza to, że dzienne spożycie alkoholu w Holandii jest niższe od średniej dla krajów UE (rys. 5).

Na podstawie wzrokowej analizy dendrogramu utworzono cztery skupienia. Im większa była odległość niż 4,5, tym zmniejszało się podobieństwo pomiędzy analizowanymi obiektami i klasy stawały się coraz bardziej heterogeniczne. Ostatecznie dendrogram przecięto w odległości równej 4,5. Do pierwszego skupienia należą Włochy i Grecja. W skład drugiego skupienia weszły następujące kraje: Węgry, Chorwacja, Estonia, Polska, Słowacja, Litwa, Czechy i Łotwa. Trzecią klasę tworzą kraje Europy Zachodniej: Holandia, Szwecja, Dania, Finlandia, Niemcy i Luksemburg. W skład czwartego, ostatniego skupienia, wchodzi: Francja, Belgia, Portugalia, Słowenia, Austria, Hiszpania, Bułgaria oraz Irlandia (rys. 5).

Dla wartości standaryzowanych średnia arytmetyczna wynosi 0. Wartości dodatnie oznaczają, iż zjawisko obserwowane w tych obiektach kształtuje się powyżej średniej, natomiast ujemne wskazują, iż wartość jest poniżej średniej. Warto zwrócić uwagę na zmienną dotyczącą aktywności fizycznej, której wartości dla trzeciego skupienia są dodatnie, natomiast dla pierwszego i drugiego, poza Węgrami, są ujemne. Analogiczna sytuacja występuje dla zmiennej opisującej odsetek społeczeństwa danego państwa, który cierpi na zaburzenia depresyjne. Dla trzeciego skupienia obserwuje się wartości dodatnie (poza Finlandią), natomiast ujemne wartości występują dla skupienia pierwszego. Analizując zmienną opisującą odsetek osób otyłych na podstawie wskaźnika BMI, można stwierdzić, iż dodatnie wartości występują dla skupienia drugiego, a ujemne dla skupienia pierwszego oraz czwartego, z wyłączeniem Słowenii. Jeśli chodzi o zmienną dotyczącą spożycia alkoholu, to w skupieniu drugim (oprócz Chorwacji) oraz w skupieniu czwartym (oprócz Belgii) występują dodatnie wartości. Z kolei ujemne występują w skupieniu pierwszym.



Rys. 5. Wizualizacja wartości standaryzowanych i podziału na skupienia (metoda Warda)

Źródło: opracowanie na podstawie badań własnych.

5. Podsumowanie

Przeprowadzenie analizy czynników wpływających na lata przeżyte w zdrowiu w państwach należących do UE pozwoliło wyciągnąć wnioski oraz odpowiedzieć na wcześniej postawione pytania badawcze. Pierwsze pytanie badawcze dotyczyło modelu ekonometrycznego. W modelu tym znalazła się tylko jedna zmienna, czyli odsetek osób otyłych na podstawie wskaźnika BMI. Pozostałe hipotetyczne symptomy nie zostały uwzględnione w ostatecznym modelu ze względu na brak istotności parametrów. Należy jednak podkreślić, że dopasowanie modelu do danych nie było zbyt wysokie, więc nie można jednoznacznie stwierdzić, iż tylko ta zmienna wpływa w pewien sposób na badane zjawisko. Ważne jest to, w jaki sposób każdy się odżywia i jaki styl życia prowadzi. Jednak ze względu na to, iż wskaźnik BMI może być niemiarodajny, gdyż nie mierzy stopnia otyłości organizmu tylko jego masę, w celu głębszego wyjaśnienia tematu należy przeprowadzić dalsze analizy, wykorzystując dodatkowe zmienne. Drugie pytanie badawcze mówiące o tym, że Polacy są podobni pod względem prowadzonego stylu życia do innych obywateli krajów położonych w Europie Środkowo-Wschodniej potwierdziło się. Wynika to z faktu, że w klasyfikacji hierarchicznej Polska pogrupowała się z następującymi krajami: Węgrami, Chorwacją, Estonią, Słowacją, Litwą, Czechami oraz Łotwą. Zatem zgodnie z obowiązującą zasadą dotyczącą klasyfikacji obiekty te są do siebie podobne, gdyż należą do jednego skupienia. Podobieństwo między tymi krajami widoczne jest przede wszystkim pod względem zmiennej opisującej odsetek osób otyłych na podstawie wskaźnika BMI, ale także pod względem zmiennej opisującej odsetek społeczeństwa danego państwa, który deklaruje podejmowanie aktywności fizycznej przynajmniej raz w tygodniu, a także pod względem zmiennej dotyczącej spożycia alkoholu, z wyłączeniem kolejno Węgier oraz Chorwacji. Do trzeciego skupienia w metodzie klasyfikacji hierarchicznej należą kraje, takie jak: Holandia, Szwecja, Dania, Finlandia, Niemcy i Luksemburg. Na podobieństwo między nimi wskazuje przede wszystkim zmienna dotycząca aktywności fizycznej, ale również zmienne dotyczące odsetka społeczeństwa danego państwa cierpiącego na zaburzenia depresyjne, z wyjątkiem Finlandii. Obiekty te są do siebie podobne, zatem drugie pytanie badawcze: „Czy konfiguracja czynników dotyczących stylu życia dla krajów Europy Zachodniej jest zbliżona do siebie i dodatkowo porównywalna z sytuacją krajów skandynawskich?”, uzyskało twierdzącą odpowiedź.

Literatura

- Achremowicz, B. (2012). Od czego zależy długość naszego życia? *Wszechświat*, 113(10-12), 285-289.
- Aeaweb.org. (2023). *JEL Classification System / EconLit Subject Descriptors*. Pobrano 4 marca 2024 z [https://www.aeaweb.org/econlit/jelCodes.php#:~:text=EconLit%](https://www.aeaweb.org/econlit/jelCodes.php#:~:text=EconLit%20)
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Bogaert, P., Van Oyen, H., Beluche, I., Cambois, E., I Robine, M.J. (2018). The Use of the Global Activity Limitation Indicator and Healthy Life Years by Member States and the European Commission. *Archives of Public Health*, 76(30). <https://doi.org/10.1186/s13690-018-0279-z>
- Dziechciarz, J. (2012). *Ekonometria: metody, przykłady, zadania*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Ec.europa.eu (2015, 9 czerwca). *Glossary: Healthy Life Years (HLY)*. Pobrano 10 grudnia 2023 z [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Healthy_life_years_\(HLY\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Healthy_life_years_(HLY))
- Ec.europa.eu. (2022, 4 kwietnia). *Daily Consumption of Fruit and Vegetables*. Pobrano 10 grudnia 2023 z [https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_fv3e\\$DV_462/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_fv3e$DV_462/default/table?lang=en)
- Ec.europa.eu. (2022, 27 kwietnia). *Obesity Rate by Body Mass Index (BMI)*. Pobrano 10 grudnia 2023 z https://ec.europa.eu/eurostat/cache/metadata/en/sdg_02_10_esmsip2.htm
- Ec.europa.eu. (2023, 16 czerwca). *Current Depressive Symptoms by Sex, Age and Educational Attainment Level*. Pobrano 10 grudnia 2023 z [https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_MH1E\\$DEFAULTVIEW/default/table](https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_MH1E$DEFAULTVIEW/default/table)
- Ec.europa.eu. (2023, 1 sierpnia). *Performing Health-Enhancing Physical Activity*. Pobrano 10 grudnia 2023 z https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_pe9e__custom_10309252/default/table?lang=en&page=time:2019
- Lalonde, M. (1974). *A New Perspective on the Health of Canadians*. Public Health Agency of Canada.
- Oecd.org. (2020). *Air pollution exposure*. Pobrano 10 grudnia 2023 z <https://data.oecd.org/air/air-pollution-exposure.htm>
- Oecd.org. (2023a). *Alcohol Consumption*. Pobrano 10 grudnia 2023 z <https://data.oecd.org/healthrisk/alcohol-consumption.htm#:~:text=Alcohol%20use%20is%20associated%20with,%2C%20violence%2C%20homicide%20and%20suicide.>
- Oecd.org. (2023b). *Health spending*. Pobrano 10 grudnia 2023 z <https://data.oecd.org/healthres/health-spending.html>
- Sauerberg, M. (2021). The Impact of Population's Educational Composition on Healthy Life Years: An Empirical Illustration of 16 European Countries. *Elsevier*, 15(100857). <https://doi.org/10.1016/j.ssmph.2021.100857>
- Sojka, E. (2020). *Selected Health Indicators and Their Economic Determinants in the Countries of the European Union*, s. 186-199. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Statsoft.pl. (b.d.). *Analiza skupień*. Pobrano 29 grudnia 2023 z https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstcluan.html

Analysis of the Lifestyle of European Union Citizens in the Context of Healthy Life Years

Abstract: The subject of the article is an analysis of the lifestyle of citizens of European Union countries and factors influencing the years lived in good health. The aim of the analysis is to identify these factors and compare European countries in terms of them. Data analysis was carried out using descriptive statistics and an econometric model in order to examine the impact of explanatory variables on the explained variable, which is years lived in health. Cluster analysis was also performed for lifestyle variables. It has been shown that Poland is similar in terms of selected factors to other countries located in Central and Eastern Europe. Moreover, their configuration is similar to Western European countries and additionally comparable to Scandinavian countries.

Keywords: healthy life years, lifestyle, modern European, econometric model, cluster analysis

Wiktoría Galarowicz

e-mail: 177503@student.ue.wroc.pl

ORCID: 0009-0006-8381-3569

Uniwersytet Ekonomiczny we Wrocławiu

Predykcja cen mieszkań w dużych miastach w Polsce z wykorzystaniem uczenia maszynowego

DOI: 10.15611/2024.53.6.05

JEL Classification: R31

© 2024 Wiktoría Galarowicz

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Galarowicz, W. (2024). Predykcja cen mieszkań w dużych miastach w Polsce z wykorzystaniem uczenia maszynowego. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 68-81). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Artykuł analizuje kluczowe zmienne wpływające na kształtowanie cen na rynku nieruchomości w dużych miastach w Polsce. Omówiono w nim pojęcie rynku nieruchomości oraz jego najważniejszych cech, szczegółowo wyjaśniono, jak zmieniały się ceny mieszkań na przestrzeni lat w Polsce. W latach 90. XX wieku ceny notowały wzrost wywołany liberalizacją i łatwym dostępem do kredytów hipotecznych. Jednakże po kryzysie finansowym z 2007 roku rynek doświadczył znacznych wahań. Pandemia COVID-19, w porównaniu z programem Bezpieczny Kredyt 2%, wywarła niewielki wpływ na ceny. Analiza cen mieszkań i kluczowych zmiennych wpływających na ich kształtowanie obejmowała przygotowanie danych, imputowanie brakujących wartości oraz identyfikację odstających obserwacji. Do prognozowania cen wykorzystano algorytmy uczenia maszynowego, Random Forest (las losowy) i XGBoost.

Słowa kluczowe: rynek nieruchomości, ceny mieszkań, uczenie maszynowe, model Random Forest, model XGBoost

1. Wstęp

Artykuł porusza tematykę rynku nieruchomości, który odgrywa kluczową rolę w gospodarce każdego kraju. Jest jednym z najważniejszych sektorów, który wpływa na życie codzienne milionów ludzi. Zrozumienie mechanizmów kształtujących ceny mieszkań w dużych miastach jest istotne zarówno dla deweloperów, inwestorów, jak i nabywców mieszkań. Motywacją do przeprowadzenia niniejszego badania była potrzeba głębszej analizy czynników wpływających na ceny mieszkań oraz ocena skuteczności metod uczenia maszynowego w prognozowaniu tych cen. Dla de-

weloperów i inwestorów wiedza o przyszłych trendach cenowych jest kluczowa w podejmowaniu decyzji inwestycyjnych. Dla osób planujących zakup mieszkania wyniki badania mogą być pomocne w podejmowaniu świadomych decyzji konsumenckich.

Celem teoretycznym artykułu jest zdobycie informacji o zmianach na rynku nieruchomości w latach 1989-2024. Natomiast celem badawczym artykułu jest stworzenie modeli uczenia maszynowego przewidującego ceny mieszkań w dużych miastach w Polsce. Oprócz tego szukana jest również odpowiedź na pytanie badawcze, jakie cechy nieruchomości i jej otoczenia najbardziej wpływają na kształtowanie się cen nieruchomości.

Skupienie się na dużych miastach wynika z ich znaczenia ekonomicznego i demograficznego. Duże miasta charakteryzują się większą dynamiką rynku nieruchomości, co czyni je interesującymi obszarami analizy. Metody uczenia maszynowego, takie jak Random Forest i XGBoost, zostały wybrane ze względu na ich wysoką skuteczność w analizie dużych zbiorów danych oraz zdolność do identyfikacji nieliniowych zależności między zmiennymi.

2. Definicja i cechy rynku nieruchomości

Na przestrzeni lat rynek nieruchomości charakteryzuje się dynamicznymi zmianami. Od 1989 roku (początek transformacji ustrojowej) nieruchomości przestały być dobrami socjalnymi i stały się przedmiotem handlu. Są to specyficzne dobra, które podlegają obrotowi prawami do nich, a nie samymi budynkami czy gruntami (Gołąbeska, 2017). Oznacza to, że transakcje dotyczą transferu różnorodnych praw użytkowania i obowiązków właścicieli. Te prawa obejmują własność, współwłasność, użytkowanie wieczyste gruntu, prawa rzeczowe ograniczone (użytkowanie, służebność), spółdzielcze własnościowe prawo do lokalu, hipotekę oraz prawa zobowiązaniowe jak najem i dzierżawa.

Obrót nieruchomościami można podzielić na dwie kategorie (Gołąbeska, 2017): rynkowe i nierynkowe. Transakcje rynkowe obejmują kupno/sprzedaż, najem/dzierżawę, użytkowanie, podnajem i zamianę, podczas gdy transakcje nierynkowe dotyczą darowizn, spadkobrania, wywłaszczenia, uwłaszczenia, eksmisji i egzekucji. Obroty nierynkowe nie mają na celu osiągnięcia zysku, ale mogą być wynikiem przeniesienia własności w różnych sytuacjach, np. spadek, darowizna, wywłaszczenie czy egzekucja. Transakcje rynkowe zwykle służą zaspokojeniu potrzeb mieszkaniowych lub prowadzeniu działalności gospodarczej. Mogą też być wynikiem inwestycji w celu osiągnięcia dodatkowych dochodów, np. poprzez zysk z transakcji kupna-sprzedaży, dochody z reklam lub dochody czynszowe z najmu nieruchomości. W zależności od rodzaju nieruchomości inwestycja może być krótko- lub długoterminowa oraz służyć zabezpieczeniu finansowemu na przyszłość.

Rynek nieruchomości jest miejscem, gdzie odbywają się transakcje kupna, sprzedaży, dzierżawy czy wynajmu, lecz jego definicja jest różna w zależności od kontekstu. Jak podaje E. Gołąbeska (2007) można go postrzegać jako interakcje między ludźmi lub instytucjami zajmującymi się nieruchomościami, wszystkie transakcje nieruchomościami, forum do zawierania umów oraz zestaw układów, gdzie ustalane są ceny i prawa własności lub wymianę prawa własności nieruchomości na inne aktywa. Niemniej, najbardziej kompletną definicją wydaje się być stwierdzenie, że rynek nieruchomości to ogół warunków, w których dokonuje się przekazywania praw do nieruchomości i zawierania umów dotyczących ich użytkowania (Gołąbeska, 2007).

Trzema głównymi czynnikami, które kształtują rynek nieruchomości (w tym również rynek mieszkaniowy) są: popyt, podaż i cena. Popyt na nieruchomości odzwierciedla siłę nabywczą osób poszukujących lokali do kupna lub wynajmu. Można wyróżnić dwa rodzaje popytu: potencjalny i efektywny (Gołąbeska, 2017). Popyt potencjalny związany jest z ludzkimi potrzebami i aspiracjami, podczas gdy popyt efektywny opiera się na dostępnych środkach finansowych. Istotną cechą popytu na nieruchomości jest ich unikalność i brak substytutów, ponieważ jak twierdzi E. Kucharska-Stasiak (2000), potrzeba posiadania własnego mieszkania lub domu nie może być zastąpiona innym dobrem. Podaż odnosi się do liczby dostępnych na rynku nieruchomości, jakie mają do zaoferowania producenci w określonym czasie i za określoną cenę. Wzrost cen nieruchomości skutkuje zwykle wzrostem podaży, ponieważ wyższe ceny zachęcają producentów do szybkiego zarobku poprzez sprzedaż lub wynajem (Gołąbeska, 2017). Podaż nieruchomości obejmuje zarówno istniejące nieruchomości na rynku wtórnym, jak i nowe inwestycje budowlane na rynku pierwotnym. Obejmuje to także różne formy transakcji, takie jak sprzedaż, wynajem, dzierżawa czy najem lokali. Deweloperzy i właściciele nieruchomości są głównymi dostawcami na rynku nieruchomości, dostarczając zarówno nowych, jak i zmodernizowanych mieszkań dostosowanych do bieżących trendów oraz oczekiwań klientów. Natomiast cena jest relacją pomiędzy popytem a podażą.

Rynek nieruchomości znacznie różni się od innych rynków, a jego stałymi cechami, jak podaje M.J. Bryx (2006) są, m.in: mała elastyczność popytu i podaży, niedoskonałość, lokalny charakter, niejednorodność, niska efektywność, niepowtarzalność i duży zakres interwencji państwa. Mała elastyczność popytu i podaży ukazuje się poprzez małą reakcję na zmiany cen. Niedoskonałość rynku objawia się w nieprawidłowym funkcjonowaniu mechanizmu ustalania ceny równowagi mimo swobodnego popytu i podaży. To wynika częściowo z niemożności zmiany lokalizacji nieruchomości, ograniczonej podaży oraz długotrwałego procesu inwestycyjnego. Lokalny charakter rynku ogranicza konkurencję, ponieważ duże znaczenie ma lokalizacja nieruchomości. Niejednorodność wynika z różnorodności rodzajów nieruchomości (czego przejawem jest podział na rynek międzynarodowy, krajowy, regionalny i lokalny). Natomiast niską efektywność potwierdza problem z właściwą wyceną nieruchomości, a informacje, które posiada inwestor, mogą nie być wystarczające,

aby opracować strategie handlowe. Dodatkowo występuje duży interwencjonizm państwa poprzez wysokość podatków, politykę czynszową i regionalną oraz ochronę zabytków. Rynek nieruchomości jest wyjątkowy nie tylko ze względu na unikatowość poszczególnych nieruchomości, ale także ze względu na ich liczbę, łączną powierzchnię i wartość. Niepowtarzalność tego rynku wynika z różnorodności popytu i podaży, która jest determinowana przez strukturę społeczeństwa i jego zamożność. Silne przywiązanie do lokalizacji sprawia, że ludzie chętniej pozostają w danym regionie, co zwiększa popyt na lokalnym rynku. Te cechy są niemal stałe niezależnie od lokalizacji, czasu czy otoczenia prawno-ekonomicznego, co wynika z charakteru samej nieruchomości jako towaru.

3. Zmiany na rynku mieszkaniowym w ostatnich trzydziestu latach w Polsce

Na wzrost i spadki cen mieszkań wpływają takie czynniki, jak rozwój gospodarczy, zmiany demograficzne i czynniki polityczne. Początek lat 90. XX wieku był okresem intensywnego rozwoju rynku nieruchomości w Polsce (Gołąbeska, 2017). Upadek komunizmu wywołał społeczną euforię i zwiększył optymizm co do przyszłości. W tym czasie ceny mieszkań dynamicznie rosły, ponieważ ludzie oczekiwali lepszego życia i chcieli posiadać własne mieszkania. Liberalizacja gospodarki, wprowadzenie wolnego rynku oraz łatwy dostęp do kredytów hipotecznych spowodowały zwiększenie popytu na mieszkania, co doprowadziło do szybkiego wzrostu cen na rynku pierwotnym i wtórnym.

W latach 2000-2005 przeważała korzystna sytuacja dla deweloperów, gdyż na rynku nieruchomości mieszkaniowych panowała hossa. Deweloperzy przejęli rolę spółdzielni mieszkaniowych budujących mieszkania. Wyjątkiem był 2002 rok i początek 2003 roku, kiedy nastąpiło spowolnienie i recesja. Ten stan nie utrzymał się długo. Ożywienie na rynku nieruchomości było w znacznym stopniu wynikiem przystąpienia Polski do Unii Europejskiej oraz wzmożonego zainteresowania możliwością zaciągania kredytów hipotecznych. W roku 2004 polski sektor bankowy doświadczył boomu hipotecznego, który był rezultatem znacznego zainteresowania społeczeństwa właśnie tą formą finansowania nieruchomości. Wzrost zadłużenia polskich gospodarstw domowych był widoczny w zwiększonej liczbie udzielanych kredytów klientom indywidualnym. Wzrost zadłużenia hipotecznego przekształcił się w Polsce w boom ekonomiczny w sektorze bankowym, co doprowadziło do powstania silnej asymetrii w systemie finansowania rynku mieszkaniowego. Deweloperzy preferowali tanie źródła finansowania, takie jak przedpłaty klientów, a ryzyko kredytowe w ich działalności było wyższe, co wymagało odpowiednich zabezpieczeń. Boom hipoteczny spowodował duży wzrost cen nieruchomości mieszkaniowych po 2005 roku, co z kolei wywołało reakcję ze strony podaży. W okresie wzrostu cen, czyli w latach 2007-2008, marże deweloperów sięgały poziomu 50-60% (Kucharska-

-Stasiak i in., 2012). Wzrost cen mieszkań nie był uzasadniony wzrostem dochodów gospodarstw domowych. Dodatkowo, czynniki makroekonomiczne, których spadające wartości przyczyniły się do kryzysu finansowego z 2007 roku, negatywnie wpłynęły na rynek nieruchomości. W drugiej połowie 2008 roku oddano do użytku mieszkania, których budowę rozpoczęto wcześniej, co spowodowało nadmierną podaż i spadek cen nieruchomości. W związku z tym deweloperzy ograniczyli nowe inwestycje i zaczęli proponować klientom upusty lub dostosowywali nieruchomości do ich potrzeb (np. przez przekształcanie mieszkań na mniejsze). Doprowadziło to do poprawy sytuacji dopiero w połowie 2009 roku, a w 2010 rynek ustabilizował się, choć nadal przeważała podaż nad popytem. W 2012 roku wprowadzono zmiany prawne, takie jak zakończenie programu „Rodzina na swoim” (Majorek, 2013) i ustawę deweloperską, co skutkowało wzrostem liczby dostępnych nieruchomości mieszkaniowych i rozpoczęciem nowych inwestycji przez deweloperów, a w 2013 roku rynek mieszkań zbliżył się do stanu równowagi, a jego dalszy rozwój zależał głównie od sytuacji makroekonomicznej. Obserwowano wtedy poprawę wskaźników sytuacji mieszkaniowej: wzrost zasobu mieszkaniowego, spadek liczby osób w mieszkaniu i zwiększenie średniej powierzchni użytkowej mieszkania na osobę. W 2015 roku odnotowano najwyższy poziom sprzedaży od 2007. Większość transakcji była finansowana przez nabywców ze środków własnych. Liczba transakcji rosła przy jednoczesnym braku wzrostu cen nieruchomości. Wzrost popytu spowodowany niskimi stopami procentowymi skłonił do większych inwestycji gotówkowych przez osoby posiadające kapitał, które szukały bardziej opłacalnych alternatyw dla nisko oprocentowanych lokat. Nieruchomości coraz częściej postrzegano nie tylko jako miejsce do zamieszkania, ale także jako atrakcyjną formę inwestycji kapitału.

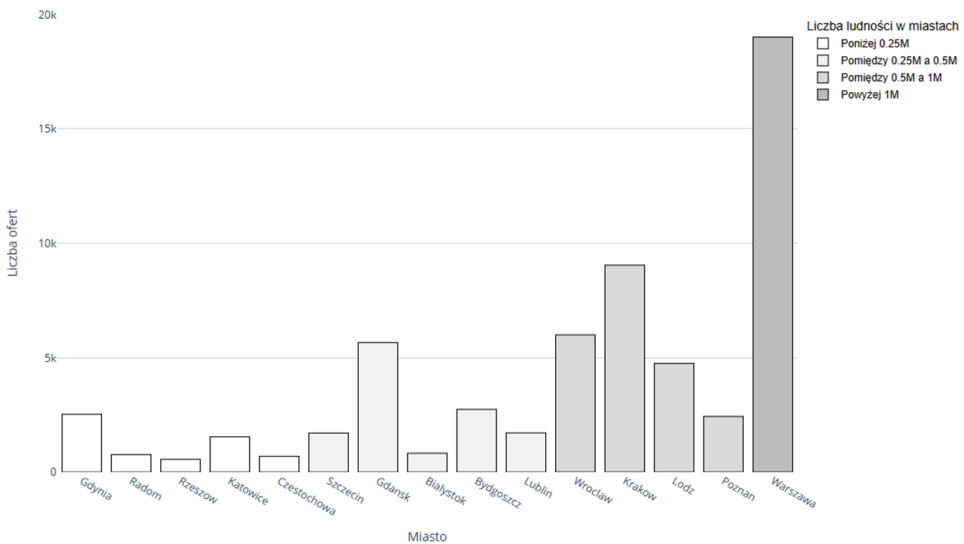
Jak podają B. Samorek i M. Cichocki (2023) zarówno rynek pierwotny, jak i wtórny, wykazały się wysoką odpornością na międzynarodowy kryzys spowodowany pandemią. W 2020 roku nie zaobserwowano istotnego wpływu pandemii na ceny nieruchomości, a dynamika zmian cen pozostała na podobnym poziomie jak w poprzednich latach. Pomimo okresowych fluktuacji rynek nieruchomości pozostawał stosunkowo stabilny, choć pandemia miała wpływ na pewne okresowe zmiany w gospodarce kraju. Końcem roku 2020, mimo zaostrzenia przepisów, ceny nieruchomości w wielu miastach powróciły do poziomu z początku roku, a zainteresowanie zakupem i sprzedażą nieruchomości wzrosło.

W lipcu 2023 roku można było zacząć składanie wniosku o Bezpieczny Kredyt 2%. Biuro Prasowe FIN-HUB (b.d.) podaje, że kredytobiorcy mogli liczyć na rządowe dopłaty do rat kredytu przez dziesięć lat. Spowodowało to spadek miesięcznych rat do 40%. Zainteresowanie tym programem miało wyraźny wpływ na ceny mieszkań. Wprowadzenie kredytu hipotecznego z rządową dopłatą zwykle powoduje wzrost cen mieszkań zarówno na rynku pierwotnym, jak i wtórnym. Można było zauważyć wyraźną tendencję podnoszenia cen nieruchomości do poziomu 700-800 tys. zł, czyli do ustalonych przez rząd limitów dopłat.

4. Analiza cen mieszkań w dużych miastach w Polsce

4.1. Opis zbioru danych

Przeprowadzono analizę, której celem było stworzenie modeli prognozujących ceny mieszkań w dużych miastach w Polsce oraz porównanie skuteczności tych modeli. Do analizy wybrano dwa algorytmy uczenia maszynowego: las losowy oraz XGBoost. Dodatkowo postanowiono zbadać, jakie czynniki najbardziej wpływają na ceny mieszkań, co pozwoli na lepsze zrozumienie rynku nieruchomości.



Rys. 1. Liczba dostępnych ofert mieszkaniowych w dużych miastach w Polsce

Źródło: opracowanie własne.

Wykorzystane w analizie dane pochodzą ze strony internetowej Kaggle.com (Jamroz, 2024) i dotyczą ceny oraz właściwości nieruchomości w piętnastu największych miastach w Polsce (Warszawa, Kraków, Wrocław, Łódź, Poznań, Gdańsk, Szczecin, Bydgoszcz, Lublin, Katowice, Białystok, Częstochowa, Rzeszów, Radom, Gdynia). Odnoszą się one do okresu sierpień 2023-styczeń 2024. Ich zbiór zawierał następujące zmienne:

- *city* – nazwa miasta, w którym znajduje się nieruchomość;
- *squareMeters* – powierzchnia mieszkania w metrach kwadratowych;
- *rooms* – liczba pokoi w mieszkaniu;
- *floor* – piętro, na którym znajduje się mieszkanie;
- *floorCount* – liczba pięter w budynku;
- *centreDistance* – odległość od centrum miasta w kilometrach;

- *poiCount* – liczba punktów usługowych w promieniu 500 metrów od mieszkania;
- *[poiName]Distance* – odległość do najbliższego punktu usługowego (szkoły, kliniki, urzędy pocztowe, przedszkola, restauracje, uczelnie, apteki);
- *ownership* – rodzaj własności nieruchomości;
- *has[features]* – informacja czy nieruchomość posiada określone cechy (miejsce parkingowe, balkon, ochronę, pomieszczenie gospodarcze);
- *price* – cena oferty w PLN.

Wśród analizowanych danych było ponad 40 000 duplikatów; po ich usunięciu pozostało około 60 000 rekordów. Zdecydowana większość ofert pochodzi z Warszawy (rys. 1). Warto zauważyć, że im większa liczba ludności w mieście, tym więcej ofert mieszkań.

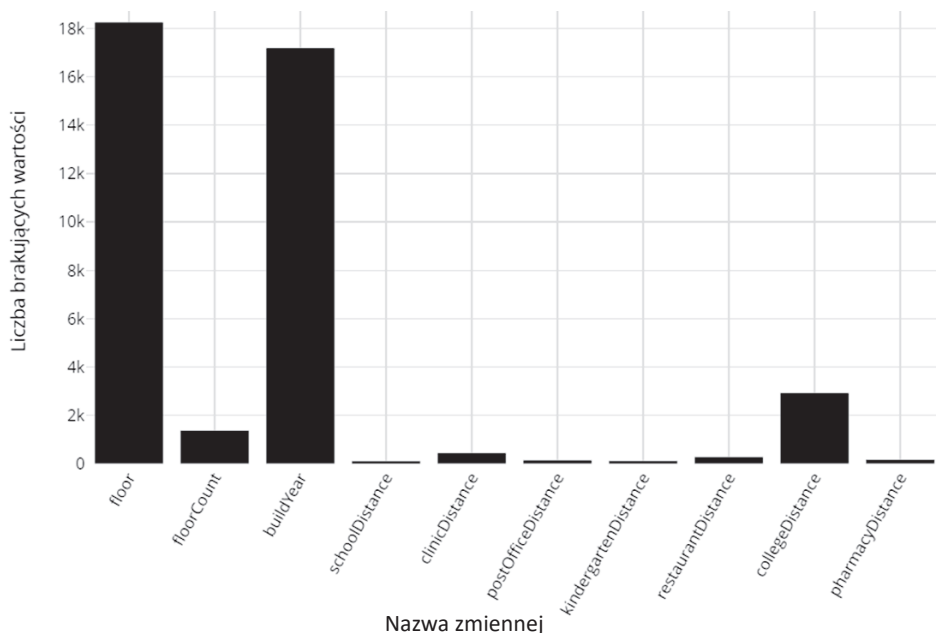
W związku z faktem, że zmienną zależną jest cena mieszkania, zdecydowano się na obliczenie współczynnika korelacji liniowej Pearsona (Sadowski, 1976). Silną zależność dodatnią zauważono dla zmiennych: liczba metrów kwadratowych mieszkania i liczba pokoi w mieszkaniu. Pozostałe zmienne charakteryzują się bardzo słabą korelacją, co wskazuje na niezasadność budowy klasycznych modeli regresji, przez co zdecydowano się na wykorzystanie algorytmów uczenia maszynowego.

4.2. Przygotowanie danych – imputacja braków danych i identyfikacja obserwacji odstających

Przed przystąpieniem do analizy właściwej konieczne było wykonanie szeregu operacji przygotowawczych. Proces ten obejmował, m.in. imputację braków danych oraz identyfikację i obsługę wartości odstających.

Dziesięć zmiennych posiadało niepełne informacje (rys. 2). Najwięcej braków mają zmienne określające rok budowy (około 17 000 braków) i numer piętra, na którym znajduje się mieszkanie (około 18 000 braków). Pozostałe charakteryzują się stosunkowo niewielką liczbą braków. Przez wzgląd na to, że każda z tych zmiennych jest numeryczna, postanowiono najpierw zestandaryzować dane, a następnie imputować braki z wykorzystaniem metody *k*-najbliższych sąsiadów (KNN). Polega ona na poszukiwaniu najbardziej podobnych przypadków ze znanymi wartościami atrybutów, które są używane do estymacji brakujących danych (Malarvizhi i Thanamani, 2012). Do uzupełnienia danych wybrano dziesięć najbardziej podobnych ofert.

Kolejnym etapem analizy była identyfikacja obserwacji odstających. Wykorzystano w tym celu metodę Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Jest to metoda, która identyfikuje punkty jako odstające, jeśli znajdują się w obszarach o niskiej gęstości (Westerlund, 2023). Następnie do automatycznej identyfikacji punktów łamania w krzywej odległości od najbliższych dziesięciu sąsiadów zastosowano algorytm KNEED. To podejście wykazało, iż około 14% danych jest obserwacjami odstającymi. Zdecydowano się na ich usunięcie z dalszej analizy.



Rys. 2. Liczba brakujących wartości w poszczególnych kolumnach

Źródło: opracowanie własne.

4.3. Modele uczenia maszynowego

Po podziale zbioru danych na zbiór testowy (20% ofert) i uczący (80% ofert) przystąpiono do modelowania. Pierwszym wykorzystanym modelem był las losowy, który opiera się na zasadzie tworzenia wielu drzew decyzyjnych (Jaiswal i Samikannu, 2017). Do optymalizacji hiperparametrów:

- *n_estimators* – liczba drzew;
- *max_depth* – maksymalna głębokość każdego drzewa;
- *max_features* – maksymalna liczba zmiennych brana pod uwagę w modelu;
- *min_samples_split* – minimalna liczba próbek wymagana do podziału węzła;

zdecydowano się wykorzystać bibliotekę Hyperopt (Bergstra i in., 2015). Celem tego procesu było minimalizowanie błędu średniokwadratowego (RMSE), będącego miarą dopasowania modelu do danych (Chicco i in., 2021). Podczas procesu doboru hiperparametrów wygenerowano 500 modeli Random Forest i wybrano najlepszy, czyli o najmniejszym błędzie RMSE.

Drugim wykonanym modelem uczenia maszynowego był XGBoost (Ramraj i in., 2016). Podobnie jak w przypadku modelu Random Forest, również tutaj wykorzy-

stano bibliotekę Hyperopt do doboru optymalnych hiperparametrów. Proces strojenia hiperparametrów obejmował przetestowanie 500 różnych wartości:

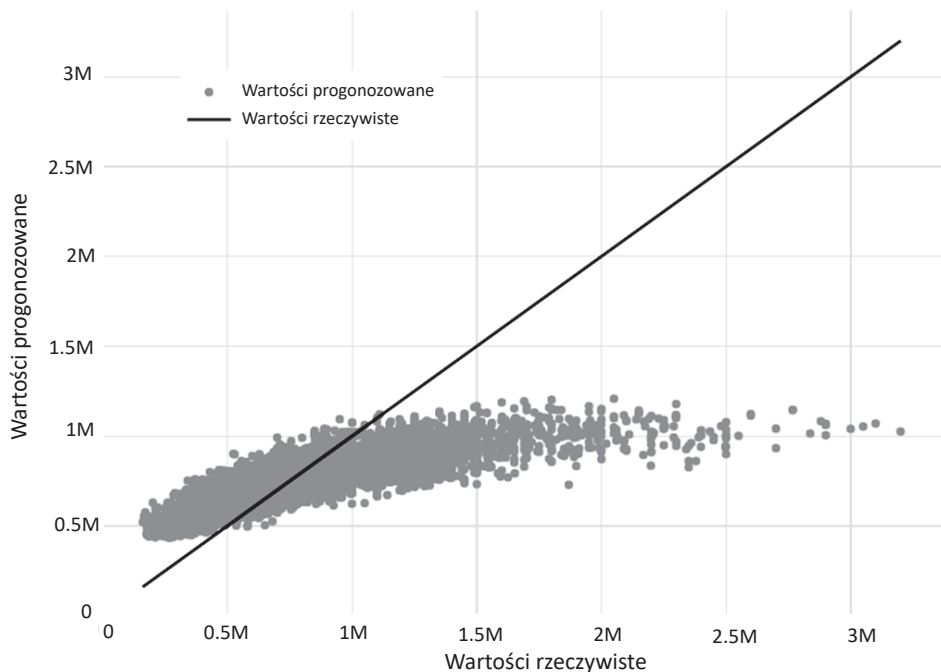
- *n_estimators* – liczba drzew decyzyjnych budowanych w sekwencji;
- *max_depth* – maksymalna głębokość każdego drzewa;
- *learning_rate* – szybkość, z jaką model się uczy;
- *gamma* – parametr, który kontroluje podział węzła podczas budowy drzewa;
- *min_child_weight* – minimalna suma wag obserwacji w danym węźle, aby ten węzeł mógł być podzielony;
- *subsample* – procent przypadków ze zbioru treningowego, który jest używany do trenowania każdego drzewa,
- *colsample_bytree* – kontroluje, ile zmiennych użyto do treningu każdego drzewa w modelu;
- *reg_alpha* – parametr regularyzacji L1 (Lasso). Dodanie składnika regularyzacji L1 pomaga modelowi bardziej skupić się na istotnych cechach;
- *reg_lambda* – parametr regularyzacji L2 (Ridge). Dodanie składnika regularyzacji L2 pomaga kontrolować wielkość wag, zapobiegając im nadmiernemu wzrostowi;

aby znaleźć optymalny zestaw dla modelu XGBoost. Spośród tych 500 modeli również wybrano ten, który osiągnął najmniejszą wartość błędu średniokwadratowego (RMSE).

4.4. Porównanie i ocena modeli uczenia maszynowego

Najlepszy model Random Forest osiągnął następujące wartości błędów: MAE wyniosło 113 032,47, co oznacza, że średnio przewidywana cena mieszkania różni się o około 113 032,47 złotych od rzeczywistej ceny; MAPE na poziomie 0,17 wskazuje, że średni względny błąd predykcji wynosi około 17%, co sugeruje, że model przewiduje ceny mieszkań z dokładnością na poziomie 83%; RMSE wyniosło 167 930,67; współczynnik determinacji osiągnął wartość 0,774, co sugeruje, że model stosunkowo dobrze dopasowuje się do danych.

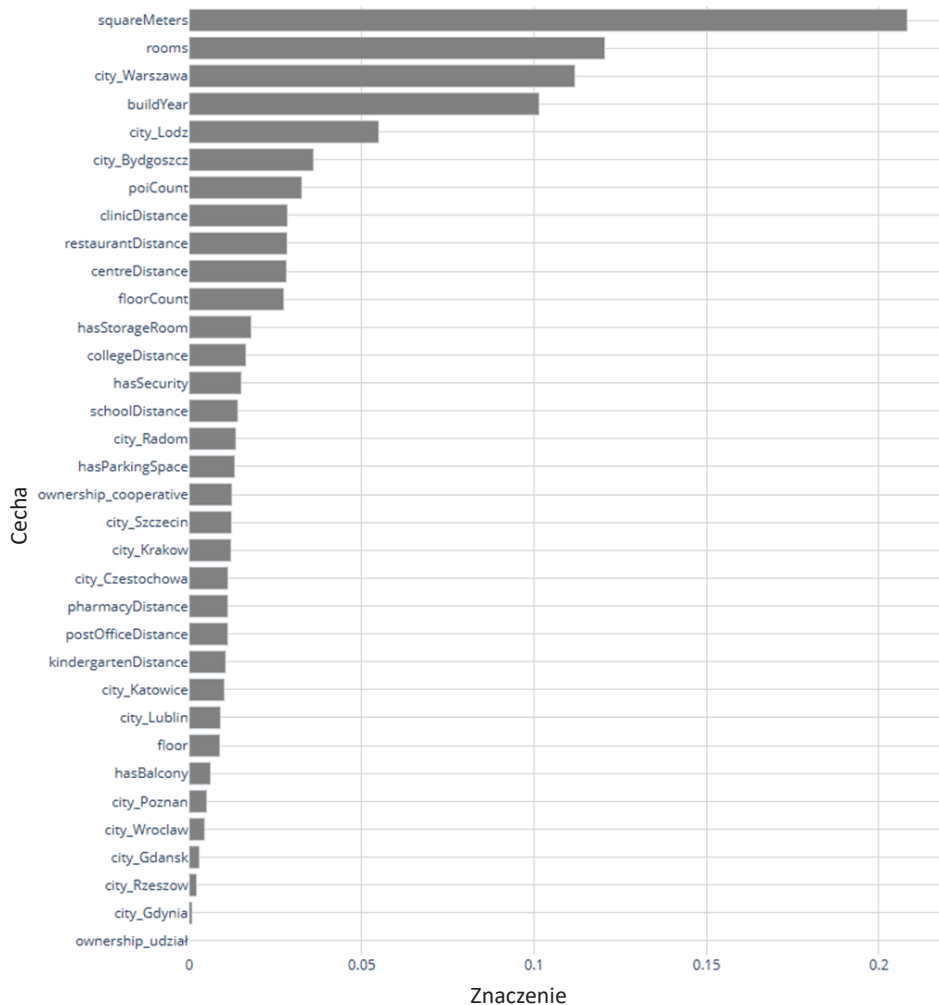
W celu jego dodatkowej oceny stworzono wykres pokazujący porównanie wartości rzeczywistych i prognozowanych (rys. 3). Idealnym dopasowaniem byłoby, gdyby punkty znajdowały się na prostej. Model dobrze przewiduje ceny mieszkań nie droższych niż 1,25 mln. W przypadku droższych mieszkań można zauważyć występowanie rozbieżności między wartościami przewidywanymi a rzeczywistością.



Rys. 3. Porównanie przewidywanych wartości modelu Random Forest z rzeczywistymi cenami mieszkań
 Źródło: opracowanie własne.

Na prognozie modelu ceny mieszkań największy wpływ miała zmienna dotycząca wielkości mieszkania (rys. 4): powierzchnia w metrach kwadratowych (20%). Duże znaczenie miała również liczba pokoi w mieszkaniu, informacja, czy mieszkanie znajduje się w Warszawie, oraz rok budowy (po około 10-12%).

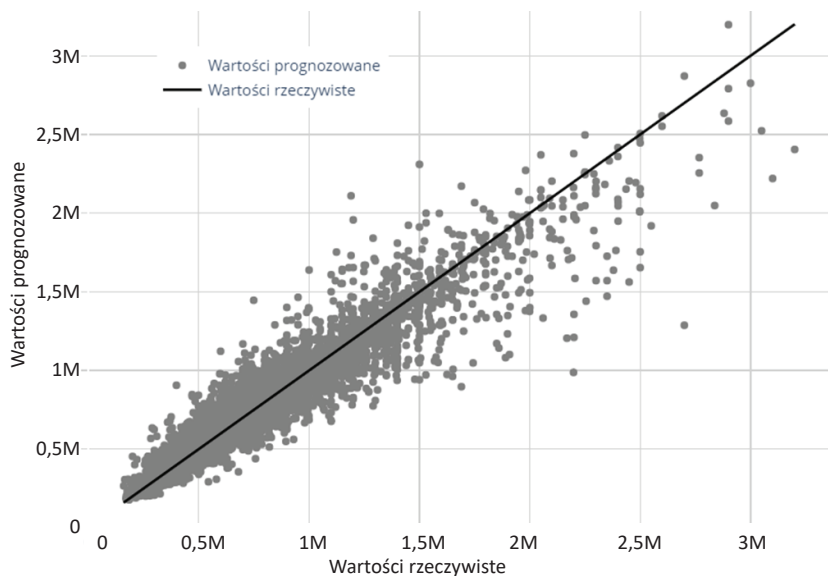
W przypadku modelu XGBoost, najlepszy z modeli osiągnął średni błąd bezwzględny (MAE) około 62 967,83, co oznacza, że przewidywane ceny mieszkań różnią się średnio o około 62 967 złotych od rzeczywistych wartości. Wartość błędów względnego (MAPE) wynosząca 0,09 wskazuje, że model przewiduje ceny z dużą dokładnością, błąd wynosi zaledwie 9%. Dodatkowo niska wartość RMSE (103 939,56) sugeruje, że model ma niską wariancję i dobrze dopasowuje się do danych treningowych. Warto również zauważyć wysoki wynik współczynnika determinacji na poziomie 0,913. Jest to znak, że model XGBoost dobrze odwzorowuje zależności między zmiennymi a cenami mieszkań, co potwierdza jego wysoką skuteczność w przewidywaniu cen mieszkań na podstawie analizowanych danych. Na rysunku 5 widać, że model dobrze przewiduje ceny mieszkań tańszych i gorszych, a punkty rozproszone są stosunkowo blisko prostej.



Rys. 4. Ważność zmiennych w modelu Random Forest

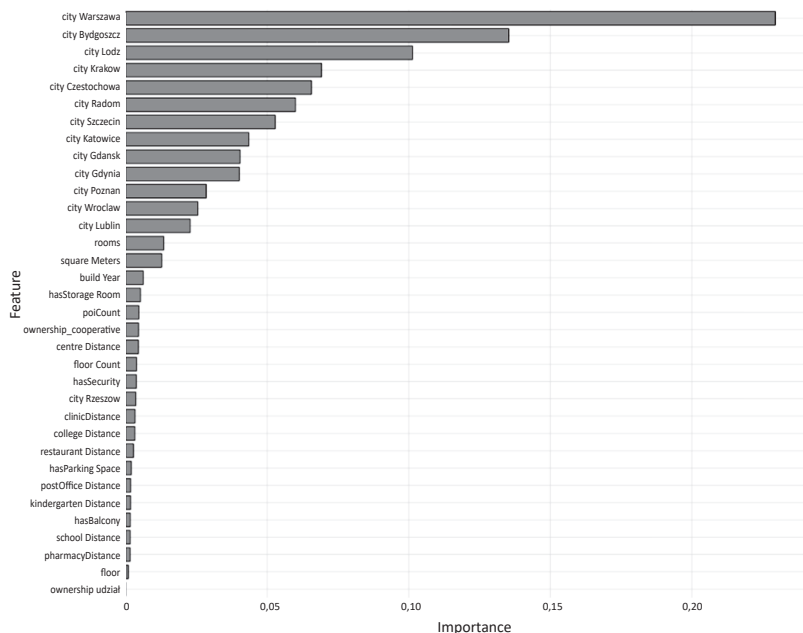
Źródło: opracowanie własne.

Największy wpływ na predykcje modelu miały zmienne dotyczące miasta, w którym znajduje się mieszkanie (rys. 6), w szczególności, czy mieszkanie znajduje się w Warszawie (23%), Bydgoszczy (13,5%) i Łodzi (10%). W znacznie mniejszym stopniu na prognozy modelu wpływała wielkość mieszkania, czyli powierzchnia w metrach kwadratowych i liczba pokoi (po około 2%).



Rys. 5. Porównanie przewidywanych wartości modelu XGBoost z rzeczywistymi cenami mieszkań

Źródło: opracowanie własne.



Rys. 6. Ważność cech w modelu XGBoost

Źródło: opracowanie własne.

Podsumowując, model XGBoost w porównaniu z Random Forest jest zdecydowanie lepiej dopasowany do danych, czyli jego przewidywania są dużo bardziej precyzyjne. Oznacza to, że na cenę mieszkań najbardziej wpływa miejscowość, w której się ono znajduje.

5. Zakończenie

Ewolucja polskiego rynku nieruchomości od lat 90. XX wieku do współczesności wyraża złożoność procesów społecznych, ekonomicznych i politycznych, które kształtują tę dziedzinę. W obliczu przemian demograficznych, zmieniających się preferencji mieszkańców oraz dynamicznych trendów rynkowych kluczowym wyzwaniem jest wycena mieszkań.

Przeprowadzone badanie dostarczyło istotnych wniosków dotyczących zastosowania metod uczenia maszynowego w prognozowaniu cen mieszkań w dużych miastach. Porównanie modeli Random Forest i XGBoost wykazało, że model XGBoost osiąga mniejsze błędy prognoz (RMSE na poziomie około 104 000 złotych) i wyższą wartość współczynnika determinacji (0,91). Model Random Forest, chociaż wykazał się RMSE wynoszącym około 168 000 złotych, dobrze prognozował ceny dla segmentu mieszkań o wartości poniżej miliona złotych, mając trudności z przewidywaniem cen wyższych, czyli mieszkań dużych lub luksusowych.

Istotnymi czynnikami wpływającymi na predykcje ceny mieszkań okazała się lokalizacja, zwłaszcza miasto, w którym znajduje się nieruchomość. Oprócz tego przeprowadzone wizualizacje potwierdziły dobrą jakość predykcji modelu XGBoost, który skutecznie przewidywał ceny mieszkań zarówno w niższym, jak i wyższym przedziale cenowym.

Literatura

- Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., i Cox, D. D. (2015). Hyperopt: A Python Library for Model Selection and Hyperparameter Optimization. *Computational Science & Discovery*, 8(1). <https://doi.org/10.1088/1749-4699/8/1/014008>
- Bryx, M. (2006). *Rynek nieruchomości system i funkcjonowanie*. Wydawnictwo Poltex.
- Chicco, D., Warrens, M. J., i Jurman, G. (2021). The Coefficient of Determination R-squared is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, (7). <https://doi.org/10.7717/peerj-cs.623>
- FIN-HUB. (b.d.). *Bezpieczny Kredyt 2%*. Pobrano z <https://finhub.pl/slownik/bezpieczny-kredyt-2>
- Gołąbeska, E. (2007). *Rynek nieruchomości i jego podmioty*. Wydawnictwo Wyższej Szkoły Finansów i Zarządzania w Białymstoku.
- Gołąbeska, E. (2017). Współczesne trendy na rynku nieruchomości mieszkaniowych. W: E. Broniewicz (red.), *Gospodarowanie przestrzeni w warunkach rozwoju zrównoważonego*, (s. 85-106). Oficyna Wydawnicza Politechniki Białostockiej.

- Jaiswal, J. K., i Samikannu, R. (2017). Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. *Conference: 2017 World Congress on Computing and Communication Technologies (WCCCT)*, 65-68. <https://doi.org/10.1109/WCCCT.2016.25>
- Jamroz, K. (2024). *Apartment Prices in Poland*. Pobrano z <https://www.kaggle.com/datasets/krzysztofjamroz/apartment-prices-in-poland?resource=download>
- Kucharska-Stasiak, E. (2000). *Nieruchomość a rynek*. Wydawnictwo Naukowe PWN.
- Kucharska-Stasiak, E., Załączna M. i Żelazowski K. (2012). *Wpływ procesu integracji Polski z Unią Europejską na rozwój rynków nieruchomości*. Wydawnictwo Uniwersytetu Łódzkiego.
- Majorek, M. (2013). Rządowy program „Rodzina na swoim”. Społeczne, ekonomiczne i medialne reperkusje wdrażania. *Państwo i Społeczeństwo*, 13(3), 11-30.
- Malarvizhi, R., i Thanamani, A. S. (2012). K-nearest Neighbor in Missing Data Imputation. *International Journal of Engineering Research and Development*, 5(1), 5-7.
- Ramraj, S., Uzir, N., Sunil, R., i Banerjee, S. (2016). Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets. *International Journal of Control Theory and Applications*, 9(40), 651-662.
- Sadowski, W. (1976). *Mała Encyklopedia Statystyki*. Państwowe Wydawnictwo Ekonomiczne Warszawa.
- Samorek, B. i Cichocki, M. (2023). *Polski rynek nieruchomości mieszkaniowych*. <https://doi.org/10.51733/opm.2023.03>
- Westerlund, O. (2023). *Cluster Analysis for Outlier Detection. A Case Study of Applying Unsupervised Machine Learning on Diesel Engine Data*. University of Turku. Department of Computing, Faculty of Technology. Pobrano z https://www.utupub.fi/bitstream/handle/10024/174378/Westerlund_Otto_opinnayte.pdf?sequence=1&isAllowed=y

Predicting Housing Prices in Major Cities in Poland Using Machine Learning

Abstract: The article provides an in-depth analysis of the real estate market in major cities in Poland, with a primary focus on predicting housing prices. It discusses the concept of the real estate market and its key features, explaining in detail how housing prices have changed over the years in Poland. In the 1990s, housing prices saw an increase due to liberalisation and easy access to mortgage loans. However, after the financial crisis of 2007, the market experienced significant fluctuations. Compared to the 2% Safe Mortgage programme, the COVID-19 pandemic had a minor impact on prices. Analysing housing prices, the article describes data preparation, imputing missing values, and identifying outliers. Machine learning algorithms, such as Random Forest and XGBoost, were used to forecast prices. These models demonstrated high effectiveness (especially XGBoost). The article analyses key variables influencing price formation in the real estate market in major cities in Poland.

Keywords: real estate market, housing prices, machine learning, Random Forest model, XGBoost model

Dominik Głowacki

e-mail: 177530@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Analiza wyceny piłkarzy za pomocą modelowania ekonometrycznego

DOI: 10.15611/2024.53.6.06

JEL Classification: C12, C51, C52, Z23

© 2024 Dominik Głowacki

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Głowacki, D. (2024). Analiza wyceny piłkarzy za pomocą modelowania ekonometrycznego. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 82-98). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: W artykule podjęto tematykę związaną z piłką nożną, czyli najpopularniejszym sportem na świecie. Główną osią badania była analiza wycen piłkarzy. W pierwszej części przedstawiono ogólne informacje na temat wycen piłkarzy, w tym najdroższe transfery w piłce nożnej. Druga część poświęcona została opisowi metody badawczej – w tym przypadku był to model ekonometryczny i jego weryfikacja. Z kolei w trzeciej części przedstawiono analizę danych, która przeprowadzona została w programie MS Excel. Celem badania było wskazanie zmiennych najsilniej wpływających na wartość piłkarza oraz stworzenie modelu ekonometrycznego pomagającego w wycenie zawodnika. Po zweryfikowaniu współczynnika determinacji, a także po przeprowadzeniu testów normalności i homoskedastyczności wektora reszt, stwierdzono, że pomimo niespełnienia założeń, osiągnięto cel, jakim było wytypowanie zmiennych mających istotny wpływ na wycenę zawodnika. W końcowej części artykułu zaproponowano możliwe działania, które poprawiłyby model pod względem dopasowania do danych oraz spełnienia założeń.

Słowa kluczowe: wycena, modelowanie ekonometryczne, piłka nożna

1. Wstęp

Bez wątpienia piłka nożna jest jednym z najpopularniejszych sportów w Polsce i na świecie. Zwłaszcza w ostatnim czasie, wobec rosnącej popularności sztucznej inteligencji oraz dzięki rozwojowi zaawansowanych metod badawczych, pojawiło się wiele opracowań dotyczących pozyskiwania danych w sporcie, w tym w piłce nożnej. Sumpter (2022, s. 181) wskazuje, że podczas meczu piłkarskiego rejestruje się ponad 200 milionów liczb. Z kolei Biermann (2022, s. 15) stwierdza, że „piłka nożna stała się grą liczb”. Pomimo tego piłka nożna wciąż uważana jest za dyscyplinę sto-

sunkowo nieskomplikowaną, a jej amatorskie uprawianie nie niesie za sobą znacznych nakładów finansowych jak w przypadku tenisa ziemnego czy sportów zimowych. Nie oznacza to jednak, że piłka nożna nie generuje znacznych pieniędzy, a beneficjentami są zarówno federacje piłkarskie, kluby, trenerzy i zawodnicy. Kontrakty najlepszych piłkarzy na świecie są absolutnie nieporównywalne z zarobkami nawet dobrze opłacanych zawodów. W roku 2023 najlepiej zarabiającym piłkarzem w polskiej ekstraklasie był zawodnik Lecha Poznań, który pobierał niemal 400 tysięcy złotych miesięcznie (Tubacki, 2023). Nierozłączną częścią piłki nożnej są transfery zawodników, czyli kupowanie i sprzedawanie piłkarzy przez kluby piłkarskie. Dotychczas (marzec 2024) najdroższym piłkarzem w historii był Brazylijczyk Neymar da Silva Santos Júnior, który w 2017 roku został kupiony przez Paris Saint-Germain z klubu FC Barcelona za 222 miliony euro. W przypadku polskiej ekstraklasy, czyli najwyższej klasy rozgrywkowej w Polsce, kwoty transferów są zdecydowanie niższe. Najdrożej sprzedanymi piłkarzami z ekstraklasy byli Kacper Kozłowski (w 2022 roku) i Jakub Moder (w 2020 roku). Obaj zostali kupieni przez angielski klub Brighton & Hove Albion za 11 milionów euro, odpowiednio z Pogoni Szczecin i Lecha Poznań. Znacznie mniejszą kwotą transferu był najdrożej kupiony piłkarz przez klub ekstraklasy. W lipcu 2023 roku do Lecha Poznań wytransferowany został Irańczyk Ali Gholizadeh. Klub z Poznania zapłacił belgijskiemu Charleroi S.C. około 1,8 miliona euro. Łatwo zauważalna jest ogromna dysproporcja pomiędzy tym, ile wydać mogą kluby z najlepszych lig świata, zwłaszcza gdy tak jak w przypadku Paris Saint-Germain są zarządzane przez katarskich biznesmanów, a tym, ile mogą zapłacić kluby z Polski. Przedmiotem niniejszego artykułu nie będzie jednak badanie tej dysproporcji, a próba sprawdzenia, skąd biorą się wyceny piłkarzy. Jako podstawowe źródło wiedzy zarówno do danych historycznych, jak i do weryfikacji statystyk posłuży portal Transfermarkt. Analizie poddane zostaną wyceny napastników (piłkarzy, którzy są tak klasyfikowani przez Transfermarkt) z ekstraklasy na dzień 31 stycznia 2024 roku.

2. Metodyka badań

W niniejszym rozdziale zaprezentowane zostaną teoretyczne aspekty dotyczące miar, metod i testów statystycznych, które zostaną zastosowane w analizie. Biorąc pod uwagę specyfikę badanego zjawiska, czyli szacowanie wartości piłkarzy, postanowiono, że dobór zmiennych objaśniających nastąpi na podstawie subiektywnej oceny autora po uwzględnieniu korelacjami między poszczególnymi zmiennymi niezależnymi.

Jednym z podstawowych sposobów badania zależności pomiędzy cechami jest współczynnik korelacji liniowej Pearsona o wzorze (Ostasiewicz i in., 2006):

$$r_{xy} = \frac{C(X, Y)}{s_x s_y} \quad (1)$$

gdzie: s_x – odchylenie standardowe zmiennej X , s_y – odchylenie standardowe zmiennej Y , $C(X,Y)$ kowariancja między cechami X i Y .

Na podstawie współczynników korelacji liniowej Pearsona utworzona zostanie macierz korelacji. Głównie na jej podstawie zostaną wytypowane zmienne, które będą tworzyły liniowy model ekonometryczny. Ogólna postać liniowego modelu ekonometrycznego przedstawiona jest poniżej (Dziechciarz, 2003).

$$y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_m + \varepsilon \quad (2)$$

Oznaczone symbolami $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$ są parametrami strukturalnymi modelu i należy je wyestymować, natomiast ε to składnik losowy. Estymacja parametrów modelu ekonometrycznego zostanie przeprowadzona za pomocą klasycznej metody najmniejszych kwadratów. Postępowanie w kolejnych krokach od budowy modelu aż po wnioskowanie przeprowadzono, opierając się na opracowaniach naukowych (Dziechciarz, 2003; Bobowski, 2004; Gładysz i Mercik, 2007).

Jednym z podstawowych kryteriów oceny dobroci dopasowania modelu do danych empirycznych jest współczynnik determinacji, który oblicza się wzorem:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (3)$$

Współczynnik ten przyjmuje wartości z przedziału od zera do jedności. Gdy R^2 osiąga wartość jeden, oznacza to idealne dopasowanie skonstruowanego modelu do danych, a co za tym idzie pełne wyjaśnienie przez model zmienności zmiennej zależnej. Dobrze skonstruowany model powinien osiągać wartość R^2 bliską jedności, należy jednak pamiętać, że zbyt dobre dopasowanie może być przejawem tzw. przeuczenia (Kufel i Lewandowski, 2023). Nie jest to zjawisko korzystne ze względu na możliwe trudności z zastosowaniem modelu do wnioskowania lub prognozowania. Po wstępnym zweryfikowaniu modelu poprzez wartość współczynnika determinacji, należy przystąpić do wykonania testu istotności statystycznej wektora parametrów strukturalnych (test Fishera). Hipoteza zerowa w tym teście zakłada, że cały wektor parametrów strukturalnych modelu jest nieistotny statystycznie. Natomiast hipoteza alternatywna w tym teście mówi, że istnieje przynajmniej jeden z parametrów strukturalnych, który jest istotnie różny od zera. Statystyka testowa:

$$F = \frac{n-m-1}{m} \cdot \frac{R^2}{1-R^2} \quad (4)$$

ma rozkład F-Snedecora z $n_1 = m$ oraz $n_2 = m - 1$ stopniami swobody, gdzie n to liczba obserwacji, a m to liczba zmiennych objaśniających (Dziechciarz, 2003). W celu zweryfikowania istotności statystycznej poszczególnych parametrów należy zastosować test Studenta. W hipotezie zerowej zakłada się, że badany parametr jest nieistotnie różny od zera, natomiast hipoteza alternatywna jest tego zaprzecze-

niem. Dla każdego kolejnego parametru strukturalnego statystyka testowa wygląda następująco (Górecki, 2010):

$$t_i = \frac{|a_i|}{S(a_i)} \quad (5)$$

i ma rozkład *t*-Studenta z $n - m - 1$ stopniami swobody, gdzie a_i to ocena parametru, natomiast $S(a_i)$ to standardowy błąd szacunku parametrów.

Testem na zgodność z rozkładem normalnym wektora reszt, który zostanie zastosowany w niniejszym artykule, jest powszechnie znany test Shapiro-Wilka. Hipoteza zerowa zakłada normalność wektora reszt, natomiast hipoteza alternatywna jest tego zaprzeczeniem. Aby przeprowadzić test, należy uporządkować reszty w sposób niemalejący, a następnie odejmować od siebie zgodnie z regułą „ostatni minus pierwszy, przedostatni minus drugi” itd. Następnie otrzymane w ten sposób różnice trzeba przemnożyć przez odpowiednie współczynniki z tablic. Statystyka testowa ma postać (Dębowska-Mróż i in., 2018):

$$W = \frac{\left[\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} a_{n,i} * (e_{(n-i+1)} - e_{(i)}) \right]^2}{\sum_{i=1}^n (e_i - \bar{e})^2} \quad (6)$$

Z tablic testu Shapiro-Wilka należy odczytać wartość krytyczną, a następnie zestawić ją z obliczoną wcześniej statystyką testową. Jeśli wartość statystyki testowej będzie większa niż wartość krytyczna, oznacza to, że nie ma podstaw do odrzucenia hipotezy zerowej, zatem reszty mają rozkład normalny (Dębowska-Mróż i in., 2018).

Ostatnim, niezbędnym do weryfikacji modelu, testem jest sprawdzenie homoskedastyczności składnika resztowego. Zastosowany zostanie w tym celu test Goldfelda-Quandt. Aby przeprowadzić test, należy podzielić obserwacje na dwie grupy, a następnie zbadać, czy wartości wariancji w otrzymanych podpopulacjach są zbliżone do siebie. Liczebności grup powinny być do siebie zbliżone. Podział następuje przeważnie arbitralnie, jednak należy uprzednio przeanalizować wektor reszt (Dziechciarz, 2003). Hipoteza zerowa mówi o tym, że wariancje reszt w obu podpopulacjach nie różnią się od siebie, natomiast hipoteza alternatywna jest negacją tego stwierdzenia. Statystyką testową jest iloraz obu wariancji, przy czym wartość większa znajduje się w liczniku. Wartość krytyczną testu należy wyznaczyć z rozkładu *F* – Snedecora z parą stopni swobody: $n_1 - m - 1$ oraz $n_2 - m - 1$. Gdy wartość statystyki testowej jest większa niż wartość krytyczna, należy odrzucić hipotezę o równości wariancji reszt, co oznacza występowanie zjawiska heteroskedastyczności (Dziechciarz, 2003). W związku z tym, że w wielu artykułach naukowych uważa się właśnie normalność i homoskedastyczność składnika resztowego za kluczowe dla poprawności modelu, postanowiono skupić się jedynie na tych dwóch testach, a ich wyniki uznać za najistotniejsze w przypadku oceny jakości modelu.

3. Wyniki badań

Analiza dotyczyła 50 najwyżej wycenianych napastników według portalu Transfermarkt. Imiona, nazwiska i kluby, które reprezentowali piłkarze, zostały zaprezentowane w tabeli 1.

Tabela 1. 50 najwyżej wycenianych piłkarzy ofensywnych (napastników) PKO BP Ekstraklasy

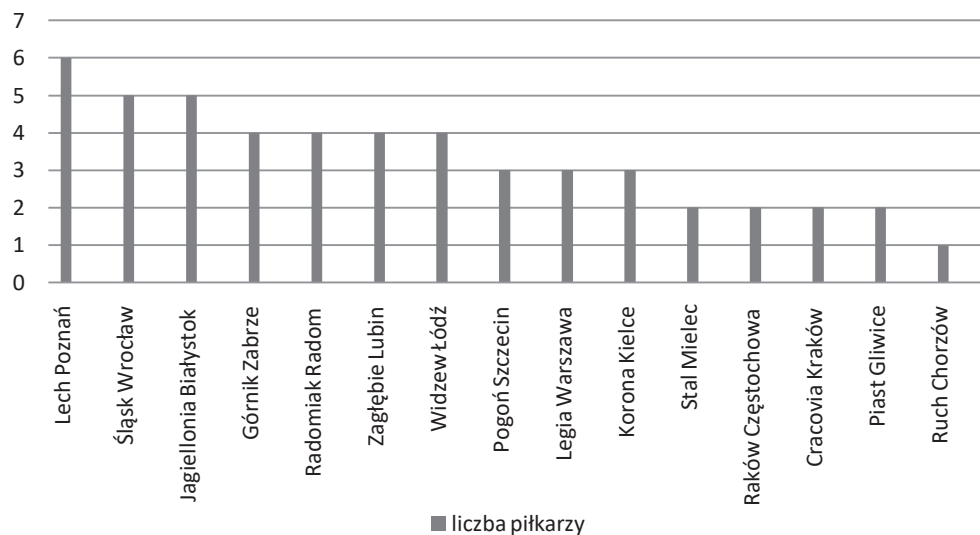
| Imię | Nazwisko | Klub |
|------------|--------------|-----------------------|
| 1 | 2 | 3 |
| Kristoffer | Velde | Lech Poznań |
| Filip | Szymczak | Lech Poznań |
| Erik | Exposito | Śląsk Wrocław |
| Patryk | Klimala | Śląsk Wrocław |
| Mikael | Ishak | Lech Poznań |
| Dino | Hotic | Lech Poznań |
| Ali | Gholizadeh | Lech Poznań |
| Efthymios | Koulouris | Pogoń Szczecin |
| Dominik | Marczuk | Jagiellonia Białystok |
| Marc | Gual | Legia Warszawa |
| Adrian | Kapralik | Górnik Zabrze |
| Pedro | Henrique | Radomiak Radom |
| Vahan | Bichakhchyan | Pogoń Szczecin |
| Ilya | Shkurin | Stal Mielec |
| Ante | Crnac | Raków Częstochowa |
| Afimico | Pululu | Jagiellonia Białystok |
| Kacper | Chodyna | Zagłębie Lubin |
| Jose | Naranjo | Jagiellonia Białystok |
| Piotr | Samiec-Talar | Śląsk Wrocław |
| Benjamin | Kallman | Cracovia Kraków |
| Adriel | Ba Loua | Lech Poznań |
| Michael | Ameyaw | Piast Gliwice |
| Mateusz | Wdowiak | Zagłębie Lubin |
| Vagner | Goncalves | Radomiak Radom |
| Maciej | Rosółek | Legia Warszawa |
| Burak | Ince | Śląsk Wrocław |
| Lisandro | Semedo | Radomiak Radom |
| Blaz | Kramer | Legia Warszawa |
| Dawid | Kurminowski | Zagłębie Lubin |
| Kristoffer | Hansen | Jagiellonia Białystok |
| Łukasz | Zwoliński | Raków Częstochowa |
| Jordi | Sanchez | Widzew Łódź |
| Daniel | Szczepan | Ruch Chorzów |
| Mateusz | Żukowski | Śląsk Wrocław |
| Kai | Meriluoto | Stal Mielec |

Tabela 1. cd.

| 1 | 2 | 3 |
|------------|-------------|-----------------------|
| Fabian | Piasecki | Piast Gliwice |
| Bartłomiej | Pawłowski | Widzew Łódź |
| Mariusz | Fornalczyk | Korona Kielce |
| Jakub | Łukowski | Korona Kielce |
| Soichiro | Kozuki | Górnik Zabrze |
| Kamil | Grosicki | Pogoń Szczecin |
| Kaan | Caliskaner | Jagiellonia Białystok |
| Kacper | Śmiglewski | Cracovia Kraków |
| Kamil | Lukoszek | Górnik Zabrze |
| Ernest | Terpiłowski | Widzew Łódź |
| Antoni | Klimek | Widzew Łódź |
| Sebastian | Musiolik | Górnik Zabrze |
| Leonardo | Rocha | Radomiak Radom |
| Juan | Munoz | Zagłębie Lubin |
| Adrian | Dalmau | Korona Kielce |

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

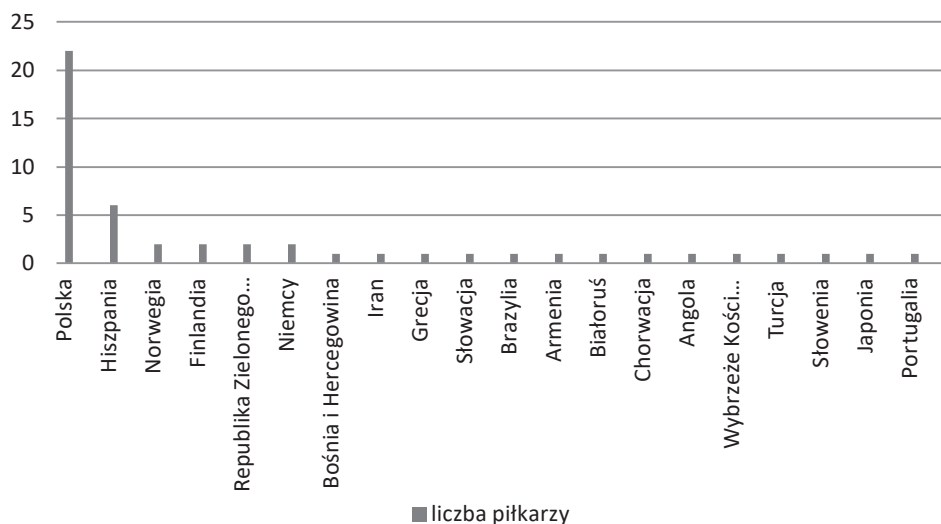
Spśród piłkarzy przedstawionych w tabeli 1, najwięcej reprezentuje Lecha Poznań (6 piłkarzy), a w dalszej kolejności Śląsk Wrocław i Jagiellonię Białystok (po 5 piłkarzy). Szczegółowe informacje co do liczby piłkarzy z poszczególnych klubów przedstawia rysunek 1.



Rys. 1. Liczba najwyżej wycenianych piłkarzy w podziale na kluby

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Jak wspomniano wcześniej, najliczniejszą grupę stanowią piłkarze z poznańskiego klubu. Wartym odnotowania jest fakt, iż spośród 18 zespołów grających w najwyższej klasie rozgrywkowej w Polsce, wśród 50 najwyżżej wycenianych napastników swojego przedstawiciela nie mają trzy kluby. Są to: Puszcza Niepołomice, ŁKS Łódź i Warta Poznań. Ponadto warto przedstawić podział według narodowości piłkarzy, który zaprezentowano na rysunku 2.



Rys. 2. Liczba najwyżżej wycenianych piłkarzy w podziale na narodowości

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Zdecydowanie najwięcej piłkarzy spośród wybranych do badania pochodzi z Polski (22 piłkarzy). Oprócz Polski licznie reprezentowanymi nacjami są: Hiszpania (6 piłkarzy) i Norwegia, Finlandia i Republika Zielonego Przylądka (po 2 piłkarzy).

Po ogólnym przedstawieniu podstawowych informacji o piłkarzach, zaprezentowane zostaną dane, które posłużą do modelowania ekonometrycznego. W celu zachowania czytelności danych, zostaną one przedstawione w dwóch częściach w tabelach 2 i 3. Oprócz omówionych wcześniej informacji o klubie, w którym gra dany zawodnik oraz jego narodowości, w tabelach 2 i 3 zawarto wiek zawodnika, długość kontraktu z obecnym klubem (wyrażoną w latach), informację, czy zawodnik jest aktualnym reprezentantem kraju (tylko reprezentacje seniorskie), procent rozegranych minut w ekstraklasie w stosunku do łącznej możliwej liczby minut z rundy jesiennej sezonu 2023/2024 (od 1 lipca do 31 grudnia 2023), liczbę strzelonych goli, liczbę asyst, miejsce drużyny na koniec rundy jesiennej sezonu 2023/2024 (według oficjalnej strony PKO BP Ekstraklasy) oraz wartość zawodnika wyrażoną w milionach euro.

W tabeli 2 zaprezentowano 25 najwyżżej wycenianych napastników grających w polskiej ekstraklasie (stan na 31 stycznia 2024).

Tabela 2. Dane do analizy – piłkarze ofensywni z miejsc 1-25 w rankingu

| Imię | Nazwisko | Klub | Narodowość | Wiek | Pozostały kontrakt (lata) | Czy reprezentant | % minut | Gole | Asysty | Miejsce drużyny | Wartość (mln euro) |
|------------|--------------|-----------------------|-------------------------------|------|---------------------------|------------------|---------|------|--------|-----------------|--------------------|
| Kristoffer | Velde | Lech Poznań | Norwegia | 25 | 2,5 | TAK | 90 | 7 | 4 | 3 | 6,0 |
| Filip | Szymczak | Lech Poznań | Polaka | 21 | 3,5 | NIE | 29 | 1 | 2 | 3 | 3,0 |
| Erik | Exposito | Śląsk Wrocław | Hiszpania | 27 | 0,5 | NIE | 100 | 14 | 3 | 1 | 2,5 |
| Patryk | Klimala | Śląsk Wrocław | Polaka | 25 | 3,5 | NIE | 9 | 1 | 0 | 1 | 1,8 |
| Miakael | Ishak | Lech Poznań | Szwecja | 30 | 1,5 | TAK | 62 | 6 | 3 | 3 | 1,5 |
| Dino | Hotic | Lech Poznań | Bośnia i Hercegowina | 28 | 2,5 | TAK | 27 | 0 | 2 | 3 | 1,5 |
| Ali | Gholizadeh | Lech Poznań | Iran | 27 | 2,5 | TAK | 10 | 0 | 0 | 3 | 1,5 |
| Eftymios | Koulouris | Pogoń Szczecin | Grecja | 27 | 2,5 | TAK | 91 | 8 | 1 | 6 | 1,5 |
| Dominik | Marczuk | Jagiellonia Białystok | Polaka | 20 | 2,5 | NIE | 72 | 2 | 9 | 2 | 1,5 |
| Marc | Gual | Legia Warszawa | Hiszpania | 27 | 2,5 | NIE | 47 | 1 | 2 | 5 | 1,3 |
| Adrian | Kapralik | Górnik Zabrze | Słowacja | 21 | 0,5 | TAK | 64 | 2 | 1 | 7 | 1,2 |
| Pedro | Henrique | Radomiak Radom | Brazylia | 27 | 2,5 | NIE | 95 | 8 | 4 | 10 | 1,2 |
| Vahan | Bichakhchyan | Pogoń Szczecin | Armenia | 24 | 1,5 | TAK | 75 | 4 | 2 | 6 | 1,2 |
| Ilya | Shkurin | Stal Mielec | Białoruś | 24 | 0,5 | NIE | 84 | 8 | 5 | 12 | 1,2 |
| Ante | Crnac | Raków Częstochowa | Chorwacja | 20 | 4,5 | NIE | 43 | 2 | 2 | 4 | 1 |
| Afimico | Pululu | Jagiellonia Białystok | Angola | 24 | 1,5 | NIE | 55 | 6 | 2 | 2 | 0,9 |
| Kacper | Chodyna | Zagłębie Lubin | Polaka | 24 | 1,5 | NIE | 76 | 3 | 5 | 8 | 0,9 |
| Jose | Naranjo | Jagiellonia Białystok | Hiszpania | 29 | 1,5 | NIE | 75 | 6 | 1 | 2 | 0,85 |
| Piotr | Samiec-Talar | Śląsk Wrocław | Polaka | 22 | 0,5 | NIE | 57 | 4 | 5 | 1 | 0,8 |
| Benjamin | Kallman | Cracovia Kraków | Finlandia | 25 | 1,5 | TAK | 83 | 5 | 4 | 13 | 0,7 |
| Adriel | Ba Loua | Lech Poznań | Wybrzeże Kości Słoniowej | 27 | 1,5 | NIE | 46 | 4 | 2 | 3 | 0,7 |
| Michael | Ameyaw | Piast Gliwice | Polaka | 23 | 2,5 | NIE | 92 | 2 | 3 | 9 | 0,65 |
| Mateusz | Wdowiak | Zagłębie Lubin | Polaka | 27 | 2,5 | NIE | 56 | 4 | 1 | 8 | 0,65 |
| Vagner | Goncalves | Radomiak Radom | Republika Zielonego Przylądka | 28 | 2,5 | TAK | 0 | 0 | 0 | 10 | 0,6 |
| Maciej | Rosolek | Legia Warszawa | Polaka | 22 | 1,5 | NIE | 36 | 0 | 2 | 5 | 0,6 |

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

W tabeli 3 przedstawiono napastników z miejsc 26-50 według wycen z portalu Transfermarkt.

Aby możliwa była analiza powyższych danych, zwłaszcza stworzenie modelu ekonometrycznego, dokonano przekodowania zmiennych opisujących narodowość piłkarzy oraz ich uczestnictwo w reprezentacjach narodowych. W pierwszym przypadku „1” oznaczono piłkarzy z Polski, natomiast „0” piłkarzy posiadających obywatelstwo innego kraju. Natomiast w drugiej zmiennej zerowej „1” oznaczono zawodników będących reprezentantami swoich krajów, natomiast „0” piłkarzy, którzy nie zagraли w seniorskich reprezentacjach. W tabelach 4 i 5 zaprezentowano dane do analizy po przekodowaniu dwóch zmiennych.

Na podstawie danych zaprezentowanych w tabelach 4 i 5 stworzono macierz korelacji (tab. 6). Na podstawie przedstawionej w tabeli 6 macierzy korelacji można stwierdzić, że zdecydowana większość wartości współczynników korelacji liniowej Pearsona jest, co do wartości bezwzględnej, niższa niż 0,4, co oznacza brak lub niską zależność liniową pomiędzy zmiennymi. Wyjątek stanowią wartości dla zmiennej opisującej procent rozegranych minut oraz liczbę goli, a także liczbę asyst. Wartości współczynnika korelacji liniowej Pearsona wynoszą w tych przypadkach odpowiednio 0,637 i 0,75. Można jednak stwierdzić, że zależność ta nie jest żadnym zaskoczeniem, gdyż logicznym wydaje się, że jeśli piłkarz grający w ofensywie przebywa dłużej na boisku, to powinien w tym czasie zanotować więcej goli i asyst.

Wobec niskich lub bardzo niskich wartości współczynników korelacji liniowej Pearsona postanowiono stworzyć model ze wszystkimi zmiennymi objaśniającymi, gdzie zmienną zależną jest zmienna wartość. Dla takiego modelu wartość statystyki Fishera wynosi 3,53, natomiast p-value 0,003. Na poziomie istotności $\alpha = 0,1$ oznacza to, że wektor parametrów strukturalnych modelu jest istotny statystycznie. Następnie przeprowadzono test istotności poszczególnych parametrów modelu – test Studenta. Wartości statystyki t i p-value dla każdej zmiennej są zamieszczone w tabeli 7.

W tabeli 7 kursywą zaznaczono te zmienne, dla których p-value wynosi mniej niż 0,1, a co za tym idzie, są one istotne statystycznie. W przypadku doboru zmiennych do modelu zdecydowano, aby przyjąć poziom istotności $\alpha = 0,1$, a nie stosowany przeważnie $\alpha = 0,05$ po to, aby do modelu włączyć zmienną *gole*, która w przypadku piłkarzy ofensywnych wydaje się być ważna. Ponowna estymacja modelu zostanie zatem przeprowadzona dla zmiennych objaśniających, które opisują: wiek, pozostały czas kontraktu, granie w reprezentacji narodowej, liczbę goli strzelonych w rundzie jesiennej PKO BP Ekstraklasy oraz miejsce drużyny w tabeli. Ponadto w modelu znajdzie się *wyraz wolny*. Przy powtórznym oszacowaniu parametrów modelu, wszystkie zmienne okazały się istotne statystycznie (na poziomie istotności $\alpha = 0,05$), co przedstawia tabela 8.

Tabela 3. Dane do analizy – piłkarze ofensywni z miejsc 26-50 w rankingu

| Imię | Nazwisko | Klub | Narodowość | Wiek | Pozostały kontrakt (lata) | Czy reprezentant | % minut | Gole | Asysty | Miejsce drużyny | Wartość (mln euro) |
|------------|-------------|-----------------------|-------------------------------|------|---------------------------|------------------|---------|------|--------|-----------------|--------------------|
| Burak | Ince | Śląsk Wrocław | Turcja | 20 | 1,5 | NIE | 22 | 2 | 1 | 1 | 0,6 |
| Lisandro | Semedo | Radomiak Radom | Republika Zielonego Przylądka | 27 | 0,5 | TAK | 61 | 2 | 2 | 10 | 0,6 |
| Blaž | Kramer | Legia Warszawa | Słowenia | 27 | 1,5 | TAK | 21 | 2 | 0 | 5 | 0,6 |
| Dawid | Kurminowski | Zagłębie Lubin | Polska | 24 | 1,5 | NIE | 63 | 3 | 0 | 8 | 0,55 |
| Kristoffer | Hansen | Jagiellonia Białystok | Norwegia | 29 | 1,5 | NIE | 39 | 5 | 1 | 2 | 0,55 |
| Łukasz | Zwoliński | Raków Częstochowa | Polska | 30 | 2,5 | NIE | 42 | 4 | 0 | 4 | 0,55 |
| Jordi | Sanchez | Widzew Łódź | Hiszpania | 29 | 0,5 | NIE | 81 | 4 | 3 | 11 | 0,55 |
| Daniel | Szczepan | Ruch Chorzów | Polska | 28 | 1,5 | NIE | 84 | 7 | 1 | 17 | 0,55 |
| Mateusz | Żukowski | Śląsk Wrocław | Polska | 22 | 0,5 | NIE | 54 | 0 | 0 | 1 | 0,5 |
| Kai | Meriluo | Stal Mielec | Finlandia | 21 | 0,5 | TAK | 17 | 2 | 0 | 12 | 0,5 |
| Fabian | Piasecki | Piast Gliwice | Polska | 28 | 3 | NIE | 29 | 0 | 0 | 9 | 0,5 |
| Bartłomiej | Pawłowski | Widzew Łódź | Polska | 31 | 2,5 | NIE | 67 | 4 | 3 | 11 | 0,5 |
| Mariusz | Formalczak | Korona Kielce | Polska | 21 | 2,5 | NIE | 14 | 2 | 0 | 16 | 0,5 |
| Jakub | Łukowski | Korona Kielce | Polska | 27 | 0,5 | NIE | 13 | 0 | 0 | 16 | 0,5 |
| Soichiro | Kozuki | Górnik Zabrze | Japonia | 23 | 1 | NIE | 10 | 0 | 0 | 7 | 0,5 |
| Kamil | Grosicki | Pogoń Szczecin | Polska | 35 | 2,5 | TAK | 95 | 8 | 8 | 6 | 0,5 |
| Kaan | Caliskaner | Jagiellonia Białystok | Niemcy | 24 | 0,5 | NIE | 24 | 0 | 0 | 2 | 0,5 |
| Kacper | Śmiglewski | Cracovia Kraków | Polska | 19 | 3,5 | NIE | 17 | 0 | 1 | 13 | 0,5 |
| Kamil | Lukoszek | Górnik Zabrze | Polska | 21 | 0,5 | NIE | 51 | 0 | 0 | 7 | 0,5 |
| Ernest | Terpifowski | Widzew Łódź | Polska | 22 | 0,5 | NIE | 38 | 1 | 0 | 11 | 0,5 |
| Antoni | Klimek | Widzew Łódź | Polska | 21 | 1,5 | NIE | 46 | 1 | 0 | 11 | 0,45 |
| Sebastian | Musioli | Górnik Zabrze | Polska | 27 | 2,5 | NIE | 50 | 2 | 4 | 7 | 0,45 |
| Leonardo | Rocha | Radomiak Radom | Portugalia | 26 | 2,5 | NIE | 19 | 2 | 0 | 10 | 0,45 |
| Juan | Munoz | Zagłębie Lubin | Hiszpania | 28 | 1,5 | NIE | 31 | 2 | 0 | 8 | 0,45 |
| Adrian | Dalmau | Korona Kielce | Hiszpania | 29 | 1,5 | NIE | 28 | 3 | 1 | 16 | 0,45 |

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Tabela 4. Dane do analizy po zakodowaniu zmiennych zerojedynkowych (zawodnicy z miejsc 1-25)

| Imię | Nazwisko | Klub | Narodo- wość | Wiek | Pozostały kontrakt (lata) | Czy reprezentant | % minut | Gole | Asysty | Miejsce drużyny | Wartość (mln euro) |
|------------|--------------|-----------------------|-----------------|------|---------------------------------|---------------------|---------|------|--------|--------------------|-----------------------|
| Kristoffer | Velde | Lech Poznań | 0 | 25 | 2,5 | 1 | 90 | 7 | 4 | 3 | 6,0 |
| Filip | Szymczak | Lech Poznań | 1 | 21 | 3,5 | 0 | 29 | 1 | 2 | 3 | 3,0 |
| Erik | Exposito | Śląsk Wrocław | 0 | 27 | 0,5 | 0 | 100 | 14 | 3 | 1 | 2,5 |
| Patryk | Klimala | Śląsk Wrocław | 1 | 25 | 3,5 | 0 | 9 | 1 | 0 | 1 | 1,8 |
| Miakael | Ishak | Lech Poznań | 0 | 30 | 1,5 | 1 | 62 | 6 | 3 | 3 | 1,5 |
| Dino | Hotic | Lech Poznań | 0 | 28 | 2,5 | 1 | 27 | 0 | 2 | 3 | 1,5 |
| Ali | Gholizadeh | Lech Poznań | 0 | 27 | 2,5 | 1 | 10 | 0 | 0 | 3 | 1,5 |
| Eftymios | Koulouris | Pogoń Szczecin | 0 | 27 | 2,5 | 1 | 91 | 8 | 1 | 6 | 1,5 |
| Dominik | Marczuk | Jagiellonia Białystok | 1 | 20 | 2,5 | 0 | 72 | 2 | 9 | 2 | 1,5 |
| Marc | Gual | Legia Warszawa | 0 | 27 | 2,5 | 0 | 47 | 1 | 2 | 5 | 1,3 |
| Adrian | Kapralik | Górnik Zabrze | 0 | 21 | 0,5 | 1 | 64 | 2 | 1 | 7 | 1,2 |
| Pedro | Henrique | Radomiak Radom | 0 | 27 | 2,5 | 0 | 95 | 8 | 4 | 10 | 1,2 |
| Vahan | Bichakhchyan | Pogoń Szczecin | 0 | 24 | 1,5 | 1 | 75 | 4 | 2 | 6 | 1,2 |
| Ilya | Shkurin | Stal Mielec | 0 | 24 | 0,5 | 0 | 84 | 8 | 5 | 12 | 1,2 |
| Ante | Crnac | Raków Częstochowa | 0 | 20 | 4,5 | 0 | 43 | 2 | 2 | 4 | 1,0 |
| Afimico | Pululu | Jagiellonia Białystok | 0 | 24 | 1,5 | 0 | 55 | 6 | 2 | 2 | 0,9 |
| Kacper | Chodyna | Zagłębie Lubin | 1 | 24 | 1,5 | 0 | 76 | 3 | 5 | 8 | 0,9 |
| Jose | Naranjo | Jagiellonia Białystok | 0 | 29 | 1,5 | 0 | 75 | 6 | 1 | 2 | 0,85 |
| Piotr | Samiec-Talar | Śląsk Wrocław | 1 | 22 | 0,5 | 0 | 57 | 4 | 5 | 1 | 0,8 |
| Benjamin | Kalliman | Cracovia Kraków | 0 | 25 | 1,5 | 1 | 83 | 5 | 4 | 13 | 0,7 |
| Adriel | Ba Loua | Lech Poznań | 0 | 27 | 1,5 | 0 | 46 | 4 | 2 | 3 | 0,7 |
| Michael | Ameyaw | Piast Gliwice | 1 | 23 | 2,5 | 0 | 92 | 2 | 3 | 9 | 0,65 |
| Mateusz | Wdowiak | Zagłębie Lubin | 1 | 27 | 2,5 | 0 | 56 | 4 | 1 | 8 | 0,65 |
| Vagner | Goncalves | Radomiak Radom | 0 | 28 | 2,5 | 1 | 0 | 0 | 0 | 10 | 0,6 |
| Maciej | Rosolek | Legia Warszawa | 1 | 22 | 1,5 | 0 | 36 | 0 | 2 | 5 | 0,6 |

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Tabela 5. Dane do analizy po zakodowaniu zmiennych zerojedynkowych (zawodnicy z miejsc 26-50)

| Imię | Nazwisko | Klub | Narodowość | Wiek | Pozostały kontrakt (lata) | Czy reprezentant | % minut | Gole | Asysty | Miejsce drużyny | Wartość (w mln euro) |
|------------|--------------|-----------------------|------------|------|---------------------------|------------------|---------|------|--------|-----------------|----------------------|
| Burak | Ince | Śląsk Wrocław | 0 | 20 | 1,5 | 0 | 22 | 2 | 1 | 1 | 0,60 |
| Lisandro | Semedo | Radomiak Radom | 0 | 27 | 0,5 | 1 | 61 | 2 | 2 | 10 | 0,60 |
| Blaz | Kramer | Legia Warszawa | 0 | 27 | 1,5 | 1 | 21 | 2 | 0 | 5 | 0,60 |
| Dawid | Kurminowski | Zagłębie Lubin | 1 | 24 | 1,5 | 0 | 63 | 3 | 0 | 8 | 0,55 |
| Kristoffer | Hansen | Jagiellonia Białystok | 0 | 29 | 1,5 | 0 | 39 | 5 | 1 | 2 | 0,55 |
| Łukasz | Zwoliński | Raków Częstochowa | 1 | 30 | 2,5 | 0 | 42 | 4 | 0 | 4 | 0,55 |
| Jordi | Sanchez | Widzew Łódź | 0 | 29 | 0,5 | 0 | 81 | 4 | 3 | 11 | 0,55 |
| Daniel | Szczepan | Ruch Chorzów | 1 | 28 | 1,5 | 0 | 84 | 7 | 1 | 17 | 0,55 |
| Mateusz | Żukowski | Śląsk Wrocław | 1 | 22 | 0,5 | 0 | 54 | 0 | 0 | 1 | 0,50 |
| Kai | Meriluoto | Stal Mielec | 0 | 21 | 0,5 | 1 | 17 | 2 | 0 | 12 | 0,50 |
| Fabian | Piasecki | Piast Gliwice | 1 | 28 | 3 | 0 | 29 | 0 | 0 | 9 | 0,50 |
| Bartłomiej | Pawłowski | Widzew Łódź | 1 | 31 | 2,5 | 0 | 67 | 4 | 3 | 11 | 0,50 |
| Mariusz | Fornalczyk | Korona Kielce | 1 | 21 | 2,5 | 0 | 14 | 2 | 0 | 16 | 0,50 |
| Jakub | Łukowski | Korona Kielce | 1 | 27 | 0,5 | 0 | 13 | 0 | 0 | 16 | 0,50 |
| Soichiro | Kozuki | Górnik Zabrze | 0 | 23 | 1 | 0 | 10 | 0 | 0 | 7 | 0,50 |
| Kamil | Grosicki | Pogoń Szczecin | 1 | 35 | 2,5 | 1 | 95 | 8 | 8 | 6 | 0,5 |
| Kaan | Caliskaner | Jagiellonia Białystok | 0 | 24 | 0,5 | 0 | 24 | 0 | 0 | 2 | 0,50 |
| Kacper | Śmiglewski | Cracovia Kraków | 1 | 19 | 3,5 | 0 | 17 | 0 | 1 | 13 | 0,50 |
| Kamil | Lukoszek | Górnik Zabrze | 1 | 21 | 0,5 | 0 | 51 | 0 | 0 | 7 | 0,50 |
| Ernest | Terpitiowski | Widzew Łódź | 1 | 22 | 0,5 | 0 | 38 | 1 | 0 | 11 | 0,50 |
| Antoni | Klimek | Widzew Łódź | 1 | 21 | 1,5 | 0 | 46 | 1 | 0 | 11 | 0,45 |
| Sebastian | Musioliik | Górnik Zabrze | 1 | 27 | 2,5 | 0 | 50 | 2 | 4 | 7 | 0,45 |
| Leonardo | Rocha | Radomiak Radom | 0 | 26 | 2,5 | 0 | 19 | 2 | 0 | 10 | 0,45 |
| Juan | Munoz | Zagłębie Lubin | 0 | 28 | 1,5 | 0 | 31 | 2 | 0 | 8 | 0,45 |
| Adrian | Dalmáu | Korona Kielce | 0 | 29 | 1,5 | 0 | 28 | 3 | 1 | 16 | 0,45 |

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Tabela 6. Macierz korelacji

| | Narodowość | Wiek | Pozostały kontrakt (lata) | Czy reprezentant | % minut | Gole | Asysty | Miejsce drużyny |
|---------------------------|------------|-------|---------------------------|------------------|---------|-------|--------|-----------------|
| Narodowość | 1,00 | -0,18 | 0,18 | -0,43 | -0,01 | -0,26 | 0,09 | 0,18 |
| Wiek | -0,18 | 1,00 | 0,05 | 0,22 | 0,22 | 0,40 | 0,12 | 0,04 |
| Pozostały kontrakt (lata) | 0,18 | 0,05 | 1,00 | -0,03 | -0,15 | -0,11 | 0,09 | -0,09 |
| Czy Reprezentant | -0,43 | 0,22 | -0,03 | 1,00 | 0,08 | 0,09 | 0,08 | -0,04 |
| % minut | -0,01 | 0,22 | -0,15 | 0,08 | 1,00 | 0,75 | 0,64 | -0,04 |
| Gole | -0,26 | 0,40 | -0,11 | 0,09 | 0,75 | 1,00 | 0,45 | -0,08 |
| Asysty | 0,09 | 0,12 | 0,09 | 0,08 | 0,64 | 0,45 | 1,00 | -0,15 |
| Miejsce drużyny obecny | 0,18 | 0,04 | -0,09 | -0,04 | -0,04 | -0,08 | -0,15 | 1,00 |

Źródło: opracowanie własne.

Tabela 7. Wyniki testu Studenta i p-value

| Zmienna | Statystyka t | p-value |
|---------------------------|--------------|---------|
| Wyraz wolny | 2,415 | 0,020 |
| Narodowość | -0,149 | 0,883 |
| Wiek | -2,101 | 0,042 |
| Pozostały kontrakt (lata) | 2,218 | 0,032 |
| Czy reprezentant | 2,050 | 0,047 |
| % Minut | 0,271 | 0,788 |
| Gole | 1,769 | 0,084 |
| Asysty | 0,022 | 0,982 |
| Miejsce drużyny | -2,208 | 0,033 |

Źródło: opracowanie własne.

Tabela 8. Wartości statystyk testu Studenta, p-value i symbole zmiennych

| Zmienna | Symbol zmiennej | Statystyka t | p-value |
|---------------------------|-----------------|--------------|---------|
| Wyraz wolny | - | 2,742 | 0,009 |
| Wiek | X_1 | -2,238 | 0,030 |
| Pozostały kontrakt (lata) | X_2 | 2,357 | 0,023 |
| Czy reprezentant | X_3 | 2,478 | 0,017 |
| Miejsce drużyny | X_4 | -2,379 | 0,022 |
| Gole | X_5 | 3,331 | 0,002 |

Źródło: opracowanie własne.

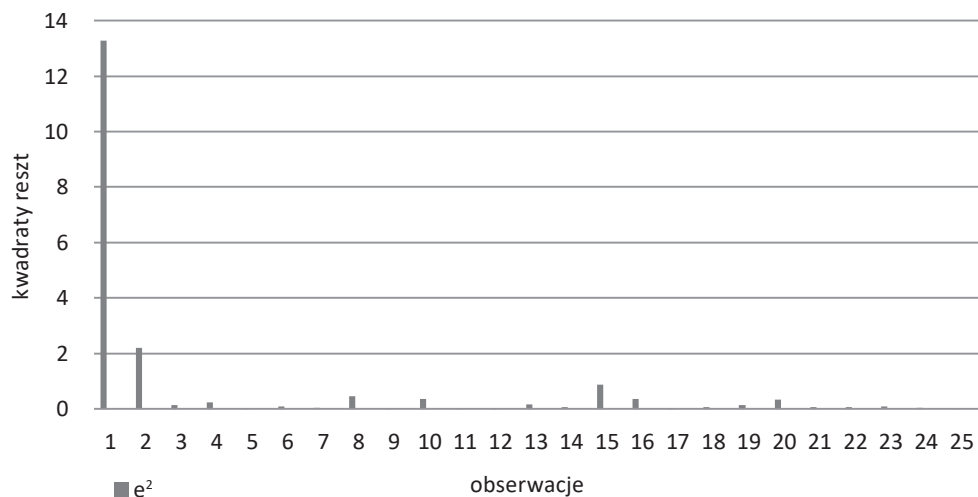
Ponadto w tabeli 8 wskazano symbole poszczególnych zmiennych. Zastosowanie symboli ma na celu uproszczenie zapisu modelu z wyestymowanymi parametrami strukturalnymi:

$$\hat{y} = 2,255 - 0,076X_1 + 0,256X_2 + 0,604X_3 - 0,056X_4 + 0,132X_5 \quad (7)$$

Współczynnik determinacji wyniósł 0,41, co świadczy o słabym (41%) dopasowaniu modelu do danych. W dalszej kolejności reszty modelu zostały poddane sprawdzeniu dwóch własności: normalności i heteroskedastyczności. Uznano, że wykonanie tych dwóch testów (Shapiro-Wilka i Goldfeldta-Quandta) pozwoli na ostateczną weryfikację modelu. Jako pierwszy przeprowadzono test normalności wektora reszt. Statystyka testowa wyniosła 0,18, co na poziomie istotności $\alpha = 0,05$ oraz wartości krytycznej równej 0,947 oznacza konieczność odrzucenia hipotezy zerowej i przyjęcia alternatywnej. Należy zatem stwierdzić, że reszty nie mają rozkładu normalnego. W dalszej kolejności przeprowadzono test homoskedastyczności składnika resztowego. W tym celu podzielono obserwacje na dwie równoliczne grupy (po 25 obserwacji). Na rysunkach 3 i 4 zaprezentowano wykresy kwadratów reszt modelu w podziale na dwie grupy po 25 obserwacji.

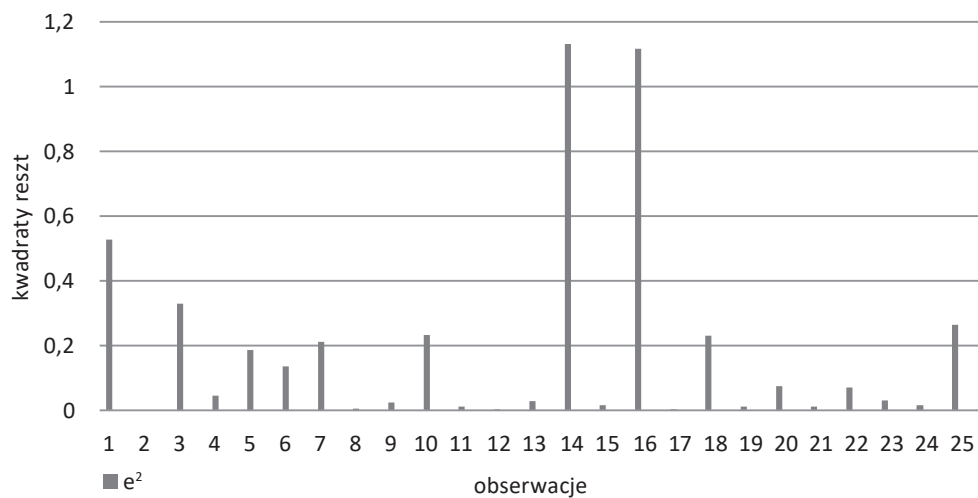
Zestawiając oba wykresy zaprezentowane poniżej, można podejrzewać występowanie zjawiska heteroskedastyczności. Jednak, aby to potwierdzić, należy zastosować odpowiedni test. Statystyka testowa wyniosła 468,76, co jest wartością zdecydowanie większą niż wartość krytyczna, która wynosi 2,01. Można zatem stwierdzić, test Goldfeldta-Quandta wskazał na poziomie istotności $\alpha = 0,05$, że w analizowanym modelu występuje zjawisko heteroskedastyczności.

Biorąc pod uwagę trzy poruszone wcześniej kryteria oceny modelu: współczynnik determinacji (dopasowanie modelu do danych), normalność i homoskedastyczność składnika resztowego, należy stwierdzić, że utworzony model jest niesatysfakcjonujący. Wspomniane kryteria nakazują dość ostrożnie podchodzić do wnioskowania na podstawie modelu. Jednak należy uznać, że założony cel – identyfikacja zmiennych mających istotny wpływ na wycenę zawodników – został osiągnięty. Co więcej, analizując znak poszczególnych zmiennych w modelu, można uznać, że mają one swoje logiczne uzasadnienie. Przeważnie wyżej wyceniani są piłkarze młodszy, gdzie często eksperci piłkarscy mówią o „kupowaniu potencjału zawodnika”. W przypadku dłuższego kontraktu również piłkarze są wyceniani wyżej, ponieważ zgodnie z tzw. prawem Bosmana piłkarz, którego kontrakt z danym klubem wygaś, może przenieść się do innego klubu za darmo (Klimczewski, 2020). W przypadku zawodnika będącego reprezentantem kraju również można zauważyć wyższe wyceny portalu Transfermarkt. Podobnie zawodnik, który strzela dużo goli jest wyżej wyceniany niż strzelający mniej, zwłaszcza że analiza skupiona była na piłkarzach ofensywnych. Analogiczny wniosek da się wysnuć w przypadku drużyn zajmujących niskie miejsce w tabeli – taki piłkarz będzie przeważnie niżej wyceniany niż zawodnik drużyny będącej na pozycji lidera tabeli. Można zatem przyjąć, że pomimo niespełnienia założeń, model jest wart uwagi. Zwłaszcza w aspekcie wytypowania dających się logicznie i intuicyjnie uzasadnić zmiennych, które wpływają na wartość zawodnika.



Rys. 3. Wykres kwadratów reszt dla pierwszej grupy

Źródło: opracowanie własne.



Rys. 4. Wykres kwadratów reszt dla drugiej grupy

Źródło: opracowanie własne.

4. Podsumowanie

Celem przeprowadzonej analizy było sprawdzenie, co w głównej mierze odpowiada za wartość piłkarza. Miał posłużyć do tego model ekonometryczny, który w przypadku bardzo dobrego dopasowania do danych oraz spełnienia warunków normalności i homoskedastyczności wektora reszt mógłby spełniać rolę pomocniczą w wycenach zawodników. Należy jednakże zauważyć, że sam dobór zmiennych był obciążony ryzykiem niepowodzenia. Do modelu zostały bowiem włączone jedynie zmienne dotyczące bieżącego sezonu. W przypadku piłkarzy można przypuszczać, że ich wartość zależy także od tego, w jakich klubach poprzednio grali, ile notowali bramek i asyst, czy odnosili poważne kontuzje oraz jaki wpływ na drużynę miała ich gra. Z drugiej strony można rozważyć teorię, iż model ekonometryczny nie jest najlepszą metodą do wyceny piłkarzy. Poważnie należałoby rozważyć zastosowanie bardziej zaawansowanych metod, np. z użyciem technik uczenia maszynowego. Innym możliwym rozwiązaniem jest zwiększenie próby i ponowna weryfikacja modelu bazującego na większej liczbie zawodników.

Literatura

- Biermann, C. (2022). *Piłkarscy hakerzy. O rewolucji w futbolu i sztuce zbierania danych*. Wydawnictwo SQN.
- Bobowski, Z. (2004). Wybrane metody statystyki opisowej i wnioskowania statystycznego. *Nauki Społeczne. Prace Dydaktyczne*, (5).
- Dębowska-Mróz, M., Ferensztajn-Galardos, E., Krajewska, R., i Rogowski, A. (2018). Ocena możliwości zastosowania rozkładu normalnego do opisu wybranych parametrów ruchu drogowego w miastach na przykładzie Radomia. *Autobusy: Technika, Eksploatacja, Systemy Transportowe*, (19). <https://doi.org/10.24136/atest.2018.419>
- Dziechciarz, J. (red.). (2003). *Ekonometria. Metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Gładysz, B. i Mercik, J. (2007). *Modelowanie ekonometryczne. Studium przypadku*. Oficyna Wydawnicza Politechniki Wrocławskiej.
- Górecki, B. (2010). *Ekonometria. Podstawy teorii i praktyki*. Key Text.
- Klimczewski, W. (2020, 14 sierpnia). *Największa rewolucja w historii futbolu – Prawo Bosmana*. Pobrano z <https://prawosportowe.pl/a/najwieksza-rewolucja-w-historii-futbolu-prawo-bosmana>
- Kufel, J., i Lewandowski, P. (red.). (2023). *Medycyna Przyszłości: rewolucja sztucznej inteligencji w opiece zdrowotnej*. ArchaeGraph.
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka: Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Sumpter, D. (2022). *Piłkomatyka. Matematyczne piękno futbolu*. Copernicus Center Press.
- Tubacki, P. (2023, 8 listopada). *Zarobki w Ekstraklasie. Ile pieniędzy krąży w polskiej lidze?* Pobrano z <https://www.sts.pl/blog/zarobki-w-ekstraklasie/>
- TransferMarkt.pl. (2024). *PKO BP Ekstraklasa*. Pobrano z https://www.transfermarkt.pl/pko-bp-ekstraklasa/marktwerte/wettbewerb/PL1/plus//galerie/0?pos=Sturm&detailpos=&altersklasse=alle&land_id=0

Analysis of Football Players Valuation Using Econometric Modeling

Abstract: The article discusses topics related to football, which is the most popular sport in the world. The main focus of the research was the analysis of football players valuations. The first part provided general information on player valuations, including the most expensive transfers in football. The second part was devoted to describing the research method – in this case, it was an econometric model and its verification. Meanwhile, the third part presented data analysis, which was conducted using MS Excel software. The aim of the study was to identify variables that most strongly influence the value of a football player and to create an econometric model to assist in player valuation. After verifying the coefficient of determination and conducting tests for the normality and homoscedasticity of the residual vector, it was found that the model did not meet basic requirements, and further inference was deemed unwarranted.

Keywords: valuation, econometric modelling, football

Julia Matysiak

e-mail: 181981@student.ue.wroc.pl

ORCID: 0009-0003-9990-1262

Uniwersytet Ekonomiczny we Wrocławiu

Sytuacja na rynku nieruchomości mieszkaniowych w Polsce na tle wybranych wskaźników społeczno-ekonomicznych

DOI: 10.15611/2024.53.6.07

JEL Classification: C38, D91, R31

© 2024 Julia Matysiak

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Matysiak, J. (2024). Sytuacja na rynku nieruchomości mieszkaniowych w Polsce na tle wybranych wskaźników społeczno-ekonomicznych. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 99-114). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Rynek nieruchomości mieszkaniowych w Polsce, z uwzględnieniem podziału na powiaty, stanowi przedmiot analizy niniejszego artykułu. Badanie koncentruje się na różnicach sytuacji rynkowej oraz wybranych wskaźnikach społeczno-ekonomicznych między poszczególnymi regionami kraju. Celem badania było wyodrębnienie czynników istotnie różnicujących powiaty i wskazanie tych lokalizacji, które odznaczają się na tle pozostałych. W tym celu wykorzystano analizę głównych składowych i na jej podstawie stworzono ranking. Uzyskane wyniki wskazują, że głównie wielkomiejskość powiatu determinuje sytuację na rynku mieszkaniowym. Pozostałe czynniki, takie jak jakość życia i bezpieczeństwo zdrowotne, wpływają głównie na zróżnicowanie wewnątrz grup, czyli miast na prawach powiatu i pozostałych powiatów. Sytuacja mieszkaniowa w 2021 roku najkorzystniejsza była w powiatach miejskich, takich jak Olsztyn, Rzeszów, Poznań, Lublin, Kraków, Łódź.

Słowa kluczowe: rynek nieruchomości, inwestycje w nieruchomości, wartość nieruchomości, analiza rynku, analiza głównych składowych

1. Wstęp

W artykule podjęto tematykę związaną z sytuacją na rynku nieruchomości mieszkaniowych w Polsce, który od lat istotnie wpływa na dynamikę gospodarczą kraju i odgrywa ważną rolę w procesach społeczno-ekonomicznych. Jest to sektor niezmienne przyciągający nowych inwestorów, jak i osoby poszukujące miejsca do życia. W obliczu wyzwań demograficznych, zmieniających się preferencji konsumentów

kich i rozwoju infrastruktury oferta rynku nieruchomości powinna być przystosowana do zmieniającego się popytu.

„Rynek nieruchomości to pewna struktura, obejmująca ukształtowany zbiór wzajemnie ze sobą powiązanych uczestników procesów oferowania i wymiany nieruchomości, polegających na przekazywaniu praw własności i/lub praw do korzystania z nieruchomości, oraz warunków, na których realizowane są te procesy” (Kałkowski, 2003, s. 15). Właściciel nieruchomości może czerpać zysk nie tylko ze sprzedaży nieruchomości, ale również z innych praw związanych z nieruchomością, na przykład z najmu. Nieruchomość stanowi również dobre zabezpieczenie dla kredytodawców, gdyż nie można w prosty i szybki sposób zmienić jej lokalizacji.

Istnieje mała elastyczność popytu i podaży nieruchomości z powodu braku substytutów oraz długotrwałego procesu inwestycyjnego. Na sytuację na rynku nieruchomości wpływają czynniki społeczne, ekonomiczne, polityczne i demograficzne. Cechuje go mała efektywność z powodu stałości, niepodzielności i małej płynności nieruchomości. Popyt jest zależny od dochodów, potrzeb mieszkaniowych, czynników cenowych i trendów rynkowych. Najważniejsze czynniki cenowe to zmiany cen mieszkań i czynszu.

Głównym celem niniejszego artykułu jest zbadanie charakterystyki rynku nieruchomości w poszczególnych powiatach, wskazanie powiatów wyróżniających się na tle pozostałych i wyodrębnienie czynników wpływających na jego zróżnicowanie. Wykorzystane w analizach zmienne dotyczą wybranych wskaźników społeczno-ekonomicznych w 380 powiatach Polski w 2021 roku i zostały pobrane z Banku Danych Lokalnych. Dla dokładnego zbadania sytuacji i uporządkowania powiatów według stopnia korzystności na rynku nieruchomości posłużono się analizą głównych składowych.

2. Rynek nieruchomości – aspekty gospodarcze i ekonomiczne

2.1. Charakterystyka rynku mieszkaniowego w Polsce

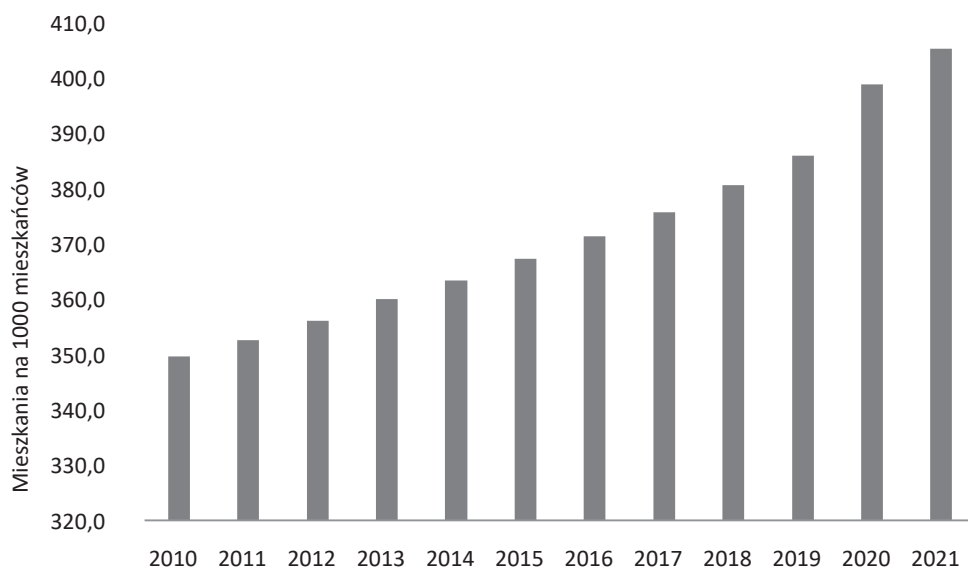
Rynek mieszkaniowy w Polsce stale ewoluuje zarówno pod względem liczby dostępnych mieszkań, jak i ich jakości. Według raportu Głównego Urzędu Statystycznego [GUS] (2022), opartego na wynikach Narodowego Spisu Powszechnego Ludności i Mieszkań 2021, liczba mieszkań w 2021 roku wzrosła o 13,7% w porównaniu z rokiem 2011. W 6 947,8 tys. budynków znajduje się 15 340,1 tys. mieszkań, co oznacza dodanie prawie 1,7 mln nowych lokali oraz zwiększenie powierzchni o 20,3%. Ten wzrost wskazuje na rozwój sytuacji mieszkaniowej w kraju przez ostatnią dekadę pomimo spadku liczby ludności. Natomiast wątpliwość może budzić jakość i trwałość budynków, bo aż 7 302,7 tys. z nich pochodzi z czasów PRL-u, czyli z lat 1945-1988.

Zgodnie z raportem Deloitte: „Property Index of European Residential Markets 11th Edition” (2022), Polska zajęła trzecie miejsce w Unii Europejskiej pod względem

liczby oddanych do użytkowania mieszkań w 2021 roku, ustępując tylko Niemcom i Francji. Spośród tych mieszkań prawie 21,6 tys. (9% ogółu) zostało wybudowanych w Warszawie. Obserwuje się znaczące różnice między dużymi a mniejszymi ośrodkami miejskimi, gdzie 36% populacji Polski mieszka w miastach wojewódzkich i ich powiatach. To właśnie tam przekazano do użytku 57% nowych lokali w 2021 roku, a najwięcej (38% ogółu) powstało w pięciu największych aglomeracjach: Warszawie, Krakowie, Wrocławiu, Poznaniu i Trójmieście.

Warunki mieszkaniowe uległy poprawie w okresie międzypisowym, a dysproporcje w wyposażeniu mieszkań między miastem a wsią zmniejszyły się. W 2021 roku 85,2% budynków mieszkalnych w kraju było podłączonych do wodociągu, a 52,7% miało dostęp do kanalizacji. Ponadto około 87% wszystkich mieszkań było wyposażonych w łazienkę (GUS, 2022).

Analiza zmian w zasobach mieszkaniowych potwierdza wzrost inwestycji mieszkaniowych oraz rozwój budownictwa mieszkaniowego. Należy jednak pamiętać, że zmniejszająca się liczba mieszkańców Polski również wpłynęła na uzyskane wyniki. W latach 2010-2021 średnia liczba mieszkań na 1000 mieszkańców wzrosła z 349,6 w 2010 roku do 405,2 w 2021 roku. Jak przedstawia rysunek 1, sytuacja stale poprawiała się w ciągu tego okresu.



Rys. 1. Liczba mieszkań na 1000 mieszkańców w latach 2010-2021 w Polsce

Źródło: opracowanie własne na podstawie danych GUS.

Wzrost liczby dostępnych mieszkań jest wynikiem licznych inwestycji deweloperskich, przebudowy istniejących nieruchomości oraz zmiany ich charakteru prze-

znaczenia. Narodowy Program Mieszkaniowy z 2016 roku zakładał osiągnięcie do 2030 wskaźnika 435 mieszkań przypadających na 1000 mieszkańców, który obecnie odpowiada średniemu poziomowi w Unii Europejskiej. Jednakże warto zauważyć, że ten wskaźnik nie zawsze odzwierciedla rzeczywisty poziom zaspokojenia potrzeb mieszkaniowych szczególnie w regionach, gdzie duży udział stanowią mieszkania wakacyjne.

Wskaźniki demograficzne opublikowane przez GUS świadczą o malejącej liczbie gospodarstw domowych i zwiększającej się liczbie mieszkań w Polsce. Te zmiany prowadzą do zmniejszenia współczynnika przeludnienia mieszkań z 46,3% w 2012 roku do 35,7% w 2021 (Eurostat, 2022a). Zmniejszenie się liczby osób przypadających na jedno mieszkanie wynika zarówno z redukcji liczby gospodarstw domowych, jak i budowy nowych mieszkań. Jednocześnie wciąż istnieje wysoki odsetek młodych osób, które, pomimo ukończenia studiów i znalezienia pracy, nadal mieszkają z rodzicami. Głównym czynnikiem tego zjawiska są wysokie w porównaniu do zarobków ceny najmu i zakupu mieszkań oraz obecnie ograniczone możliwości kredytowe. Dodatkowo mieszkanie z rodzicami pozwala młodym osobom na obniżenie kosztów utrzymania mieszkania, które w Polsce stanowią średnio 17,9% dochodu, jak wynika z danych Eurostatu (2022b).

Mieszkanie pełni kluczową rolę jako dobro podstawowe, dlatego zmiany w zakresie zasobów mieszkaniowych, wyposażenia czy podłączenia do instalacji mają istotny wpływ na rozwój społeczeństwa. Należy pamiętać, że również inne czynniki, takie jak sytuacja na rynku pracy, polityka państwa, starzejące się społeczeństwo, oprocentowanie kredytów hipotecznych oraz koszty utrzymania mają duży wpływ zarówno na poprawę, jak i na pogorszenie sytuacji mieszkaniowej.

2.2. Rynek mieszkaniowy w Polsce na tle innych krajów Unii Europejskiej

Współczesna mobilność sprawia, że zmiana miejsca zamieszkania zarówno w kraju, jak i za granicą staje się coraz łatwiejsza. Jak czują się Polacy, mieszkając w swoich miastach? Czy istnieją cechy wspólne charakteryzujące miasta, gdzie ludzie są szczęśliwi? Raport „Szczęśliwy dom. Miasto dobre do życia” (Thinkco.pl, 2022), opracowany przez firmę ThinkCo we współpracy z Otodom, analizuje te kwestie. Zgodnie z wynikami ankiety, około połowa mieszkańców Polski czuje się szczęśliwa w środowisku miejskim. Najwyżej ocenione pod względem poziomu szczęścia miasta to Gdańsk, Gdynia i Zielona Góra, podczas gdy najniżej oceniono Częstochowę, Kalisz i Rudę Śląską.

W raporcie zauważono, że poziom szczęścia wzrasta wraz z wielkością miasta, ale tylko w miastach o populacji od 200 do 500 tys. mieszkańców. Najmniej zadowolone są osoby zamieszkujące miasta o wielkości od 50 do 100 tys. mieszkańców. Badanie podkreśla także czynniki istotne dla zadowolenia mieszkańców, takie jak dostępność sklepów, usług zdrowotnych, kultury i rozrywki, koszty życia, komunikacja miejska i stan środowiska. Te informacje są ważne dla deweloperów, biur nieruchomości i władz miast, aby lepiej odpowiadać na potrzeby mieszkańców.

W większości państw europejskich obserwuje się kryzys mieszkaniowy charakteryzujący się wzrostem cen nieruchomości i stawek najmu przy braku skutecznych rozwiązań zadowalających właścicieli, nabywców i najemców. Średnio około 70% mieszkańców Europy jest właścicielami nieruchomości. Jednak w krajach Europy Środkowej i Północnej, takich jak Szwajcaria, Niemcy i Austria, odsetek osób posiadających własne nieruchomości jest niższy niż średnia europejska. W niektórych krajach Europy Wschodniej: Chorwacji, Rumunii, Bułgarii i Słowacji, mniej niż 20% obywateli mieszka w wynajmowanych nieruchomościach (Mattarocci i Roberti, 2020).

Warto pamiętać, że również pandemia COVID-19 istotnie wpłynęła na rynek nieruchomości. Wraz ze spadkiem aktywności gospodarczej dochody pracowników niektórych branż uległy pomniejszeniu, a w pewnych przypadkach nawet zostały całkowicie utracone. To z kolei spowodowało problemy z regulowaniem kredytów hipotecznych, płaceniem czynszów i podatków związanych z nieruchomościami. W odpowiedzi na ten negatywny wpływ na rynek nieruchomości mieszkaniowych niektóre rządy wprowadziły specjalne środki, takie jak ulgi hipoteczne czy ochrona najemców przed eksmisją. Przykłady takich działań można znaleźć w Niemczech, Portugalii i Włoszech (Kholodilin, 2020).

W Turcji zanotowano wzrost popytu na grunty i gospodarstwa rolne na obszarach wiejskich. Blisko połowa wszystkich transakcji nieruchomościowych w 2020 roku dotyczyła nowych nieruchomości. Ponadto obserwuje się tam wzrost liczby gospodarstw domowych posiadających więcej niż jedną nieruchomość. Spodziewane są zmiany w podejściu do inwestycji oraz podejmowaniu decyzji dotyczących planowania i budowy osiedli mieszkaniowych (Tanrivermiş, 2020).

W Polsce i Bułgarii pierwotny rynek nieruchomości mieszkalnych wyraźnie zareagował na sytuację pandemiczną. Na początku zanotowano spadek wartości indeksów cen wskazujący na niepewność co do dalszego rozwoju sytuacji pandemii. Po lepszym zrozumieniu reakcji głównych wskaźników gospodarczych i alternatywnych rynków inwestycyjnych uczestnicy rynku nieruchomości dostosowali się do nowych warunków, a w drugiej połowie 2020 roku dynamika wzrostu cen się zwiększyła (Palicki i Kostov, 2022).

Na wtórnym rynku nieruchomości mieszkalnych w Polsce zaobserwowano przejściowe zmiany. W pierwszym kwartale 2020 roku nastąpił gwałtowny wzrost cen (o 5,9% w porównaniu do poprzedniego kwartału). W kolejnych kwartałach dynamika wzrostu cen stopniowo osłabiała się, a na koniec czwartego kwartału 2020 roku wzrost wyniósł jedynie 0,49% w porównaniu do poprzedniego kwartału. W Bułgarii zarówno rynek wtórny, jak i pierwotny nieruchomości mieszkalnych doświadczyły wpływu sytuacji pandemicznej, co spowodowało znaczne wahania trendów (Palicki i Kostov, 2022).

Pandemia COVID-19 wywarła negatywny wpływ na dotychczasowy rozwój zjawisk na rynkach nieruchomości w krajach Unii Europejskiej, powodując ich gwałtowne zaburzenie. Jednocześnie przyspieszyła procesy jak zdalną pracę i rozwój

cyfrowy, które rozpoczęły się już wcześniej. Pojawiły się również nowe wymagania dotyczące mieszkań, które prawdopodobnie będą miały wpływ na rynek nieruchomości w najbliższych latach.

2.3. Wartość ekonomiczna nieruchomości

Wartość ekonomiczna nieruchomości zależy nie tylko od jej fizycznych cech, ale także od czynników środowiskowych, ekonomicznych, społecznych i prawnych (Kucharska-Stasiak, 2016). Te czynniki mogą mieć znacznie większy wpływ na wartość nieruchomości niż czynniki fizyczne. Podczas poszukiwania nieruchomości istotne są takie czynniki, jak wiek budynku, jego stan techniczny, wielkość działki (w przypadku domów), udogodnienia funkcjonalne i technologiczne, dostępność oraz bezpieczeństwo okolicy. Zmiany w poziomie zatrudnienia i dochodów ludności mogą również wpływać na popyt na nieruchomości – przy wzroście bezrobocia i spadku dochodów wartość nieruchomości może maleć.

Wpływ na wartość nieruchomości mają również czynniki, takie jak ulgi podatkowe, ochrona praw własności, plany zagospodarowania terenu, opłaty związane z posiadaniem nieruchomości oraz wymogi dotyczące ochrony środowiska (Kucharska-Stasiak, 2006). Także restrykcyjne regulacje dotyczące ogrzewania mogą wpłynąć na atrakcyjność nieruchomości. Sytuacja demograficzna danego regionu również odgrywa istotną rolę – wyż demograficzny może zwiększyć popyt na nieruchomości, zwłaszcza na te położone w pobliżu szkół i przedszkoli. W większych miastach, gdzie znajdują się uczelnie wyższe, liczba studentów może wpływać na ceny najmu i wartość nieruchomości.

Trudno wskazać jednoznacznie, który czynnik ma największy wpływ na wartość nieruchomości, ponieważ różne rodzaje nieruchomości mogą być wielorako oceniane przez uczestników rynku. Wartość nieruchomości kształtuje się w wyniku oddziaływania różnych czynników, a nie tylko na podstawie jej fizycznych cech.

3. Metodyka badań

3.1. Rodzaj i źródła danych

Analiza sytuacji na rynku nieruchomości mieszkaniowych jest zadaniem złożonym, ponieważ na jego kształtowanie wpływa wiele czynników. Aby dokładnie zrozumieć tę sytuację, konieczne jest uwzględnienie zmiennych z różnych kategorii, które oddziałują na rynek nieruchomości, a także decyzje podejmowane przez inwestorów i potencjalnych nabywców mieszkań.

Wybrano 11 wskaźników społeczno-ekonomicznych charakteryzujących polskie powiaty do zbadania sytuacji na rynku nieruchomości mieszkaniowych w Polsce. Są one następujące:

- X_1 – liczba sprzedanych lokali mieszkalnych w ramach transakcji rynkowych na 1000 mieszkańców;

- X_2 – typ powiatu (0 – powiat, 1 – miasto na prawach powiatu);
- X_3 – średnia cena za 1 m² lokalu mieszkalnego sprzedanego w transakcji rynkowej [zł];
- X_4 – przeciętna powierzchnia jednego sprzedanego lokalu mieszkalnego [m²];
- X_5 – przeciętne miesięczne wynagrodzenie brutto [zł];
- X_6 – stopa bezrobocia rejestrowanego [%];
- X_7 – przestępstwa stwierdzone przez policję przeciwko mieniu na 1000 mieszkańców;
- X_8 – przestępstwa stwierdzone przez policję przeciwko życiu i zdrowiu na 1000 mieszkańców;
- X_9 – udział powierzchni terenów zieleni w powierzchni ogółem [%];
- X_{10} – lekarze (personel pracujący ogółem) na 1000 mieszkańców;
- X_{11} – emisja zanieczyszczeń gazowych na 1 km² [t/km²].

Dane dla 380 powiatów, dotyczące 2021 roku pobrano z Banku Danych Lokalnych (<https://bdl.stat.gov.pl>). Dla zmiennych X_1 i X_3 wystąpił brak danych dla powiatu miasta Koszalin. Chcąc uniknąć usuwania jednego obiektu z analizy, uogólnione zostały dla niego dane z powiatu koszalińskiego dla tych dwóch zmiennych. Dla zmiennej X_{11} wystąpiły braki danych dla piętnastu powiatów. Do zastąpienia brakujących danych określoną wartością liczbową wykorzystano imputację. Braki danych dotyczące wartości zostały uzupełnione średnią wartością zmiennej, która została obliczona na podstawie danych z powiatów sąsiadujących z powiatem, w którym brakowało danych.

3.2. Metody badawcze

W celu wyodrębnienia kilku kluczowych czynników reprezentujących w dostatecznym stopniu obszerny zbiór zmiennych zastosowana została analiza głównych składowych. Jest ona techniką eksploracyjną umożliwiającą badanie relacji w zbiorze zmiennych i stosowaną, by uprościć reprezentację za pomocą nowych zmiennych, będących kombinacjami liniowymi zmiennych pierwotnych (Afifi i in., 2012).

Analizę rozpoczyna się od wyboru zmiennych związanych z badanym zjawiskiem, które zostaną nią objęte. Następnie dane wejściowe są transformowane poprzez standaryzację, aby zapobiec trudnościom wynikającym z różnic w ich skali (Balicki, 2013). Na podstawie zestandaryzowanych danych tworzona jest macierz korelacji.

Aby zredukować wielowymiarowość złożonego zjawiska, jakim jest sytuacja na rynku nieruchomości mieszkaniowych, zastępuje się liczbę obserwowanych zmiennych mniejszą liczbą nieobserwowalnych zmiennych, czyli głównych składowych. W celu określenia oczekiwanej liczby głównych składowych stosuje się następujące kryteria:

- 70-90% całkowitej zmienności powinno być wyjaśniane przez p pierwszych składowych;

- ograniczenie liczby głównych składowych do tych, które mają własności własne większe od 1;
- wybór liczby głównych składowych na podstawie wykresu osypiska.

Przy określaniu liczby głównych składowych należy zwracać uwagę na sensowność interpretacji uzyskanych wyników analizy.

Wykorzystując macierz korelacji i wektor własny odpowiadający największej wartości własnej, można utworzyć ranking. Jednocześnie warto dokonać sprawdzenia, czy wagi stanowiące pierwszy wektor własny macierzy korelacji dla zmiennych będących stymulantami przyjmują wartości dodatnie, a dla destymulant wartości ujemne. Jeśli znaki się nie zgadzają, trzeba zamienić je na przeciwne. Wyznaczone wartości pierwszej głównej składowej są porządkowane malejąco.

4. Analiza rynku nieruchomości mieszkaniowych w Polsce

4.1. Statystyki opisowe

Dokładne zbadanie wpływu wybranych czynników społeczno-ekonomicznych i ogólnej charakterystyki rynku nieruchomości w poszczególnych powiatach wymaga wcześniejszej analizy sytuacji w skali całego kraju i zidentyfikowania powiatów, które wyraźnie wyróżniają się na tle pozostałych.

W tabeli 1 przedstawiono statystyki opisowe dla zmiennych ilościowych, które wskazują, że w 2021 roku średnia liczba sprzedanych mieszkań wyniosła 4,59 sztuk na 1000 mieszkańców. Najwięcej mieszkań w przeliczeniu na 1000 mieszkańców sprzedano w powiecie kołobrzeskim, a najmniej w powiecie ostrołęckim. Najwyższa średnia cena za metr kwadratowy mieszkania wynosiła 14 888 zł i została odnotowana w powiecie miasta Sopot. Najtaniej natomiast było w powiecie lwóweckim. Zmienności wszystkich cech są powyżej 10%.

Tabela 1. Statystyki opisowe dla zmiennych ilościowych

| Zmienna | Średnia | Mediana | Minimum | Maksimum | Współczynnik zmienności |
|----------|---------|---------|---------|-----------|-------------------------|
| X_1 | 4,59 | 3,64 | 0,05 | 25,86 | 88,13 |
| X_3 | 4219,84 | 3992,50 | 2060,00 | 14 888,00 | 33,02 |
| X_4 | 54,22 | 53,01 | 44,29 | 118,25 | 12,43 |
| X_5 | 5211,63 | 5070,22 | 4244,56 | 10 076,64 | 12,27 |
| X_6 | 8,06 | 7,10 | 1,60 | 26,30 | 55,15 |
| X_7 | 9,19 | 7,86 | 2,52 | 52,00 | 61,99 |
| X_8 | 0,36 | 0,34 | 0,08 | 0,88 | 41,59 |
| X_9 | 1,74 | 0,33 | 0,04 | 22,50 | 198,69 |
| X_{10} | 2,57 | 1,89 | 0,48 | 15,81 | 86,42 |
| X_{11} | 2438,21 | 66,69 | 0,03 | 79 864,64 | 387,33 |

Źródło: opracowanie własne na podstawie danych GUS.

W powiecie lubińskim zanotowano najwyższe średnie wynagrodzenie miesięczne brutto, co jest związane z obecnością takich firm, jak KGHM Polska Miedź, Zakłady Górnicze Lubin, Huta Miedzi Cedynia i innych działających w sektorze miedziowym. Natomiast powiat poznański i powiat miasto Poznań charakteryzowały się najniższą stopą bezrobocia rejestrowanego wynoszącą 1,6% w roku 2021. Te powiaty niewątpliwie przyciągały nowych mieszkańców ze względu na swoją korzystną sytuację ekonomiczną.

Najwięcej przestępstw przeciwko mieniu na 1000 mieszkańców odnotowano w powiecie miasto Katowice, a przeciwko życiu i zdrowiu w powiecie gryfickim. Pod względem emisji zanieczyszczeń gazowych na 1 km² najgorzej wypadł powiat miasto Ostrołęka, w którym funkcjonuje Zespół Elektrowni Ostrołęka. W powiatach, które nie są miastami na prawach powiatu, można zaobserwować zdecydowanie niższy udział terenów zieleni w powierzchni ogółem i mniejszą liczbę lekarzy na 1000 mieszkańców. W dalszej części analizy zostanie sprawdzane, czy te niedogodności zaobserwowane dla czynników społecznych istotnie wpłyną na uzyskiwane wyniki.

4.2. Analiza głównych składowych

Do identyfikacji kluczowych czynników wpływających na rynek nieruchomości mieszkaniowych wykorzystano analizę głównych składowych, która opierała się na zmiennych od X_2 (typ powiatu) do X_{11} (emisja zanieczyszczeń gazowych na 1 km²).

Wyniki analizy głównych składowych wykazały, że sytuację na rynku nieruchomości mieszkaniowych różnicowały cechy związane z pierwszymi czterema głównymi składowymi, które wyjaśniały łącznie 73,23% łącznej wielowymiarowej zmienności cech – tabela 2. Do zbadania siły i kierunku zależności wykorzystano dane zawarte w tabeli 3.

Tabela 2. Wartości oraz skumulowane wartości własne głównych składowych

| Nr wartości | Wartość własna | % ogółu wariancji | Skumulowane wartości własne | Skumulowane % wariancji |
|-------------|----------------|-------------------|-----------------------------|-------------------------|
| 1 | 4,187266 | 41,87266 | 4,18727 | 41,8727 |
| 2 | 1,406660 | 14,06660 | 5,59393 | 55,9393 |
| 3 | 0,933053 | 9,33053 | 6,52698 | 65,2698 |
| 4 | 0,796454 | 7,96454 | 7,32343 | 73,2343 |
| 5 | 0,688754 | 6,88754 | 8,01219 | 80,1219 |
| 6 | 0,638284 | 6,38284 | 8,65047 | 86,5047 |
| 7 | 0,560347 | 5,60347 | 9,21082 | 92,1082 |
| 8 | 0,410357 | 4,10357 | 9,62117 | 96,2117 |
| 9 | 0,226765 | 2,26765 | 9,84794 | 98,4794 |
| 10 | 0,152061 | 1,52061 | 10,00000 | 100,0000 |

Źródło: opracowanie własne na podstawie danych GUS.

Tabela 3. Ładunki czynnikowe dla czterech głównych składowych

| Zmienna | Czynnik 1 (wielkowiejskość powiatu) | Czynnik 2 (jakość życia) | Czynnik 3 (bezpieczeństwo zdrowotne) | Czynnik 4 (pozostałe) |
|----------|--|-----------------------------|---|--------------------------|
| X_2 | -0,870807 | -0,186643 | -0,157711 | -0,167242 |
| X_3 | -0,653954 | 0,315814 | 0,240604 | -0,149431 |
| X_4 | -0,009819 | 0,773866 | 0,022720 | 0,224904 |
| X_5 | -0,654693 | 0,192328 | 0,034582 | 0,396185 |
| X_6 | 0,515566 | -0,522339 | -0,073970 | -0,123800 |
| X_7 | -0,728972 | -0,081145 | 0,232644 | -0,034166 |
| X_8 | -0,284148 | -0,528519 | 0,603085 | 0,429427 |
| X_9 | -0,833227 | -0,140166 | -0,103787 | -0,267241 |
| X_{10} | - | -0,007681 | -0,034919 | -0,205398 |
| X_{11} | | -0,240346 | -0,642869 | 0,473395 |

Źródło: opracowanie własne na podstawie danych GUS.

Pierwsza główna składowa, która wyjaśnia 41,87% całkowitej zmienności, była najsilniej skorelowana z typem powiatu (-0,87), udziałem terenów zieleni w powierzchni powiatu ogółem (-0,83), liczbą lekarzy na 1000 mieszkańców (-0,88). Wymienione cechy w największym stopniu różnicowały analizowane powiaty. Z współzależności tych zmiennych można wnioskować, że w powiatach niebędących miastami na prawach powiatu występuje mniej terenów zielonych i przyjmuje mniej lekarzy, więc dostęp do służby zdrowia może być trudniejszy.

Dodatkowo biorąc pod uwagę skorelowanie pierwszej głównej składowej z średnią ceną za 1 m² (-0,65), przeciętną wysokością wynagrodzenia miesięcznego brutto (-0,65) i przestępstwa przeciwko mieniu na 1000 mieszkańców (-0,73), można przyjąć, że najistotniejszy czynnik determinujący rynek nieruchomości mieszkaniowych to **wielkowiejskość powiatu**. Wskazują na to zawarte w tym czynniku zmienne z tym samym znakiem zarówno te pozytywne – jak udział terenów zieleni w powierzchni ogółem, przeciętna miesięczna wysokość wynagrodzenia, jak i te negatywne, czyli liczba przestępstw przeciwko mieniu oraz życiu i zdrowiu na 1000 mieszkańców. Pozostałe czynniki w mniejszym stopniu determinują rynek nieruchomości mieszkaniowych.

Druga główna składowa przede wszystkim odzwierciedlała średnią powierzchnię sprzedawanych lokali mieszkalnych (0,77), stopę bezrobocia rejestrowanego (-0,52) oraz liczbę przestępstw przeciwko życiu i zdrowiu (-0,52), wyjaśniając 14,07% całkowitej zmienności. Stanowi ona ważny czynnik różnicujący **jakość życia** w miastach na prawach powiatu. Dowodzi to tego, że większe mieszkania mają pozytywny wpływ na jakość życia, podczas gdy wyższe stopy bezrobocia i większa liczba przestępstw przeciwko zdrowiu mają negatywny wpływ.

Trzecia główna składowa, wyjaśniając 9,33% zmienności całkowitej, wykazywała silne powiązania z przestępstwami przeciwko życiu i zdrowiu na 1000 mieszkań-

ców (0,60) oraz emisją zanieczyszczeń gazowych ogółem na 1 km² (-0,64). Przeciwnie znaki dla tych dwóch negatywnych zjawisk komplikują interpretację, ponieważ w miastach na prawach powiatu przestępczość i emisja zanieczyszczeń są z reguły wyższe w porównaniu do innych powiatów, co negatywnie wpływa na stan zdrowia.

Następnie sporządzono ranking w oparciu o analizę głównych składowych w celu określenia, w których powiatach sytuacja społeczno-ekonomiczna wpływająca na rynek nieruchomości mieszkaniowych w 2021 roku była najlepsza. Zmienne będące stymulantami to bycie miastem na prawach powiatu (wartość 1 dla zmiennej X_2), przeciętna powierzchnia jednego sprzedanego mieszkania, przeciętne miesięczne wynagrodzenie brutto, udział terenów zieleni w powierzchni ogółem i lekarze na 1000 mieszkańców. Destymulanty to średnia cena za 1 m² lokalu mieszkalnego, stopa bezrobocia rejestrowanego, przestępstwa przeciwko mieniu, życiu i zdrowiu, emisja zanieczyszczeń gazowych na 1 km². Wyniki przedstawione w tabeli 4 świadczą o tym, że miasta na prawach powiatu przewodzą, co potwierdza kluczową rolę tego czynnika. Pozostałe różnice występują wewnątrz tych grup, tj. miast na prawach powiatu i pozostałych powiatów.

Tabela 4. Ranking powiatów na podstawie analizy głównych składowych

| Powiat | Wartość w rankingu |
|----------------------------|--------------------|
| Powiat m. Olsztyn | 4,586042153 |
| Powiat m. Rzeszów | 4,140554454 |
| Powiat m. Poznań | 3,864188972 |
| Powiat m. Lublin | 3,769191892 |
| Powiat m. Kraków | 3,315553905 |
| Powiat m. Łódź | 3,263654537 |
| Powiat m. Jastrzębie-Zdrój | 3,231341082 |
| Powiat m. Białystok | 3,220969121 |
| Powiat m. Szczecin | 3,049103221 |
| Powiat m.st. Warszawa | 2,962495410 |
| Powiat m. Wrocław | 2,942239152 |
| ... | ... |
| Powiat lipnowski | -1,337223926 |
| Powiat zawierciański | -1,369185044 |
| Powiat koszaliński | -1,404786393 |
| Powiat gryficki | -1,405961749 |
| Powiat choszczeński | -1,415088722 |
| Powiat sławieński | -1,432766842 |
| Powiat górowski | -1,498084796 |
| Powiat nowodworski | -1,775024821 |
| Powiat tatrzański | -2,057450384 |
| Powiat kamieński | -2,886644577 |

Źródło: opracowanie własne na podstawie danych GUS.

Najlepsze wyniki osiągnął powiat miasto Olsztyn, będący centralnym ośrodkiem gospodarczym, edukacyjnym i kulturalnym w regionie. Charakteryzuje się on niską stopą bezrobocia rejestrowanego (2,2%) oraz średnim wynagrodzeniem brutto o 13% wyższym od średniej krajowej. Dodatkowo, obszary zielone zajmują aż 22,06% całkowitej powierzchni tego miasta. To połączenie korzystnych czynników, wraz z konkurencyjnymi cenami za metr kwadratowy, korzystnie wpływa na rynek nieruchomości mieszkaniowych w tym powiecie.

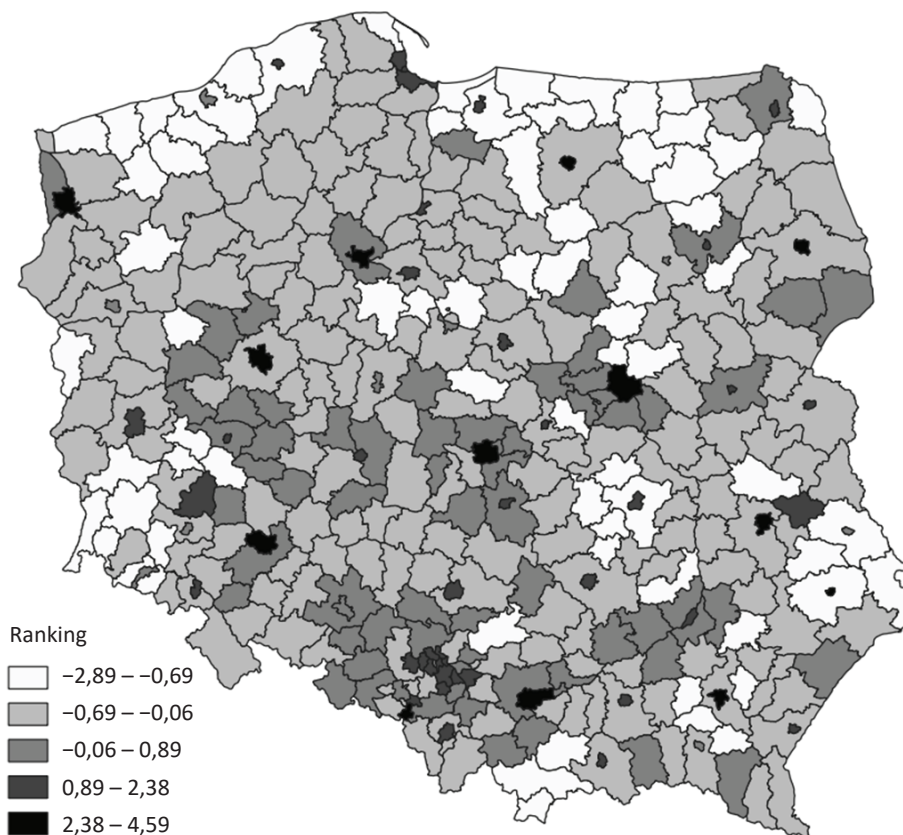
Pozostałe powiaty z czołówki rankingu wykazują podobną charakterystykę. Wysokie nakłady na utrzymanie terenów zielonych powodują, że ich udział w całkowitej powierzchni jest znacznie większy niż w innych powiatach. Koncentracja dużych firm oraz placówek edukacyjnych przyciąga do osiedlania się w tych powiatach, co sprzyja zakupowi mieszkań. Daje to możliwość uzyskiwania wyższych dochodów, łatwiejszego znalezienia pracy i dostępu do bardziej specjalistycznych usług medycznych.

W większych miastach obserwuje się zwiększoną liczbę przestępstw przeciwko mieniu, włączając w to przypadki kradzieży. Dlatego też władze tych obszarów powinny skupić się na zapewnieniu mieszkańcom poczucia bezpieczeństwa. Właściciele i zarządcy nieruchomości również powinni rozważyć dodatkowe środki zabezpieczające ich posiadłości, a także ewentualnie zakupienie ubezpieczenia od kradzieży i włamania.

Najgorszy wynik w rankingu osiągnął powiat kamieński położony w województwie zachodnio-pomorskim. Pomimo niewielkiej emisji zanieczyszczeń gazowych i niedużej liczby przestępstw przeciwko zdrowiu i życiu na 1000 mieszkańców pozostałe warunki społeczno-ekonomiczne nie sprzyjają osiedlaniu się w tym regionie. Wysokie ceny mieszkań, prawdopodobnie związane z dużym zainteresowaniem ze strony inwestorów, którzy wykupują nieruchomości na wynajem turystom lub korzystają z nich w celu wypoczynkowym, sprawiają, że ta lokalizacja nie jest atrakcyjna dla potencjalnych mieszkańców. Dodatkowo, średnia powierzchnia sprzedawanych mieszkań w tym regionie wynosi jedynie 45,41 m², co jest wartością znacząco niższą od średniej krajowej.

Uzyskane w rankingu wyniki zostały dodatkowo zaprezentowane na kartogramie – rysunek 2.

Powiaty zajmujące niższe pozycje w rankingu, cechujące się mniej korzystnymi warunkami społeczno-ekonomicznymi w porównaniu do innych regionów, przede wszystkim znajdują się w północnej Polsce, głównie w województwach zachodnio-pomorskim i warmińsko-mazurskim. To częściowo wynik nadal istniejącego wysokiego bezrobocia zwłaszcza na Mazurach, gdzie wcześniej duży odsetek ludności pracował w Państwowych Gospodarstwach Rolnych, które po zmianie ustroju zostały zlikwidowane, pozostawiając wielu ludzi bez pracy. Oprócz tego regiony te zajmują się głównie turystyką letnią, a w pozostałych miesiącach aktywność gospodarcza jest ograniczona.



Rys. 2. Kartogram dla rankingu na podstawie analizy głównych składowych

Źródło: opracowanie własne na podstawie danych GUS.

Najlepszą sytuację prezentują powiaty zlokalizowane w województwie śląskim. Konurbacja górnośląska, z dużym skupiskiem miast na prawach powiatu, stanowi istotne centrum gospodarcze, gdzie czynniki ekonomiczne mają korzystny wpływ na rynek nieruchomości. Te obszary posiadają liczne miejsca pracy związane głównie z przemysłem wydobywczym oraz instytucjami edukacyjnymi, co w połączeniu z przystępnymi cenami za metr kwadratowy mieszkania może przyciągać nabywców. Jedynym potencjalnym problemem jest wysoka emisja zanieczyszczeń gazowych w tym regionie.

Warto zauważyć, że powiaty sąsiadujące z miastami na prawach powiatu zazwyczaj mają z tego względu korzystną sytuację, co pozytywnie wpływa na lokalny rynek nieruchomości. Na tych obszarach powstają nowe mieszkania, odległość do większych aglomeracji miejskich jest niewielka, a ceny nieruchomości są atrakcyjniejsze niż w centrum dużych miast. Poza tym dostęp do usług medycznych jest ła-

twiejszy, a duże strefy przemysłowe często znajdują się na obrzeżach miast, umożliwiając mieszkańcom sąsiednich powiatów łatwiejszy dojazd do pracy.

5. Zakończenie

Rynek nieruchomości mieszkaniowych w Polsce, podzielony na powiaty, wykazuje znaczne zróżnicowanie. Pomimo wykorzystania wskaźników społeczno-ekonomicznych w analizie nie oddaje on całej złożoności rynku ani indywidualnych preferencji potencjalnych nabywców. Decyzja o lokalizacji zamieszkania często zależy od miejsca pracy, dostępności szkół wyższych i osobistych preferencji. Przewidywane zmiany na rynku nieruchomości mieszkaniowych mogą wynikać z pandemii COVID-19, która zmieniła nawyki ludzi i zwiększyła popularność pracy zdalnej. Istotny wpływ ma także napływ ludności z Ukrainy, co stanowi nową sytuację na polskim rynku nieruchomości. Jednocześnie Polska zmagą się z problemem niżu demograficznego, którego skutki będą odczuwalne w przyszłości.

Przeprowadzona analiza pozwala zidentyfikować kluczowe czynniki wpływające na rynek nieruchomości mieszkaniowych i jego atrakcyjność dla potencjalnych nabywców. Chociaż znaczenie tych czynników może się zmieniać w zależności od sytuacji, nadal odgrywają one istotną rolę i kształtują atrakcyjność rynku nieruchomości mieszkaniowych w poszczególnych powiatach.

Analiza głównych składowych wyodrębniła cztery kluczowe grupy czynników oddziałujących na rynek nieruchomości mieszkaniowych. Najważniejszym z nich okazała się wielkomiejskość powiatu. Miasta na prawach powiatu wyróżniają się większym udziałem terenów zielonych, łatwiejszym dostępem do wysoko wykwalifikowanego personelu medycznego, wyższymi zarobkami i większą liczbą ofert pracy. Posiadają one również lepszy dostęp do szkół wyższych, wydarzeń kulturalnych i atrakcyjnych opcji zatrudnienia, co istotnie wpływa na jakość życia. Przeważają nad pozostałymi powiatami także pod względem rynku nieruchomości mieszkaniowych.

Ranking opracowany na podstawie analizy głównych składowych potwierdził dominującą rolę wielkomiejskości w kształtowaniu rynku nieruchomości mieszkaniowych. Pozostałe czynniki, w tym jakość życia i bezpieczeństwo zdrowotne, głównie różnicują sytuację wewnątrz poszczególnych grup, tj. miast na prawach powiatu i pozostałych powiatów.

Najbardziej korzystna sytuacja na rynku nieruchomości mieszkaniowych występuje przede wszystkim w dużych miastach na prawach powiatu. Stanowią one wyjątek na tle pozostałych powiatów, oferując nie tylko możliwości rozwoju, ale także zapewniając szeroki dostęp do usług oraz starając się o zapewnienie bezpieczeństwa swoim mieszkańcom.

W przypadku powiatów, takich jak kamieński, tatrzański, nowodworski, szydłowiecki i zamojski, konieczne byłoby podjęcie działań mających na celu poprawę sytuacji na rynku nieruchomości mieszkaniowych, ponieważ aktualnie zajmują one

ostatnie miejsca w rankingu. W miastach: Opolu, Ostrołęce, Płocku, Katowicach i Koninie, władze powinny rozważyć środki mające na celu zmniejszenie emisji zanieczyszczeń gazowych, ponieważ wysoki poziom zanieczyszczenia środowiska ma negatywny wpływ na zdrowie obecnych mieszkańców i może odstraszać potencjalnych nowych lokatorów.

Literatura

- Afifi A., May S., i Clark V.A. (2012). *Practical Multivariate Analysis*. CRC Press Taylor & Francis Group.
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Deloitte. (2022). *Property Index Overview of European Residential Markets 11th edition*. Pobrano z https://www2.deloitte.com/content/dam/Deloitte/cz/Documents/real-estate/Property_Index_2022.pdf
- Eurostat. (2022a). *Wskaźnik przeludnienia mieszkań*. Pobrano z https://ec.europa.eu/eurostat/data-browser/view/ILC_LVHO05A/default/table?lang=en1
- Eurostat. (2022b). *Wydatki gospodarstw domowych na utrzymanie mieszkania*. Pobrano z https://ec.europa.eu/eurostat/databrowser/view/HBS_EXP_T123B/default/table?lang=en&category=livcon.hbs.hbs_exp
- Główny Urząd Statystyczny. (2022). *Narodowy spis powszechny ludności i mieszkań 2021. Raport z wstępnych wyników*. Pobrano z <https://stat.gov.pl/spisy-powszechnie/nsp-2021/nsp-2021-wyniki-wstepne/raport-zawierajacy-wstepne-wyniki-nsp-2021,6,1.html>
- Kałkowski, L. (red.). (2003). *Rynek nieruchomości w Polsce*. TWIGGER
- Kholodilin, K. A. (2020). Housing Policies Worldwide During Coronavirus Crisis: Challenges and Solutions. *DIW focus*, (2). Deutsches Institut für Wirtschaftsforschung. Pobrano z <https://hdl.handle.net/10419/219095>
- Kucharska-Stasiak, E. (2006). *Nieruchomości w gospodarce rynkowej*. Wydawnictwo Naukowe PWN.
- Kucharska-Stasiak, E. (2016). *Ekonomiczny wymiar nieruchomości*. Wydawnictwo Naukowe PWN.
- Mattarocci, G., i Roberti, S. (2020). Real Estate and the Effects of the COVID-19 Pandemic in Europe, W: C. Bagnoli (red.), *A New World Post COVID-19. Lessons for Business, the Finance Industry and Policy Makers*. Edizioni Ca' Foscari – Digital Publishing. <https://doi.org/10.30687/978-88-6969-442-4>
- Palicki, S. i Kostov, I. (2022). Rynek mieszkaniowy w erze pandemii COVID-19. Studium porównawcze Poznania i Warny. W: P. Bartkowiak (red.), *Tendencje rozwoju współczesnego rynku nieruchomości mieszkaniowych*, 147-161. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. <https://dx.doi.org/10.18559/978-83-8211-124-8/9>
- Tanrıvermiş, H. (2020). Possible Impacts of COVID-19 Outbreak on Real Estate Sector and Possible Changes to Adopt: A Situation Analysis and General Assessment on Turkish Perspective. *Journal of Urban Management*, 9(3), 263-269. <https://doi.org/10.1016/j.jum.2020.08.005>
- Thinkco.pl. (2022). *Szczęśliwy dom. Miasto dobre do życia*. Pobrano z <https://thinkco.pl/raport-szczesliwy-dom/>

Situation on the Residential Real Estate Market in Poland against the Background of Selected Socio-economic Indicators

Abstract: The residential real estate market in Poland, taking into account the division into counties, is the subject of the analysis in this article. The study focuses on differences in the market situation and selected socio-economic indicators between individual regions of the country. The aim of the study was to highlight factors that significantly differentiate counties and to identify those locations that stand out from the rest. For this purpose, principal components analysis was used and a ranking was created based on it. The obtained results indicate that the metropolitan nature of the district mainly determines the situation on the housing market. Other factors, such as quality of life and health security, mainly influence the differences within groups, i.e. cities with county rights and other counties. The most favourable housing situation in 2021 was in urban counties such as Olsztyn, Rzeszow, Poznan, Lublin, Krakow, and Lodz.

Keywords: real estate market, real estate investments, real estate value, market analysis, principal components analysis

Łukasz Meller

e-mail: 181993@student.ue.wroc.pl

ORCID: 0009-0001-4365-0500

Uniwersytet Ekonomiczny we Wrocławiu

Analiza atrakcyjności ryнку lokali mieszkalnych w województwie mazowieckim

DOI: 10.15611/2024.53.6.08

JEL Classification: C38, R3

© 2024 Łukasz Meller

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Meller, Ł. (2024). Analiza atrakcyjności rynku lokali mieszkalnych w województwie mazowieckim. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 115-128). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Wybór miejsca zamieszkania to dylemat, z którym musi się zmierzyć każda osoba. Ostateczna decyzja o wyborze miejsca lokalu mieszkalnego ma wpływ na nasze codzienne życie i dlatego powinna być dobrze przemyślana. Artykuł skupia się na analizie rynku nieruchomości mieszkaniowych w jednym z najbardziej rozwiniętych regionów Polski. Celem artykułu jest zbadanie atrakcyjności rynku lokali mieszkalnych w powiatach województwa mazowieckiego. Przeprowadzone badania mają za zadanie pomóc potencjalnemu nabywcy nieruchomości mieszkalnej w województwie mazowieckim wybrać odpowiednie miejsce zamieszkania. Scharakteryzowano czynniki wpływające na atrakcyjność rynku lokali mieszkalnych oraz wskazano powiaty o podobnej atrakcyjności za pomocą analizy głównych składowych. Powiaty leżące blisko Warszawy odznaczają się zazwyczaj lepszą sytuacją na rynku nieruchomości. Płock i Siedlce można uznać za atrakcyjne miejsce dla osób chcących zamieszkać w innym mieście województwa mazowieckiego niż Warszawa.

Słowa kluczowe: analiza głównych składowych, analiza atrakcyjności, rynek nieruchomości mieszkalnych, województwo mazowieckie

1. Wstęp

Każda osoba zmaga się z decyzją dotyczącą wyboru swojego miejsca zamieszkania. Wszystko zależy od indywidualnych preferencji i potrzeb nabywcy lokum. Niektórzy wolą mieszkać w metropoliach, inni natomiast preferują spokojne, wiejskie okolice, gdzie panuje cisza i spokój. Ostateczna decyzja o wyborze miejsca lokalu mieszkalnego ma wpływ na nasze codzienne życie i dlatego powinna być dobrze przemyślana.

W artykule podjęto temat oceny atrakcyjności rynku lokali mieszkalnych w województwie mazowieckim. Ze względu na bardzo dużą liczbę potencjalnych czynników wpływających na ocenę danego miejsca zamieszkania, wybrano tylko najważniejsze determinanty. Zbiór zmiennych wykorzystanych w pracy jest subiektywną decyzją autora popartą przesłankami z przeprowadzonego przeglądu literatury. Celem badawczym artykułu jest analiza atrakcyjności rynku lokali mieszkalnych w powiatach województwa mazowieckiego. Za pomocą analiz zdecydowano wyjaśnić:

- Czy miasto stołeczne Warszawa powinno być traktowane jak obserwacja ekstremalnie odstająca, przez co przypadek ten powinno się poddać oddzielnej analizie?
- Czy powiaty sąsiadujące ze stolicą Polski charakteryzują się specyficzną strukturą atrakcyjności rynku mieszkalnego?
- Czy miasta na prawach powiatu można uznać za atrakcyjne pod względem wyboru miejsca zamieszkania?

W badaniu, korzystając z analizy głównych składowych, wskazano powiaty o podobnej charakterystyce atrakcyjności dla potencjalnego mieszkańca regionu.

2. Analiza atrakcyjności

2.1. Podstawowe definicje

Prezentując rynek lokali mieszkalnych, niezbędne jest dokładne opisanie niektórych pojęć często używanych w artykule, w celu uniknięcia błędnej interpretacji. Głównym zadaniem tego punktu pracy jest jednoznaczne zdefiniowanie atrakcyjności danego rynku lokali mieszkalnych.

Definicja samodzielnego lokalu mieszkalnego zgodnie z Ustawą z dnia 24 czerwca 1994 r. o własności lokali, art. 2 ust. 2 wskazuje, że jest to „(...) wydzielona trwałymi ścianami w obrębie budynku izba lub zespół izb przeznaczonych na stały pobyt ludzi, które wraz z pomieszczeniami pomocniczymi służą zaspokajaniu ich potrzeb mieszkaniowych”. Konstytucja Rzeczypospolitej Polskiej uznaje lokal mieszkalny za dobro społeczne i rynkowe (Główny Urząd Statystyczny, 2023), które służy zaspokajaniu podstawowej potrzeby ludzkiej, zapewniając schronienie i poczucie bezpieczeństwa (Żelazowski, 2013) oraz wpływa na warunki życia człowieka.

Według E. Kucharskiej-Stasiak (2006) rynek lokali mieszkalnych to przestrzeń ekonomiczna, na której nabywcy (zgłaszający zapotrzebowanie) i sprzedawcy (reprezentujący podaż) podejmują transakcje dotyczące kupna, sprzedaży, wymiany lub wynajmu tych nieruchomości. Na rynku lokali mieszkalnych działa wiele podmiotów, takich jak deweloperzy, inwestorzy, agencje nieruchomości, pośrednicy oraz osoby prywatne, które poszukują bądź oferują lokale mieszkalne. Rynek ten podlega wpływom różnych czynników, takich jak sytuacja gospodarcza kraju, poziom inflacji, polityka fiskalna, stopy procentowe, trendy demograficzne oraz lokal-

na sytuacja rynkowa. Cena i popyt na lokale mieszkalne zależą od wielu czynników, takich jak lokalizacja, metraż, standard wykończenia, dostępność infrastruktury, stan techniczny budynku i inne.

Słowo „atrakcyjny” wywodzi się od łacińskiego *attractio*, które oznacza „przyciągać”. Przymiotnik ten określa pewną cechę lub zbiór cech sprawiających, że określony rynek lokali mieszkalnych uważamy za przyciągający nabywców, inwestorów. Atrakcyjność jest postrzegana w sposób subiektywny, ponieważ to, co jedna osoba uważa za atrakcyjne, może nie przyciągać uwagi innej osoby. Jednakże istnieją pewne uniwersalne cechy, które pozwalają uznać dany rynek lokali mieszkalnych za atrakcyjny.

Zatem atrakcyjność rynku lokali mieszkalnych to poziom zainteresowania i popytu na nieruchomości mieszkalne, który jest determinowany przez szereg czynników, takich jak lokalizacja, cena, powierzchnia użytkowa lokalu, jakość budynku i wyposażenia, dostępność usług i infrastruktury, stan rynku mieszkaniowego, trendy oraz preferencje konsumentów.

2.2. Czynniki określające atrakcyjność danego rynku lokali mieszkalnych

Przyjęte w artykule podejście jest skoncentrowane tylko na wybranych cechach określających atrakcyjność danego rynku lokali mieszkalnych z racji tego, że liczba czynników, która może wpływać na ocenę danego rynku lokali mieszkalnych, jest bardzo duża. Również preferencje osób zainteresowanych kupnem mieszkania mogą być zróżnicowane (Plebankiewicz i Biel, 2017) ze względu na takie cechy jak: wiek, zarobki, wykształcenie, liczba posiadanych dzieci itp. Ponadto odpowiednie zaznajomienie się przez nabywcę z czynnikami przestrzennymi opisującymi miejsce zamieszkania ma kluczowe znaczenie w podejmowaniu decyzji (Słodczyk i Rajchel, 2004). Z racji tego postarano się uwzględnić względnie uniwersalne czynniki, które jednoznacznie definiują, czy dany rynek lokali mieszkalnych można uznać za atrakcyjny, czy też nie.

Wiodącym czynnikiem określającym atrakcyjność danego lokalu mieszkalnego jest jego cena. Kryterium cenowe zawsze będzie ograniczało wybór dostępnych lokali, które mogą być nabyte przez kupującego. Osoba nieposiadająca wystarczających środków finansowych, aby pozyskać lokum o pożądanym cechach w wybranej przez niego lokalizacji, będzie zmuszona do kupna lokalu mieszkalnego o mniejszej powierzchni użytkowej lub zmiany lokalizacji zamieszkania. Również standard warunków zamieszkania zależy od możliwości finansowych nabywcy (Dąbrowska-Milewska, 2007). Lokale wyróżniające się niską, atrakcyjną ceną za metr kwadratowy prawdopodobnie będą charakteryzowały się zaniżonym standardem lub też innymi niedogodnościami związanymi z codziennym życiem. Również należałoby mieć tę zależność na uwadze przy selekcji nieruchomości zaspokajających potrzeby mieszkaniowe. Poziom zarobków, oszczędności oraz zdolność kredytowa wpływają ostatecznie na decyzję o wyborze lokalu pomimo indywidualnych upodobań (Dąbrowski, 2018).

Następną cechą determinującą atrakcyjność rynku lokali mieszkalnych jest jego lokalizacja. Wybór miejsca zamieszkania przekłada się na poziom życia. Czynniki te wpływają na wiele aspektów codziennego życia, takich jak dostępność do szkoły, przedszkoli, żłobków, sklepów, miejsc rekreacyjnych, opieki medycznej, transportu publicznego. Wybór miejsca zamieszkania zapewniającego szybki dostęp do obiektów często odwiedzanych, może przyczynić się do zmniejszenia stresu związanego z długimi dojazdami i pozwolić na więcej czasu spędzanego w domu lub na innych aktywnościach. Miasta w oczywisty sposób zapewniają lokalnej ludności lepsze warunki do życia (Batty i Yichun, 1998). Obszary większych aglomeracji miejskich (o liczbie ludności co najmniej 500 000 mieszkańców) koncentrują miejsca pracy, ośrodki naukowe, placówki medyczne, miejsca wydarzeń kulturowych (Jałowiecki, 2001). Wybór miasta jako miejsca zamieszkania niesie za sobą wiele udogodnień. Na obszarach gęsto zaludnionych obserwuje się lepszą komunikację publiczną, bardziej zróżnicowany rynek pracy, dostęp do szerokiego zakresu usług. Jednak życie w większych aglomeracjach niesie za sobą również gorszą jakość powietrza oraz zwiększone natężenie hałasu niż na peryferiach (Nove Olendry, b.d.).

Kolejnymi czynnikami opisującymi atrakcyjność danego rynku lokali mieszkalnych są informacje dotyczące liczby oddanych nowych mieszkań do użytkowania w owym regionie oraz ich przeciętna powierzchnia użytkowa. Osoby preferujące zamieszkanie w budynkach nowoczesnych oraz przestronnych powinny zwrócić szczególną uwagę na wysoki wskaźnik liczby oddanych nowych lokali mieszkalnych na 1000 mieszkańców, który wskazuje na duży popyt na nieruchomości mieszkaniowe w danej lokalizacji oraz dynamiczny rozwój regionu. Powstające nowe mieszkania mogą przyciągać inwestorów i firmy, a tym samym zwiększać szanse na poprawę okolicznej infrastruktury. W wyniku czego w danym miejscu powstają nowe miejsca pracy, szkoły, punkty usługowe.

Następne wybrane czynniki określające atrakcyjność lokali mieszkalnych przede wszystkim koncentrują się na osobach we wczesnym stadium swojego życia (osoby do 13. roku życia), jak i w podeszłym wieku (osoby powyżej 60. roku życia), gdyż decyzja o wyborze miejsca zamieszkania ma częściowy związek z wiekiem osoby lub strukturą wieku rodziny (House i in., 2013). Małoletnie dzieci oraz seniorzy przeważnie wyróżniają się niższym poziomem samodzielności oraz potrzebą częstszych wizyt u lekarza. Taka charakterystyka owych osób powoduje zwiększenie atrakcyjności miejsc zamieszkania w obszarach z łatwym dostępem do usług medycznych oraz infrastrukturą drogową zapewniającą bezpieczne przejścia dla pieszych. Dodatkowo warto również zwrócić uwagę na to, czy okolica miejsca zamieszkania jest wystraszająco bezpieczna. Niestety seniorzy są grupą społeczeństwa, która najczęściej pada ofiarami kradzieży (Policja.pl, 2019). Ponadto osoby posiadające dzieci szczególnie zwracają uwagę na bliskość miejsca, w którym podczas ich nieobecności mogą pozostawić potomka pod dobrą opieką. Aby posiadać taką możliwość, również liczba miejsc w pobliskim żłobku czy klubie dziecięcym musi być wystarczająca w rozpatrywanym obszarze.

Potencjalny nabywca lokalu mieszkalnego, wybierający miejsce zamieszkania musi wyznaczyć, które cechy determinujące atrakcyjność danego rynku są dla niego najważniejsze. Zaprezentowany zestaw cech określających rynki nieruchomości mieszkalnych prezentuje jedynie wąski zakres wszystkich możliwych determinant opisujących atrakcyjność. Część kupujących może zwracać uwagę jeszcze na inne aspekty związane z lokalami mieszkalnymi, które nie zostały ujęte w rozważaniach.

3. Wyniki badania

3.1. Przesłanki badania

Wykonanie analiz ma na celu wskazanie powiatów w województwie mazowieckim atrakcyjnych pod względem miejsca zamieszkania. Skupiono się na wytypowaniu interesujących regionów w pobliżu miasta stołecznego Warszawa, dla osób chcących mieszkać nieopodal stolicy Polski oraz na przedstawieniu reszty powiatów, które swoją specyfiką mogą wydawać się godne rozważenia pod względem miejsca zamieszkania.

Przeprowadzenie analizy głównych składowych pozwoli na zredukowanie liczby zmiennych do dwóch nieskorelowanych czynników ukrytych, zachowujących bardzo dużą część informacji o zmiennych pierwotnych opisujących atrakcyjność rynku lokali mieszkalnych w danych powiatach województwa mazowieckiego, a także ma na celu zobrazowanie struktury oraz wykrycie ogólnych reguł między zmiennymi. Powyższa metoda pozwala na opisanie i skategoryzowanie powiatów w przestrzeniach określonych przez nowo powstałe czynniki.

3.2. Opis i uzasadnienie wyboru zmiennych poddanych analizie

Pierwotny zbiór danych składa się z 11 zmiennych i 41 obserwacji – powiatów województwa mazowieckiego. Dane uznano za wiarygodne ze względu na ich źródło pochodzenia (Główny Urząd Statystyczny [GUS], 2023), ze względu na brak wartości dla powiatu ostrołęckiego zdecydowano się odrzucić go z dalszych rozważań. Ostateczny zbiór danych składa się z 40 obserwacji. Część zmiennych przekształcono, dzieląc przez liczbę mieszkańców i mnożąc razy 1000 (zmiennie na 1000 mieszkańców). Operacja ta wykonana została w celu zniwelowania różnic wynikających z populacji danych powiatów – dane stały się porównywalne. Obserwacje pochodzą z 2020 roku. Dane z lat wcześniejszych były niepełne i nie zawierały obserwacji dla niektórych zmiennych biorących udział w analizie, dlatego zdecydowano się na wybór ostatniego roku zawierającego pełne informacje.

Zmienne X1-X6 dotyczą danych bezpośrednio opisujących stan rynku lokali mieszkalnych w danym powiecie. Zmienne X7-X11 opisują atrakcyjność danego regionu pod względem bezpieczeństwa, opieki zdrowotnej, zamożności mieszkańców oraz opieki nad dziećmi. Wskaźnik wykrywalność sprawców przestępstw stwierdzonych opisuje „relację liczby przestępstw wykrytych w danym roku (łącznie z wykry-

tymi po podjęciu z umorzenia) do liczby przestępstw stwierdzonych w danym roku, powiększonej o liczbę przestępstw stwierdzonych w podjętych postępowaniach, a umorzonych w latach poprzednich z powodu niewykrycia sprawców” (GUS, b.d.).

Tabela 1. Lista wykorzystanych zmiennych wraz ze statystykami opisowymi

| | Zmienna | Średnia | Współczynnik zmienności [%] | Górny kwartyl Q3 | Rozstęp kwartylowy IQR |
|-----|--|-----------|-----------------------------|------------------|------------------------|
| X1 | Średnia cena za 1 m ² lokali mieszkalnych sprzedanych w ramach transakcji rynkowych [zł] | 4128,1 | 31,8 | 4641,0 | 1298,0 |
| X2 | Mieszkania oddane do użytkowania na 1000 mieszkańców [szt./1000 os.] | 5,4 | 62,6 | 6,0 | 2,5 |
| X3 | Przeciętna wartość lokali mieszkalnych sprzedanych w ramach transakcji rynkowych [zł] | 23 8408,8 | 41,9 | 26 3134,0 | 85 120,0 |
| X4 | Mieszkania na 1000 mieszkańców [szt./1000 os.] | 380,5 | 11,0 | 394,9 | 35,8 |
| X5 | Przeciętna powierzchnia użytkowa 1 mieszkania oddanego do użytkowania [m ²] | 110,4 | 21,1 | 130,4 | 31,4 |
| X6 | Średnia liczba osób zamieszkujących jedno mieszkanie [os.] | 2,7 | 10,6 | 2,9 | 0,4 |
| X7 | Przeciętne miesięczne wynagrodzenia brutto w relacji do średniej krajowej (Polska = 100) [%] | 90,8 | 13,3 | 93,9 | 10,8 |
| X8 | Wypadki drogowe na 1000 mieszkańców [liczba wypadków/1000 os.] | 0,7 | 41,5 | 0,8 | 0,4 |
| X9 | Wskaźnik wykrywalności sprawców przestępstw stwierdzonych [%] | 76,2 | 11,7 | 81,7 | 10,6 |
| X10 | Liczba lekarzy na 1000 mieszkańców [liczba lekarzy/1000 os] | 4,5 | 69,2 | 4,8 | 2,0 |
| X11 | Miejsca w żłobkach i klubach dziecięcych na 1000 dzieci w wieku do lat 3 [liczba miejsc/1000 dzieci] | 103,0 | 59,8 | 135,0 | 77,0 |

Źródło: opracowanie własne na podstawie danych GUS.

Wyliczenie podstawowych statystyk opisowych ma na celu prezentację zróżnicowania wartości zmiennych w województwie mazowieckim. Ważnym jest, aby zmienna biorąca udział w analizie była niejednorodna. Według D. Tarki i A.M. Olszewskiej (2018) zmienna posiadająca współczynnik zmienności (odchylenie standardowe podzielone przez średnią arytmetyczną) na poziomie mniejszym niż 10 % wykazuje zróżnicowanie statystycznie nieistotne. Ponadto, aby dowiedzieć się, jak prezentuje się wartość zmiennej w danym powiecie na tle innych powiatów, potrzebny jest punkt odniesienia, jakim jest przeciętny poziom danej cechy wszystkich

powiatów. W tabeli 1 przedstawiono podstawowe statystyki opisowe zmiennych wykorzystanych do analizy głównych składowych. Pod względem dalszej analizy danych najważniejszy jest fakt, że zmienność każdej ze zmiennych jest wyższa niż 10%. Najbardziej zróżnicowany rozkład cechy posiadają zmienne: X2, X10 i X11 (tab. 1). To oznacza, że te trzy zmienne wykazują największą zmienność. Warto zwrócić szczególną uwagę na te zmienne podczas analizy, ponieważ to właśnie te cechy będą wskazywały największe rozbieżności pomiędzy badanymi powiatami.

3.3. Obserwacje odstające

Według J. Tukeya – amerykańskiego statystyka i matematyka – wartości odstające to punkty danych, które leżą poza następującym zakresem:

$$\{Q_1 - k \times IQR, Q_3 + k \times IQR\}$$

gdzie: Q_1 – pierwszy kwartył danych, Q_3 – trzeci kwartył danych, IQR – rozstęp kwartyłowy, k – nieujemna stała

Tukey zaproponował, że przy $k = 1,5$ wartości znajdujące się poza przedziałem określonym wzorem są traktowane jako odstające, a przy $k = 3$ – jako ekstremalnie odstające (Adil i Irshad, 2015).

Na podstawie analizy tabeli 1 należy usunąć ze zbioru przypadków miasto stołeczne Warszawa, gdyż powiat ten jest obserwacją ekstremalnie odstającą dla zmiennych X1, X3, X4, X10 oraz odstającą dla zmiennych X2, X6, X9, X11. Nieusuwanie tego przypadku mogłoby znacznie zafałszować istniejące zależności między zmiennym (Długosz, 2014), co byłoby niekorzystne dla dalszej analizy atrakcyjności rynku lokali mieszkalnych w województwie mazowieckim.

Miasto stołeczne Warszawa należy poddać oddzielnej analizie ze względu na znacznie inną specyfikę rynku lokali mieszkalnych. Powiat ten cechuje się niemal dwuipółkrotnie wyższymi średnimi cenami za 1 m² lokali mieszkalnych sprzedanych w ramach transakcji rynkowych w porównaniu z przeciętną średnią ceną za 1 m² wszystkich powiatów. Przeciętna powierzchnia użytkowa mieszkania oddanego do użytkowania jest prawie dwa razy mniejsza w porównaniu ze średnią powierzchnią mieszkania w powiatach województwa mazowieckiego. Również można zauważyć znacząco większą liczbę mieszkań na 1000 mieszkańców. Niewątpliwie wyjątkowość rynku warszawskiego wiąże się ze statusem miasta jako stolicy Polski. Popyt na lokale mieszkalne w tej aglomeracji jest na tyle wysoki (Welcome Poland 2, 2022), że ludzie są w stanie zapłacić więcej za 1 m² powierzchni mieszkaniowej oraz mieszkać w przeciętnie mniejszych powierzchniowo lokalach.

3.4. Analiza głównych składowych

Na początku procedury zestandaryzowano dane, by wartości zmiennych były tego samego rzędu i były porównywalne. Następnie obliczono współczynnik Kaisera-Mayera-Olkina (KMO), który „(...) wykorzystywany jest do sprawdzania stopnia skorelowania zmiennych pierwotnych, czyli siły dowodów świadczących na rzecz sensowności przeprowadzania analizy składowych głównych” (PQStat – Baza Wiedzy, 2022). Wartość współczynnika KMO dla zestawu zmiennych wykorzystanych do analizy wynosi 0,728. Keiser ocenia wartość powyżej 0,7 jako bardzo dobrą (Norusis, 1994). Im większa jest wartość tego wskaźnika, tym mocniejsze są podstawy do stosowania analizy głównych składowych. Powyższe założenie zostało spełnione, aby uzasadnić wykorzystanie zestawu danych do dalszej analizy. Po wyznaczeniu wektorów własnych i wartości własnych można określić, które główne składowe najlepiej reprezentują dane. Główne składowe z wyższymi wartościami własnymi mają większy wpływ na zmienność danych, a te o mniejszych wartościach własnych zostaną pominięte.

Tabela 2. Wartości własne

| Główne składowe | Wartości własne | Odsetek wyjaśnianej wariancji [%] | Skumulowany odsetek wyjaśnianej wariancji [%] |
|-----------------|-----------------|-----------------------------------|---|
| 1 | 5,69 | 51,74 | 51,74 |
| 2 | 2,03 | 18,45 | 70,19 |
| 3 | 0,98 | 8,87 | 79,06 |
| 4 | 0,61 | 5,55 | 84,61 |
| 5 | 0,48 | 4,38 | 88,99 |
| 6 | 0,37 | 3,40 | 92,39 |
| 7 | 0,32 | 2,94 | 95,33 |
| 8 | 0,26 | 2,33 | 97,66 |
| 9 | 0,15 | 1,41 | 99,07 |
| 10 | 0,09 | 0,86 | 99,93 |
| 11 | 0,01 | 0,07 | 100,00 |

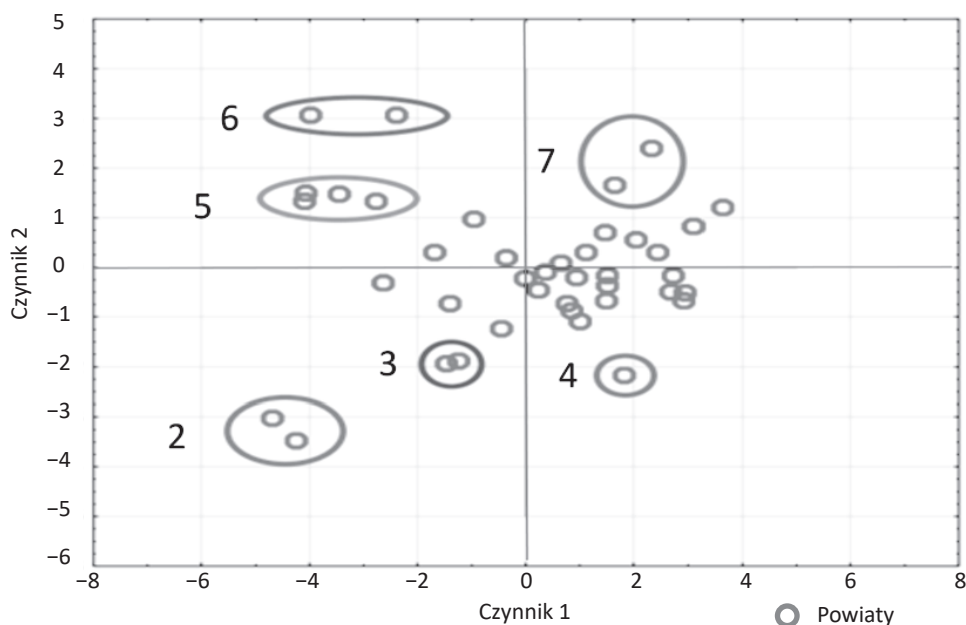
Źródło: opracowanie własne z wykorzystaniem Statistica 13.3.

Tabela 2 przedstawia wartości własne macierzy korelacji. Na tej podstawie można stwierdzić „jaka część całkowitej zmienności jest tłumaczona przez daną składową główną” (Kowalik i Gruszewska, 2015, s. 150). Do określenia liczby głównych składowych wykorzystanych w dalszej analizie wybrano dwa kryteria: subiektywne i obiektywne. Kryterium subiektywne – należy wybrać tyle głównych składowych, aby wyjaśniały ustalony procent ogólnej wariancji. Za wystarczający poziom można uznać 70% (Balicki, 2013), co z kolei prowadzi do wyboru dwóch głównych składowych. Kryterium obiektywne – należy wybrać te główne składowe, dla których wartości własne są większe od 1. Trzecia wartość własna ma wartość własną mniejszą niż 1, z tego powodu należy wybrać dwie pierwsze główne składowe do dalszej analizy.

Tabela 3. Wektory własne

| Zmienna | Czynnik 1 | Czynnik 2 |
|---------|-----------|-----------|
| X1 | -0,360 | 0,180 |
| X2 | -0,299 | 0,399 |
| X3 | -0,304 | 0,401 |
| X4 | -0,266 | -0,483 |
| X5 | 0,191 | 0,282 |
| X6 | 0,345 | 0,269 |
| X7 | -0,333 | -0,053 |
| X8 | 0,253 | -0,197 |
| X9 | 0,289 | -0,277 |
| X10 | -0,286 | -0,363 |
| X11 | -0,349 | -0,111 |

Źródło: opracowanie własne z wykorzystaniem Statistica 13.3.



Rys. 1. Powiaty w przestrzeni dwuwymiarowej ze względu na 1 czynnik i 2 czynnik

Źródło: opracowanie własne na podstawie danych GUS z wykorzystaniem Statistica 13.3.

Tabela 3 przedstawia wektory własne macierzy korelacji. Na jej podstawie możemy poznać wagi poszczególnych składowych:

$$y_1 = -0,360x_1 - 0,299x_2 - 0,304x_3 - 0,266x_4 + 0,191x_5 + 0,345x_6 - 0,333x_7 + 0,253x_8 + 0,289x_9 - 0,286x_{10} - 0,349x_{11}$$

$$y_2 = +0,180x_1 + 0,399x_2 + 0,401x_3 - 0,483x_4 + 0,282x_5 + 0,269x_6 - 0,053x_7 - 0,197x_8 - 0,277x_9 - 0,363x_{10} - 0,111x_{11}$$

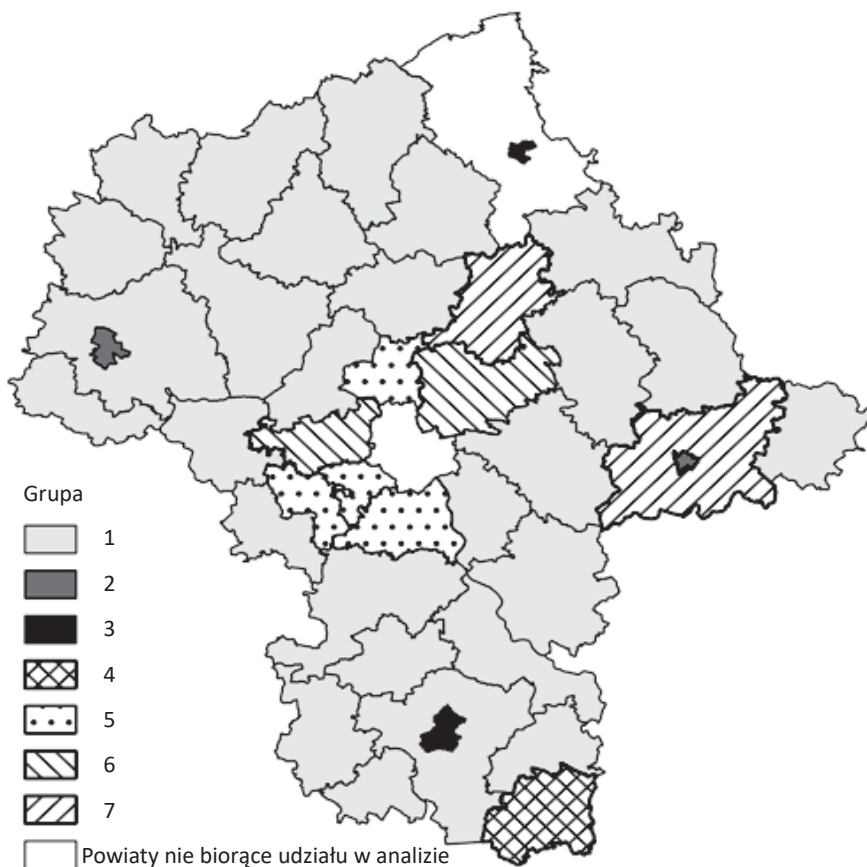
Zmienne y są fikcyjnymi zmiennymi – stworzonymi w celu redukcji wymiaru – nie są to zmienne obserwowane. Miarą odtworzenia informacji będzie wariancja zmiennych $y - \hat{y}$.

Tabela 4. Pogrupowane powiaty ze względu na czynniki 1 i 2

| Numer grupy | Powiaty |
|-------------|---|
| 1 | białobrzeski, ciechanowski, garwoliński, gostyniński, grójecki, kozienicki, łosicki, makowski, miński, mławski, nowodworski, ostrowski, otwocki, płocki, płoński, przasnyski, przysuski, pułtuski, radomski, sierpecki, sochaczewski, sokołowski, szydłowiecki, węgrowski, zwoleński, żuromiński, żyrardowski |
| 2 | m. Płock, m. Siedlce |
| 3 | m. Ostrołęka, m. Radom |
| 4 | lipski |
| 5 | grodziski, legionowski, piaseczyński, pruszkowski |
| 6 | warszawski zachodni, wołomiński |
| 7 | siedlecki, wyszkowski |

Źródło: opracowanie własne.

Na podstawie rysunku 1 można wyodrębnić siedem grup powiatów województwa mazowieckiego ze względu na 1 i 2 czynnik. Grupa 1 znajduje się blisko punktu (0;0). Rysunek 2 pozwala na wizualizację rozmieszczenia grup na mapie chronometrycznej. Miasto stołeczne Warszawa i powiat ostrołęcki nie należą do żadnej z grup, gdyż zostały wyłączone z analizy. Można zauważyć, że większość powiatów otaczających Warszawę i miasta na prawach powiatu wyróżniają się na tle pozostałych powiatów województwa mazowieckiego, tworząc oddzielne grupy. Dodatkowo powiat lipski, siedlecki, wyszkowski oraz grodziski zostały przypisane do innych grup niż grupa 1. Grupa 5 (powiat grodziski, legionowski, piaseczyński, pruszkowski) cechuje się wysoką średnią ceną za 1 m², dużą liczbą mieszkań oddanych do użytku w 2020 roku, niską liczbą wypadków drogowych, ponad przeciętną liczbą lekarzy na 1000 mieszkańców oraz dużą liczbą miejsc w żłobkach. Grupa ta wydaje się być niezwykle atrakcyjna. Można wywnioskować, że jest to najbardziej dynamicznie rozwijający się region pod względem rynku lokali mieszkalnych w województwie mazowieckim. Bardzo podobna do grupy 6 jest grupa 5 (tab. 4). Różnice, które wydają się wyraźne, to mniejsza liczba lekarzy na 1000 mieszkańców oraz mniejsza liczba miejsc w żłobkach na 1000 dzieci do lat 3 w grupie 6. Zatem można uznać, że grupa 5 wydaje się być najbardziej atrakcyjna pod względem rynku lokali mieszkalnych dla potencjalnych nabywców chcących mieszkać niedaleko stolicy Polski. Grupę 2 i 3 stanowią miasta na prawach powiatów. Miasto Płock oraz Siedlce (grupa 2) cechują się zdecydowanie większą liczbą lekarzy i miejsc w żłobkach niż Ostrołęka



Rys. 2. Wizualizacja położenia grup powiatów

Źródło: opracowanie własne z wykorzystaniem programu QGIS 3.30.1.

i Radom (grupa 3). Ciekawy przypadek stanowi grupa 4 zawierająca powiat lipski. Powiat ten cechuje się najmniejszą liczbą mieszkań oddanych do użytku na 1000 mieszkańców w 2020 roku. Również średnia cena za 1 m² powierzchni użytkowej jest na bardzo niskim poziomie. Liczba wypadków drogowych na 1000 mieszkańców jest na najwyższym poziomie wśród reszty powiatów. Można wywnioskować, że jest to najwolniej rozwijający się rynek lokali mieszkalnych w województwie mazowieckim.

3.5. Podsumowanie wyników analizy głównych składowych

Podziału powiatów według atrakcyjności danego rynku lokali mieszkalnych można dokonać za pomocą położenia względem miasta stołecznego Warszawy. Powiaty leżące blisko niej odznaczają się zazwyczaj lepszą sytuacją na rynku nieruchomości.

Takie wnioski mogą wynikać z charakteru niektórych powiatów, jako powiatów satelitarnych dla stolicy województwa mazowieckiego, które pełnią głównie funkcję sypialną (Zawierucha, 2012, s. 44) dla osób dojeżdżających do pracy w stolicy Polski. Ponadto analiza wyodrębniła dwie grupy, które zawierały miasta na prawach powiatu. Niewątpliwie grupę 2 (tab. 4) można uznać za atrakcyjniejszą dla osób chcących zamieszkać w stosunkowo dużym jak na Polskę innym mieście województwa mazowieckiego niż Warszawa. Miasto Płock oraz Siedlce cechują się zdecydowanie większą liczbą lekarzy i miejsc w żłobkach niż Ostrołęka i Radom (grupa 3). Dodatkowo zidentyfikowano powiat lipski jako najmniej atrakcyjne miejsce pod względem zamieszkania w województwie mazowieckim.

4. Zakończenie

Artykuł miał za zadanie zbadać atrakcyjność rynku lokali mieszkalnych w powiatach województwa. Mnogość zastosowanych w analizie aspektów opisujących atrakcyjność powiatów sprawia, że każda osoba znająca swoje zapotrzebowania i możliwości finansowe będzie w stanie znaleźć dla siebie odpowiednie miejsce zamieszkania, które spełni jej wymagania i oczekiwania.

W artykule odpowiedziano na postawione pytania badawcze, jednak wybór zmiennych zastosowanych w badaniu może mieć duży wpływ na końcowe wnioski. Przeprowadzona analiza wykazała, że miasto stołeczne Warszawa pod wytypowanymi cechami wyróżnia się (jest obserwacją odstającą) na tle pozostałych powiatów województwa mazowieckiego. Z tego powodu przypadek stolicy Polski należy poddać oddzielnej ocenie ze względu na znacząco różną specyfikę rynku lokali mieszkalnych. Ponadto wykonanie analizy głównych składowych potwierdziło, że powiaty sąsiadujące ze stolicą Polski charakteryzują się specyficzną strukturą atrakcyjności rynku mieszkalnego (rys. 2). Obszar ten cechuje się dużą liczbą nieruchomości mieszkalnych oddanych do użytkowania w 2020 roku na tle pozostałych powiatów województwa mazowieckiego. Powiaty otaczające Warszawę zaspokajają głównie popyt na tańsze mieszkania dla osób chcących pracować w stolicy Mazowsza.

Literatura

- Adil, I., i Irshad, A. (2015). A Modified Approach for Detection of Outliers. *Pakistan Journal of Statistics and Operation Research*, 11(1). <https://doi.org/10.18187/pjsor.v11i1.500>
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Batty, M., i Yichun, X (1998). Self-Organized Criticality and Urban Development. *Discrete Dynamics in Nature and Society*, 3(2-3), 109-124.
- Dąbrowska-Milewska, G. (2007). Propozycja klasyfikacji standardu zabudowy mieszkaniowej wielorodzinnej w odniesieniu do zasobów powstałych w Polsce po roku 1990. *Problemy Rozwoju Miast*, (3), 56-65.

- Dąbrowski, S. (2018). *Wpływ struktury krajobrazu na atrakcyjność obszaru zabudowy mieszkalnej w Poznaniu*. [Rozprawa doktorska, Uniwersytet Im. Adama Mickiewicza w Poznaniu]. Repozytorium AMU. <https://hdl.handle.net/10593/24495>
- Długosz, M. M. (2014). *Laboratorium XI: Analiza składowych głównych*. Pobrano 3 stycznia 2023 z http://home.agh.edu.pl/~mmd/_media/dydaktyka/adp/analiza_skladowych_glownych.pdf
- Główny Urząd Statystyczny. [GUS]. (2023). *Bank Danych lokalnych*. Pobrano 15 marca 2023 z <https://bdl.stat.gov.pl/bdl/dane/podgrup/temat>
- Główny Urząd Statystyczny. (b.d.). *Wskaźnik wykrywalności sprawców przestępstw*. Pobrano 4 stycznia 2023 z <https://stat.gov.pl/metainformacje/slownik-pojec/pojecia-stosowane-w-statystyce-publicznej/3853,pojecie.html>
- Gov.pl. (2023). *Polityka mieszkaniowa*. Pobrano 7 marca 2023 z <https://www.gov.pl/web/rozwoj-technologie/polityka-mieszkaniowa>
- House, J. S., Kessler R. C., Herzog, A. R., Mero, R. P., Kinney, A. M., i Breslow, M. J. (2013). Social Stratification, Age, and Health. *Aging, Health Behaviors, and Health Outcomes*, 13-44.
- Jałowiecki, B. (2001). *Spółeczna przestrzeń metropolii*. Wydawnictwo Scholar.
- Kowalik, J. i Gruszewska, E. (red.). (2015). Innowacyjność gospodarki polskiej na tle innych państw UE. W: E. Gruszewska (red.), *Współczesne wyzwania rozwoju gospodarczego: polityka i kreacja potencjału. Kreacja – innowacyjność – handel zagraniczny*, cz. 1 (s. 143-157), Wydawnictwo Uniwersytetu w Białymstoku.
- Kucharska-Stasiak, E. (2006). *Nieruchomość w gospodarce rynkowej*. Wydawnictwo Naukowe PWN.
- Norusis, M. J. (1994). *SPSS Professional Statistics, Version 6.1 (SPSS for Windows 6.1)*. SPSS Inc.
- Nove Olendry, (b.d). *Zalety mieszkania w mieście*. Pobrano 13 marca 2023 z <https://www.noveolendry.pl/news/zalety-mieszkania-w-miescie/>
- Plebankiewicz, E. i Biel, S. (2017). Analiza preferencji potencjalnych nabywców nieruchomości mieszkaniowych w Krakowie. *Świat Nieruchomości – World of Real Estate Journal*, 4(102), s. 55-60. <https://doi.org/10.14659/worej.2017.102.07>
- Policja.pl. (2019). *Bezpieczeństwo seniorów*. Pobrano 13 marca 2023 z <https://policja.pl/pol/kgp/biu-ro-prewencji/wydzial-profilaktyki-sp/bezpieczenstwo-osob-sta/174854,BEZPIECZENSTWO-SENIOROW.html>
- PQStat – Baza Wiedzy. (2022). *Zasadność stosowania Analizy składowych głównych*. Pobrano 1 marca 2023 z <http://manuals.pqstat.pl/statpqpl:redpl:pcapl:zasadnpl>
- Słodczyk, J. i Rajchel, D. (2004). *Przemiany demograficzne i jakość życia ludności miast*. Wydawnictwo Uniwersytetu Opolskiego.
- Tarka, D. i Olszewska, A. M. (2018). *Elementy statystyki, Opis statystyczny*. Oficyna Wydawnicza Politechniki Białostockiej
- Ustawa z dnia 24 czerwca 1994 r. o własności lokali (Dz. U. z 1994 r. Nr 85, poz. 388).
- Welcome Poland 2. (2022). *Czym charakteryzuje się warszawski rynek nieruchomości?* Pobrano 5 stycznia 2023 z <https://welcome2poland.eu/artukul/czym-charakteryzuje-sie-warszawski-rynek-nieruchomosci-173415>
- Zawierucha, S. (2012). *Struktura funkcjonowania gmin wiejskich a możliwości ich rozwoju*. [Rozprawa doktorska, Uniwersytet Ekonomiczny we Wrocławiu]. Pobrano z <https://www.wir.ue.wroc.pl/docstore/download.seam;jsessionid=EE13E995E08ACDF0EA9BAB7F1786A0C3?fileId=WUTbc3e20a9e502432c8568c66afddd56dc>
- Żelazowski, K. (2013). Ocena atrakcyjności rynków mieszkaniowych w Polsce. *Acta Universitatis Lodzianis. Folia Oeconomica*, (289), 115-127.

Analysis of the Attractiveness of the Housing Market in the Masovian Voivodeship

Abstract: Choosing a place of residence is a dilemma that everyone must face. The final decision on the location of a residential property affects our daily lives and should therefore be well thought out. This article focuses on analysing the housing real estate market in one of the most developed regions of Poland. The aim of the article is to examine the attractiveness of the residential property market in the counties of the Masovian Voivodeship. The conducted research is intended to help potential homebuyers in the Masovian Voivodeship choose the right place to live. Factors influencing the attractiveness of the residential property market are characterised, and counties with similar attractiveness characteristics are identified using principal component analysis. Counties near Warsaw generally have a better real estate market situation. Plock and Siedlce can be considered attractive locations for those wishing to live in a city within the Masovian Voivodeship other than Warsaw.

Keywords: Principal Component Analysis, analysis of the attractiveness, residential real estate market, Masovian Voivodeship

Julia Mikołajczak

e-mail: 182011@student.ue.wroc.pl

ORCID: 0009-0006-9886-4377

Uniwersytet Ekonomiczny we Wrocławiu

Analiza jakości zarządzania budżetami i funduszami unijnymi w województwach w Polsce w 2020 roku

DOI: 10.15611/2024.53.6.09

JEL Classification: H7, R50

© 2024 Julia Mikołajczak

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Mikołajczak, J. (2024). Analiza jakości zarządzania budżetami i funduszami unijnymi w województwach w Polsce w 2020 roku. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 129-145). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Celem badania jest przeprowadzenie wielowymiarowej analizy porównawczej (WAP) jakości zarządzania budżetem i funduszami unijnymi w szesnastu jednostkach samorządu terytorialnego w Polsce w roku 2020. WAP służy, m.in. do wykrywania prawidłowości w zbiorowościach. Analiza ta pozwoli zidentyfikować te województwa, które najlepiej zarządzają swoim budżetem, oraz te, które radzą sobie pod tym względem najgorzej. W opracowaniu przeanalizowano rolę wydatków i przychodów z wyszczególnieniem funduszy unijnych w budżetach województw. Ponadto celem jest zbadanie zależności pomiędzy analizowanymi zmiennymi oraz redukcja wymiarowości zbioru danych. Analizę wykonano z wykorzystaniem kilku metod, w tym miar zależności i analizy składowych głównych (PCA). Na podstawie uzyskanych wyników stwierdzono, że pomiędzy badanymi zmiennymi występują zależności, a ich strukturę przedstawiono za pomocą trzech składowych głównych. Najlepiej wykorzystującym swój budżet województwem okazało się dolnośląskie, które w 2020 roku otrzymało najwięcej środków unijnych, zaś najgorzej wypadło mazowieckie. Zidentyfikowano również województwa, w których zarządzanie budżetem nie okazało się tak efektywne, jak można by się spodziewać. Wśród takich regionów znalazło się, m.in. województwo lubelskie, które posiada największe zadłużenie w Polsce.

Słowa kluczowe: budżet województwa, zarządzanie budżetem, analiza głównych składowych, analiza zależności, fundusze europejskie

1. Wstęp

Budżetem województwa określić można plan finansowy, który opisuje przewidywane przychody i wydatki województwa. Jest on podstawą funkcjonowania samodzielnej gospodarki finansowej każdego z województw w państwie polskim. Tema-

tem niniejszego artykułu jest wielowymiarowa analiza porównawcza jakości zarządzania budżetami oraz funduszami unijnymi w każdej z szesnastu jednostek terytorialnych, jakimi są województwa. Celem tej analizy jest zbadanie zależności pomiędzy wybranymi zmiennymi dotyczącymi przychodów i wydatków województw oraz redukcja wymiarów wyjściowego zbioru danych do mniejszego, który pozwoli na przedstawienie danych w graficznej formie.

Ponadto oprócz celu głównego wyznaczone zostały również cele szczegółowe, a wśród nich znalezienie województwa, które najlepiej zarządza swoim budżetem i tego, które radzi sobie z tym najgorzej. Przed przeprowadzeniem analizy autorka stawia hipotezę, iż w województwie mazowieckim zarządzanie budżetem jest najbardziej efektywne oraz że to województwo dostaje najwięcej środków unijnych w związku ze znajdującą się na jego terenie stolicą – Warszawą. Drugą z hipotez jest to, iż można zauważyć znaczącą różnicę w wielkości zasobów oraz środków europejskich między województwami znajdującymi się w Polsce Wschodniej i Zachodniej, czyli tzw. Polską „A” i Polską „B”, co może wynikać z historycznych uwarunkowań, np. zaborów. Obie hipotezy poddane zostaną weryfikacji w dalszej części analizy. Można przypuszczać również, iż między przychodami a wydatkami województw zachodzi bezpośredni wpływ, tj. im większe są przychody, tym większe są średnie wydatki. W związku z powyższym sformułowano także kilka pytań badawczych, które brzmią następująco:

- Jakie województwo zarządza swoim budżetem oraz środkami unijnymi najbardziej efektywnie?
- Czy istnieją silne zależności między badanymi zmiennymi?
- Czy lepiej rozwinięte regiony w kraju otrzymują więcej środków unijnych?

2. Pojęcie budżetu województwa

Budżet województwa to plan finansowy, który określa przewidywane przychody i wydatki województwa. Według art. 134 ust. 1 Statutu Województwa Dolnośląskiego jest to podstawa funkcjonowania samodzielnej gospodarki finansowej każdego z województw (ISAP, b.d.). Według Ustawy z dnia 5 czerwca 1998 r. o samorządzie województwa budżet województwa podzielić można na cztery elementy:

- planowane wydatki i przychody,
- źródła finansowanie deficytu budżetowego,
- przeznaczenie nadwyżki budżetowej,
- wydatki związane z wieloletnimi programami.

Celem budżetu województwa jest zapewnienie jasnego zrozumienia sytuacji finansowej regionu, umożliwienie podejmowania świadomych decyzji i pomaganie mu w osiągnięciu stabilności i wzrostu gospodarczego.

Budżet województwa uchwalany jest jako część uchwały budżetowej na rok budżetowy przez sejmik województwa. Zarząd województwa odpowiada za realizację jego uchwał oraz zarządzanie finansami województwa. Jego zadaniem jest również

prezentowanie informacji o realizacji budżetu w danym roku w formie Opisowego Sprawozdania z Wykonania Budżetu Województwa. W związku z brakiem formalnego wzoru sprawozdanie może wyglądać inaczej w każdym województwie (Nagórek, 2023). Priorytetem tego dokumentu jest przedstawienie działań, które zostały podjęte, aby osiągnąć założone cele, oraz weryfikacja osiągniętych efektów. W skład sprawozdania z budżetu wchodzi, m.in.:

- dochody budżetowe,
- wydatki budżetowe,
- zadania z zakresu administracji rządowej oraz inne zadania zlecone ustawami realizowane przez województwa,
- realizacja planów inwestycyjnych,
- przychody i rozchody.

2.1. Dochody województw

Według Głównego Urzędu Statystycznego [GUS], zgodnie z Ustawą z dnia 27 sierpnia 2009 r. o finansach publicznych, w skład dochodów budżetów województw wchodzi:

- dochody własne, które pochodzą, m.in. z podatków od nieruchomości, udziałów z podatków dochodowych od osób fizycznych oraz prawnych, opłat skarbowych czy też zysków ze spadków, zapisów i darowizn na rzecz jednostek samorządu terytorialnego [jst];
- dochody majątkowe, które obejmują środki przeznaczone na inwestycje, oraz dochody ze sprzedaży majątku;
- dotacje, w skład których wchodzi wszelka, bezzwrotna pomoc finansowa udzielona przez państwo, Unię Europejską [UE], przedsiębiorstwa, organizacje pozarządowe czy też osoby prywatne;
- subwencja ogólna, czyli świadczenia państwa na rzecz jst;
- inne dochody należne województwu.

2.2. Fundusze unijne

Słownik Języka Polskiego PWN definiuje słowo „fundusz” jako wszelkie pieniądze, które są gromadzone z przeznaczeniem na dany cel, bądź jako instytucję, która jest powołana do zarządzania takimi pieniędzmi.

Unia Europejska jako organizacja międzynarodowa finansuje działania dążące do zmniejszania różnic w rozwoju między państwami członkowskimi. Środki z jej budżetu stanowią duży procent wszystkich dochodów województw w Polsce. Zgodnie z informacjami zawartymi na Portalu Funduszy Europejskich (2022) Polska była w latach 2014-2020 największym beneficjentem funduszy unijnych i otrzymała w sumie 82,5 mld euro. Pieniądze te miały pomóc zwiększyć konkurencyjność gospodarki w Polsce w ramach Strategii Europa 2020, której głównym celem jest zapewnienie

krajom UE inteligentnego i zrównoważonego rozwoju społeczno-gospodarczego (Gasz, 2014).

2.2.1. Struktura funduszy unijnych

Zgodnie z założeniami Strategii Europa 2020 funkcjonuje pięć głównych funduszy europejskich, którymi UE wspiera rozwój gospodarczy i społeczny państw członkowskich. Należą do nich (Gov.pl, 2021):

- Europejski Fundusz Rozwoju Regionalnego, którego głównym celem jest wyrównywanie poziomu rozwoju wszystkich regionów UE;
- Fundusz Spójności, który przeznaczony jest jedynie dla tych państw członkowskich, których dochód narodowy brutto (na mieszkańca) jest mniejszy niż 90% średniej w UE (w ramach tego funduszu realizowane są duże inwestycje, głównie z zakresu infrastruktury oraz ochrony środowiska);
- Europejski Fundusz Społeczny, którego najważniejszym zadaniem jest zwalczanie bezrobocia;
- Europejski Fundusz Rolny na rzecz Rozwoju Obszarów Wiejskich, którego celem jest wspomaganie obszarów wiejskich i zmiana struktury rolnictwa;
- Europejski Fundusz Morski i Rybacki, którego zadaniem jest restrukturyzacja rybołówstwa.

Rzeczpospolita Polska korzysta ze wszystkich wymienionych powyżej funduszy, ponieważ poziom PKB na mieszkańca w Polsce wynosi zaledwie 70% średniego poziomu UE (Gov.pl, b.d.).

2.2.2. Programy krajowe i regionalne

Za wdrażanie Funduszy Europejskich w Polsce w formie programów krajowych odpowiada przede wszystkim Minister Funduszy i Polityki Regionalnej, zaś jeśli chodzi o programy regionalne, jest to zadanie zarządu województwa.

Do programów krajowych należą: Program Infrastruktura i Środowisko, Program Inteligentny Rozwój, Program Wiedza Edukacja Rozwój, Program Polska Cyfrowa, Program Polska Wschodnia oraz Program Pomoc Techniczna.

Programy regionalne w odróżnieniu od krajowych mogą być znacznie bardziej zróżnicowane. Jest to zależne, m.in. od bieżących potrzeb danego regionu oraz jego budżetu.

2.3. Wydatki województw

Według GUS-u, zgodnie z Ustawą z dnia 27 sierpnia 2009 r. o finansach publicznych, w skład wydatków budżetów województw zalicza się:

- wydatki bieżące, których największą część stanowią świadczenia na rzecz osób fizycznych, wynagrodzenia oraz wydatki związane z zakupem materiałów i usług;

- wydatki majątkowe pochodzące głównie ze środków przeznaczonych na inwestycje oraz pieniądze na zakup i objęcie akcji;
- wydatki na obsługę długu publicznego;
- wydatki z dotacji ogółem, które obejmują wszelkie środki przeznaczone na cele inwestycyjne długookresowe lub krótkookresowe.

3. Opis zastosowanych metod

3.1. Miary zależności

W celu zbadania występowania zależności między zmiennymi wyznaczony zostanie współczynnik korelacji liniowej Pearsona, zgodnie z następującym wzorem (Ręklewski, 2020):

$$r = \frac{\text{cov}(X, Y)}{s_x * s_y} \quad (1)$$

gdzie $\text{cov}(X, Y)$ oznacza kowariancję między X a Y , natomiast s_x, s_y są odchyleniami standardowymi tych zmiennych.

3.2. Analiza głównych składowych

Analiza głównych składowych (ang. *Principal Component Analysis* [PCA]) jest metodą statystyczną, której głównym celem jest redukcja wymiarowości zbioru danych składającego się z dużej liczby powiązanych ze sobą zmiennych. Redukcja ta następuje poprzez przeprowadzenie transformacji oryginalnych zmiennych na nowe, nazywane głównymi składowymi przy zachowaniu jak największej ilości informacji, które niósł ze sobą pierwotny zbiór danych (Jolliffe, 2002). Nowo powstałe główne składowe są kombinacjami liniowymi badanych zmiennych i powinny być ortogonalne. Oznacza to, iż są one ze sobą nieskorelowane oraz uporządkowane w taki sposób, aby pierwsza główna składowa wyjaśniała największą część łącznej wariancji wszystkich oryginalnych zmiennych. Każda zaś kolejna składowa powinna być ortogonalna w stosunku do poprzedniej. Głównych składowych, koniecznych do wyjaśnienia całej wspólnej wariancji, może być maksymalnie tyle, ile jest wyjściowo zmiennych w zbiorze danych (Górniak, 1998). Analiza głównych składowych pozwala na, m.in.:

- redukcję liczby zmiennych;
- zbadanie i zweryfikowanie struktury i prawidłowości między zmiennymi;
- klasyfikację obiektów w nowych przestrzeniach zdefiniowanych przez utworzone czynniki;
- graficzną prezentację struktury wielowymiarowego zbioru danych na płaszczyźnie. Analizę głównych składowych wykonuje się dla zmiennych podanych co najmniej na skali przedziałowej oraz dla zmiennych, między którymi zachodzą związki

liniowe (Górniak, 1998). Korelacja między wyjściowymi zmiennymi powinna być badana za pomocą współczynnika korelacji r -Pearsona (Kim i Mueller, 1978). Analizie głównych składowych, w największej liczbie przypadków, poddaje się zmienne standaryzowane. Im większe są korelacje między zmiennymi, tym większe jest uzasadnienie stosowania analizy głównych składowych.

Istnieje bardzo dużo różnych kryteriów, które stosuje się przy podejmowaniu decyzji o liczbie głównych składowych w analizie. Wśród nich znajdują się, m.in.: kryterium wartości własnej Keisera, kryterium Jolliffe'a, kryterium wystarczającej proporcji wyjaśnionej wariancji i kryterium ospyska. Przed podjęciem decyzji o liczbie głównych składowych można sugerować się jednym bądź kilkoma kryteriami naraz.

4. Wyniki badania

Efektem końcowym badań jest udzielenie odpowiedzi, na zawarte we wstępie pytania, a także weryfikacja postawionych hipotez badawczych.

Wykorzystane w analizie dane zaczerpnięte są ze strony GUS, a konkretnie z Bazy Danych Lokalnych (<https://bdl.stat.gov.pl/bdl/start>). Zbiór danych zawiera obserwacje, które pochodzą z 2020 roku. W analizie wykorzystano dziewięć zmiennych dotyczących dochodów i wydatków każdego z województw na terenie Polski.

Zmienne poddane analizie:

- X_1 – dochody majątkowe ogółem [zł],
- X_2 – dochody własne razem [zł],
- X_3 – dotacje ogółem [zł],
- X_4 – subwencja ogólna razem [zł],
- X_5 – środki z UE na finansowanie programów i projektów unijnych [zł],
- X_6 – wydatki bieżące ogółem [zł],
- X_7 – wydatki na obsługę długu ogółem [zł],
- X_8 – wydatki z dotacji ogółem [zł],
- X_9 – wydatki majątkowe ogółem [zł].

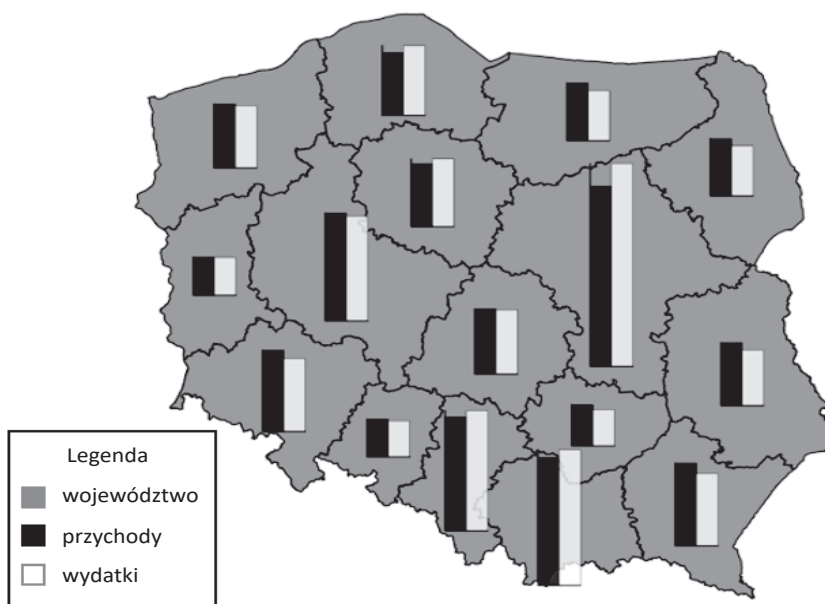
4.1. Prezentacja zmiennych

W tabeli 1 przedstawione zostały zsumowane przychody (zmienne X_1, X_2, X_3, X_4, X_5) oraz wydatki (zmienne X_6, X_7, X_8, X_9) województw w Polsce w 2020 roku, a także ich wynik finansowy oraz procent udziału, jaki stanowiły środki unijne w całym dochodzie każdego województwa w 2020 roku, zaś na rysunku 1 przedstawiona została wizualizacja wysokości przychodów i wydatków w każdym z województw.

Tabela 1. Budżet województw w 2020 roku

| Województwo | Suma przychodów [zł] | Suma wydatków [zł] | Udział środków z UE w przychodach [%] | Wynik finansowy [zł] |
|---------------------|----------------------|--------------------|---------------------------------------|----------------------|
| dolnośląskie | 1 689 204 087,72 | 1 512 925 059,84 | 6 | 176 279 027,88 |
| kujawsko-pomorskie | 1 307 292 507,65 | 1 404 576 568,06 | 3 | -97 284 060,41 |
| lubelskie | 1 304 963 671,19 | 1 153 200 826,80 | 4 | 151 762 844,39 |
| lubuskie | 794 124 422,55 | 783 784 495,60 | 3 | 10 339 926,95 |
| łódzkie | 1 356 204 255,10 | 1 338 322 206,50 | 4 | 17 882 048,60 |
| małopolskie | 2 658 923 533,58 | 2 824 281 360,08 | 4 | -165 357 826,50 |
| mazowieckie | 3 744 726 490,49 | 4 212 894 791,71 | 2 | -468 168 301,22 |
| opolskie | 786 407 432,25 | 745 781 262,51 | 4 | 40 626 169,74 |
| podkarpackie | 1 721 214 850,07 | 1 502 711 543,05 | 4 | 218 503 307,02 |
| podlaskie | 1 190 642 742,41 | 1 037 567 552,04 | 3 | 153 075 190,37 |
| pomorskie | 1 312 382 055,91 | 1 453 686 235,30 | 3 | -141 304 179,39 |
| śląskie | 2 364 151 222,39 | 2 498 288 961,64 | 3 | -134 137 739,25 |
| świętokrzyskie | 854 500 914,16 | 752 089 329,12 | 3 | 102 411 585,04 |
| warmińsko-mazurskie | 1 197 550 898,16 | 1 037 344 100,20 | 3 | 160 206 797,96 |
| wielkopolskie | 2 240 990 902,23 | 2 175 018 857,41 | 2 | 65 972 044,82 |
| zachodniopomorskie | 1 329 271 247,05 | 1 296 529 300,37 | 4 | 32 741 946,68 |

Źródło: opracowanie własne na podstawie danych z GUS.

**Rys. 1.** Kartodiagram dla przychodów i wydatków w województwach

Źródło: opracowanie własne na podstawie danych z GUS.

Jak widać w tabeli 1 oraz na rysunku 1 największe przychody osiągają województwa mazowieckie, małopolskie, a także wielkopolskie, najmniejszą zaś ilością środków charakteryzuje się województwo opolskie. Województwa podkarpackie oraz dolnośląskie osiągnęły w 2020 roku najlepszy wynik finansowy. To właśnie w województwie dolnośląskim udział środków unijnych okazał się być największy i wyniósł aż 6% wszystkich dochodów w badanym roku. Najgorzej, jeśli chodzi o wynik finansowy, wypada województwo mazowieckie, zaś po nim kolejno województwa małopolskie i pomorskie. W tych województwach można było zaobserwować ujemny wynik finansowy, a zatem w 2020 roku przyniosły one straty. Co zaskakujące, najmniejszym udziałem środków europejskich wśród przychodów w całym roku pochwalić się mogą województwa mazowieckie i wielkopolskie, gdzie ilość tych środków oscyluje na poziomie ok. 2%.

4.2. Analiza zależności

Pomiędzy wszystkim analizowanymi zmiennymi wyznaczone zostały współczynniki korelacji liniowej Pearsona. Zależności między wszystkimi zmiennymi zostały przedstawione w formie macierzy korelacji w tabeli 2.

Tabela 2. Macierz korelacji

| Zmienna | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | X_9 |
|---------|-------------|-------------|-------------|-------|-------------|-------------|-------|-------------|-------------|
| X_1 | 1,00 | 0,07 | 0,84 | -0,07 | 0,51 | 0,18 | -0,08 | 0,44 | 0,74 |
| X_2 | 0,07 | 1,00 | 0,31 | -0,50 | 0,35 | 0,97 | 0,51 | 0,84 | 0,69 |
| X_3 | 0,84 | 0,31 | 1,00 | -0,27 | 0,64 | 0,46 | 0,01 | 0,74 | 0,76 |
| X_4 | -0,07 | -0,50 | -0,27 | 1,00 | -0,14 | -0,43 | 0,04 | -0,52 | -0,34 |
| X_5 | 0,51 | 0,35 | 0,64 | -0,14 | 1,00 | 0,44 | 0,21 | 0,60 | 0,48 |
| X_6 | 0,18 | 0,97 | 0,46 | -0,43 | 0,44 | 1,00 | 0,48 | 0,91 | 0,74 |
| X_7 | -0,08 | 0,51 | 0,01 | 0,04 | 0,21 | 0,48 | 1,00 | 0,30 | 0,21 |
| X_8 | 0,44 | 0,84 | 0,74 | -0,52 | 0,60 | 0,91 | 0,30 | 1,00 | 0,83 |
| X_9 | 0,74 | 0,69 | 0,76 | -0,34 | 0,48 | 0,74 | 0,21 | 0,83 | 1,00 |

Źródło: Opracowanie własne na podstawie danych z GUS.

Można zaobserwować, iż przedstawiona w tabeli 2 macierz korelacji jest bardzo zróżnicowana pod względem występujących w niej wartości. Korelacje wahają się od 0,01 do nawet 0,97. Wszystkie pogrubione wartości przedstawiają tzw. silne, czyli takie, których wartość bezwzględna współczynnika r zawiera się w przedziale od 0,6 do 1. Najniższe zależności można zaobserwować dla zmiennej X_7 z pozostałymi zmiennymi, czyli dla wydatków na obsługę długu ogółem. Jej korelacje ze zmiennymi X_3 (dotacje ogółem) i X_4 (subwencja ogólna razem) wynoszą kolejno 0,01 i 0,04. Oznacza to w zasadzie brak zależności pomiędzy tymi badanymi zmiennymi. Najbardziej skorelowane ze sobą są zmienne X_6 (wydatki bieżące ogółem) i X_2 (dochody

własne razem), a także X_6 z X_8 (wydatki z dotacji ogółem). Współczynnik korelacji dla tych zmiennych wynosi kolejno 0,97 i 0,91. Można zatem powiedzieć, że występuje między tymi zmiennymi bardzo silna zależność, a wraz ze wzrostem wydatków bieżących ogółem będą rosły średnie dochody własne razem oraz średnie wydatki z dotacji ogółem. Warto również wspomnieć, że zmienna X_4 (subwencja ogólna razem) jest skorelowana ujemnie z każdą pozostałą zmienną z wyjątkiem zmiennej X_7 (wydatki na obsługę długu ogółem). Oznacza to, że jeśli subwencja ogólna będzie rosła, to maleć będą średnie wartości pozostałych zmiennych z nią skorelowanych.

4.3. Analiza głównych składowych

4.3.1. Ocena wartości własnych na podstawie kryteriów

Celem analizy głównych składowych jest, m.in. redukcja liczby zmiennych, które opisują dane zjawisko, oraz zbadanie, jakie prawidłowości zachodzą między tymi zmiennymi. W związku z powyższym obliczone zostały wartości własne macierzy korelacji i procent wyjaśnianej wariancji, które są przedstawione w tabeli 3.

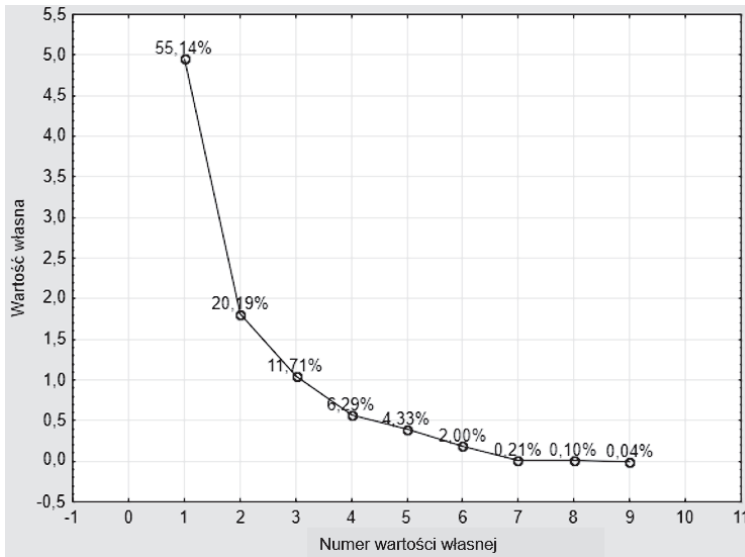
Tabela 3. Wartości własne macierzy korelacji

| Numer wartości | Wartości własne | Procent ogółu wariancji | Skumulowane wartości własne | Skumulowany procent ogółu wariancji |
|----------------|-----------------|-------------------------|-----------------------------|-------------------------------------|
| 1 | 4,96 | 55,14 | 4,96 | 55,14 |
| 2 | 1,82 | 20,19 | 6,78 | 75,32 |
| 3 | 1,05 | 11,71 | 7,83 | 87,03 |
| 4 | 0,57 | 6,29 | 8,40 | 93,32 |

Źródło: opracowanie własne na podstawie danych z GUS.

Wartości własne macierzy korelacji są niezbędne do wyboru liczby głównych składowych. Biorąc pod uwagę kryterium obiektywne, należy wybrać tyle głównych składowych, dla ilu wartości własne są większe od 1. W tym przypadku pierwsze trzy wartości są większe od 1, wskazane jest zatem zredukowanie badanej przestrzeni do trzech wymiarów (3W).

Według kryterium subiektywnego należy wyznaczyć tyle głównych składowych, ile wyjaśnia ustalony procent ogólnej wariancji (Górniak, 1998). Wysokość tego procentu zależy od rodzaju badanych zmiennych oraz ich dziedziny. Analizowane w tym przypadku zjawisko dotyczy finansów i ma charakter mierzalny, a zatem można przyjąć, że procent wyjaśnianej zmienności powinien wynosić około 85%. Jak widać w tabeli 3 pierwszy wymiar wyjaśnia 55,14%, drugi wyjaśnia 75,32%, a trzeci 87,03%. Zatem według kryterium subiektywnego sugerowane jest, że model powinien być odwzorowany przez trzy wymiary.



Rys. 2. Wykres osypiska (wykres wartości własnych)

Źródło: opracowanie własne na podstawie danych z GUS.

Trzecie kryterium odnosi się do wykresu osypiska, który został przedstawiony na rysunku 2. Kryterium osypiska (Cattella) sugeruje, iż należy zostawić tyle czynników, ile tworzy zbocze na wykresie, zaś nie należy brać pod uwagę tych, które tworzą tzw. osypisko. Wykres ten jednak potrafi być trudny do interpretacji. Na rysunku 2 widać, iż do analizy powinien być wybrany wymiar 2W lub 3W, aczkolwiek można zauważyć, że wyplaszczenie zaczyna się bardziej dla wymiaru trzeciego. Wykres ten jest niejednoznaczny, dlatego kryterium to nie będzie brane pod uwagę przy wyborze liczby głównych składowych.

Na podstawie kryteriów obiektywnego i subiektywnego, należy stwierdzić, iż trzeci wymiar najlepiej będzie odwzorowywał dane w modelu.

4.3.2. Analiza wektorów własnych

W następnym kroku analizy niezbędne jest wyznaczenie trzech głównych składowych. W tym celu najpierw wyznaczono wektory własne macierzy korelacji, które posłużą do określenia wag głównych składowych.

Główne składowe, które zostały wyznaczone za pomocą wektorów własnych macierzy korelacji, prezentują się następująco:

$$Z_1 = -0,26x_1 - 0,37x_2 - 0,35x_3 + 0,22x_4 - 0,30x_5 - 0,40x_6 - 0,16x_7 - 0,43x_8 - 0,41x_9,$$

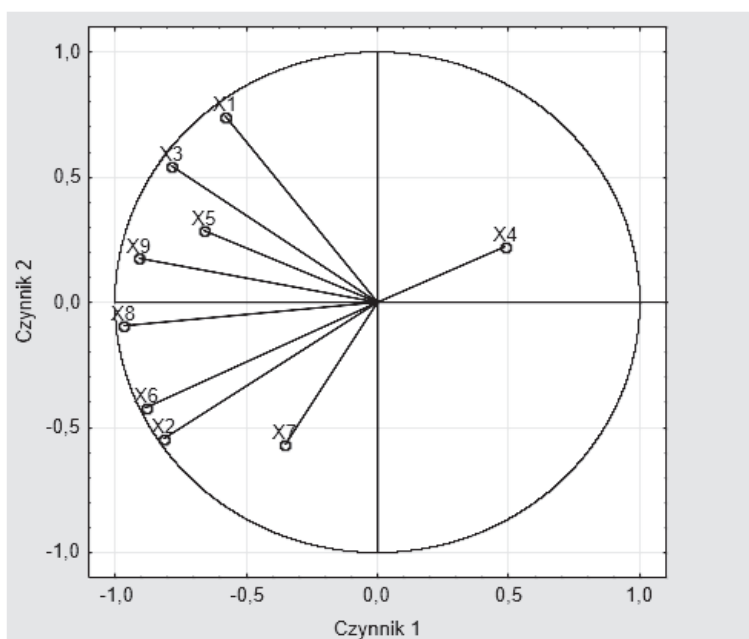
$$Z_2 = 0,55x_1 - 0,40x_2 + 0,41x_3 + 0,17x_4 + 0,21x_5 - 0,31x_6 - 0,42x_7 - 0,07x_8 + 0,13x_9,$$

$$Z_3 = 0,10x_1 - 0,05x_2 - 0,001x_3 + 0,72x_4 + 0,29x_5 + 0,01x_6 + 0,61x_7 - 0,08x_8 - 0,006x_9.$$

Zgodnie z definicją Górniaka (1999), przedstawione powyżej trzy główne składowe są liniową kombinacją badanych zmiennych.

4.3.3. Przedstawienie danych na wykresach

Graficzna prezentacja obliczonych wartości współczynnika koleracji pozwala na podział zmiennych na grupy, co zostało przedstawione na rysunku 3.



Rys. 3. Wykres klasyfikacji badanych zmiennych

Źródło: opracowanie własne na podstawie danych z GUS.

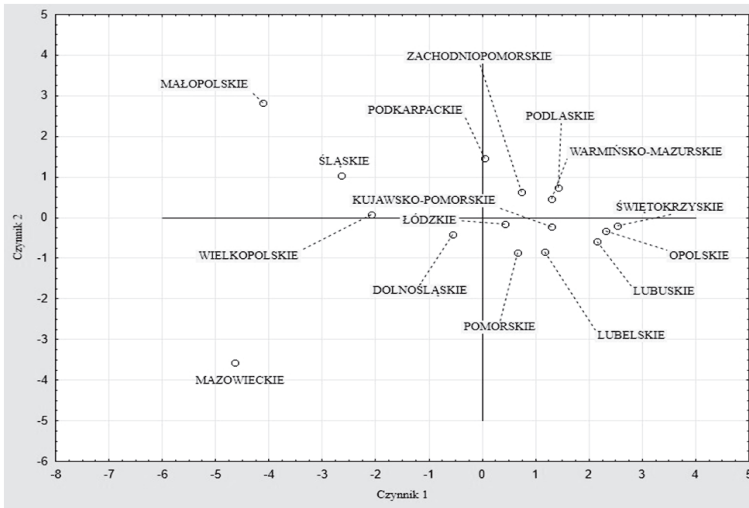
Na podstawie rysunku 3 można zauważyć, że wszystkie zmienne, z wyjątkiem zmiennej X_4 – subwencji ogólnej, ze względu na pierwszą główną składową (czynnik 1) są umieszczone po lewej stronie. Oznacza to, iż oprócz zmiennej X_2 , są one między sobą skorelowane ujemnie.

Ze względu na drugą główną składową, czyli czynnik 2, można dokonać podziału zmiennych na kilka grup:

- grupa 1: dochody własne, wydatki na obsługę długu, wydatki bieżące;
- grupa 2: wydatki z dotacji;
- grupa 3: wydatki majątkowe, subwencja ogólna, środki z UE na finansowanie programów i projektów unijnych;
- grupa 4: dochody majątkowe oraz dotacje ogółem.

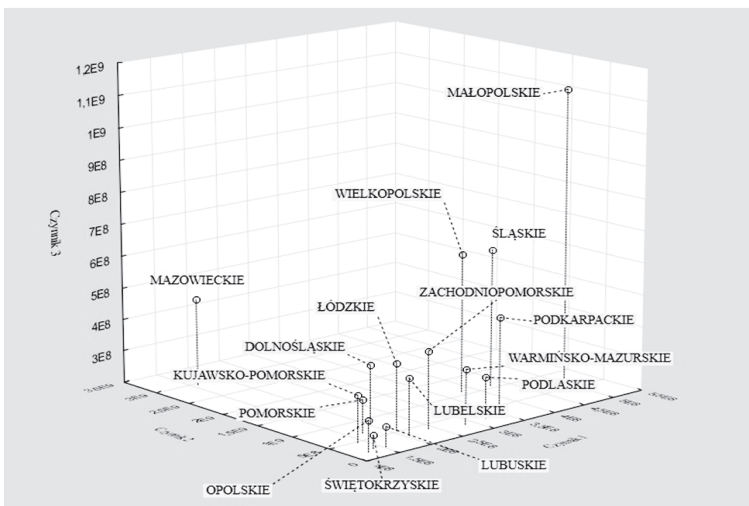
Podział ten świadczy o tym, że w województwach zmienne, które znajdują się w obrębie grupy, współwystępują ze sobą. Najdłuższe wektory widoczne na wykresie

się posiadają zmienne X_2 (dochody własne), X_6 (wydatki bieżące) oraz X_8 (wydatki z dotacji). Oznacza to, że ich stopień odtworzenia wariancji jest największy i są one bardzo dobrze reprezentowane przez dwie pierwsze główne składowe tworzące układ współrzędnych. Na rysunku 4 przedstawiony został rzut przypadków w przestrzeni dwuwymiarowej.



Rys. 4. Wykres przypadków w przestrzeni dwuwymiarowej

Źródło: opracowanie własne na podstawie danych z GUS.



Rys. 5. Wykres przypadków w przestrzeni trójwymiarowej

Źródło: opracowanie własne na podstawie danych z GUS.

Na rysunku 4 dane są odtworzone w 75,32%. Jednoznacznie można stwierdzić, że województwa mazowieckie i małopolskie są obserwacjami odstającymi. Na wykresie również znajduje się kilka innych punktów, które można podejrzewać, że są obserwacjami odstającymi. Takim punktem jest województwo śląskie o współrzędnych $(-2,65;1,03)$ oraz województwo wielkopolskie o współrzędnych równych $(-2,09;0,07)$.

Rysunek 5 przedstawia rzut punktów w przestrzeni trójwymiarowej.

W modelu trójwymiarowym procent odtworzenia danych wynosi 87,03%, czyli zwiększył się o prawie 12 p.p. w stosunku do przestrzeni dwuwymiarowej. Tym razem również można zaobserwować wyraźne odstawanie, względem 1 i 3 czynnika, województwa małopolskiego od innych obserwacji. Innym, znacząco odstającym (w stosunku do 2 czynnika) od reszty województwem jest mazowieckie. Podobnie jak w przypadku wykresu 2W województwo śląskie oraz wielkopolskie można uznać za obserwacje odstające.

4.3.4. Ocena odtworzenia jakości punktów

W tabeli 4 przedstawione zostały zasoby zmienności wspólnej dla wszystkich zmiennych. Informują nas one o tym, jak każda z głównych składowych określa część opisaną wariancji, a także, jaka część informacji została utracona poprzez redukcję liczby wymiarów danych.

Tabela 4. Zasoby zmienności wspólnej

| Zmienna | Czynnik 1 | Czynnik 2 | Czynnik 3 | Czynnik 4 | Czynnik 5 | Czynnik 6 | Czynnik 7 | Czynnik 8 | Czynnik 9 |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| X_1 | 0,339 | 0,893 | 0,904 | 0,946 | 0,980 | 0,994 | 0,999 | 1,000 | 1,000 |
| X_2 | 0,664 | 0,962 | 0,964 | 0,970 | 0,986 | 0,996 | 0,998 | 0,998 | 1,000 |
| X_3 | 0,618 | 0,917 | 0,917 | 0,917 | 0,921 | 0,998 | 0,998 | 1,000 | 1,000 |
| X_4 | 0,238 | 0,288 | 0,837 | 0,877 | 0,998 | 1,000 | 1,000 | 1,000 | 1,000 |
| X_5 | 0,441 | 0,522 | 0,609 | 0,972 | 0,980 | 1,000 | 1,000 | 1,000 | 1,000 |
| X_6 | 0,776 | 0,953 | 0,953 | 0,958 | 0,993 | 0,995 | 0,998 | 0,998 | 1,000 |
| X_7 | 0,124 | 0,443 | 0,841 | 0,841 | 0,998 | 1,000 | 1,000 | 1,000 | 1,000 |
| X_8 | 0,938 | 0,946 | 0,953 | 0,955 | 0,969 | 0,993 | 0,995 | 1,000 | 1,000 |
| X_9 | 0,825 | 0,855 | 0,855 | 0,963 | 0,964 | 0,992 | 0,999 | 1,000 | 1,000 |

Źródło: opracowanie własne na podstawie danych z GUS.

Z tabeli 4 wynika, że pierwszy wymiar, wyznaczony na podstawie pierwszej głównej składowej, wyjaśnia zaledwie dwie zmienne (X_8, X_9) na wystarczającym poziomie. Przestrzeń dwuwymiarowa, choć wystarczająco wyjaśnia większość zmiennych, to jednak przy zmiennych X_4 (subwencja ogólna), X_5 (środki z Unii Europejskiej na finansowanie programów i projektów unijnych) oraz X_7 (wydatki na obsługę długu) to odtworzenie jest za niskie, ponieważ wynosi kolejno (29%, 52% i 44%). Patrząc natomiast na przestrzeń trójwymiarową, wyjaśnia ona zdecydowanie lepiej

wszystkie zmienne. Najstąbiej w tej przestrzeni, bo w 60% odtwarzana jest zmienna określająca środki z UE na finansowanie programów i projektów unijnych. Najlepiej odtwarzaną zmienną w 3W jest zmienna X_2 , czyli dochody własne.

W tabeli 5 przedstawiona została jakość odtworzenia punktów na rysunku w przestrzeni dwuwymiarowej i trójwymiarowej.

Tabela 5. Jakość odtworzenia punktów w 2W i 3W

| Województwo | Czynnik 1 | Czynnik 2 | Czynnik 3 | 2W | 3W |
|---------------------|-----------|-----------|-----------|--------|--------|
| dolnośląskie | 0,0470 | 0,0269 | 0,0519 | 0,0739 | 0,1258 |
| kujawsko-pomorskie | 0,4531 | 0,0127 | 0,0738 | 0,4658 | 0,5396 |
| lubelskie | 0,1145 | 0,0607 | 0,7554 | 0,1752 | 0,9306 |
| lubuskie | 0,7652 | 0,0567 | 0,1103 | 0,8219 | 0,9322 |
| łódzkie | 0,2110 | 0,0262 | 0,2299 | 0,2372 | 0,4671 |
| małopolskie | 0,6519 | 0,3047 | 0,0064 | 0,9566 | 0,9629 |
| mazowieckie | 0,6148 | 0,3650 | 0,0005 | 0,9798 | 0,9803 |
| opolskie | 0,6907 | 0,0147 | 0,2537 | 0,7055 | 0,9592 |
| podkarpackie | 0,0006 | 0,5095 | 0,2480 | 0,5101 | 0,7582 |
| podlaskie | 0,5311 | 0,1409 | 0,0010 | 0,6720 | 0,6730 |
| pomorskie | 0,1971 | 0,3234 | 0,4118 | 0,5205 | 0,9323 |
| śląskie | 0,8050 | 0,1210 | 0,0000 | 0,9259 | 0,9259 |
| świętokrzyskie | 0,8776 | 0,0060 | 0,0830 | 0,8835 | 0,9666 |
| warmińsko-mazurskie | 0,6395 | 0,0814 | 0,1109 | 0,7209 | 0,8318 |
| wielkopolskie | 0,6220 | 0,0006 | 0,0375 | 0,6226 | 0,6601 |
| zachodniopomorskie | 0,4246 | 0,3057 | 0,0201 | 0,7303 | 0,7504 |

Źródło: opracowanie własne na podstawie danych z GUS.

Na podstawie tabeli 5 można stwierdzić, że najgorsze odtworzenie punktów w przestrzeni dwuwymiarowej da się zauważyć dla województwa dolnośląskiego i lubelskiego (0,07 i 0,18). Dla porównania w przestrzeni trójwymiarowej jakość odtworzenia tych punktów wzrosła do kolejno 0,13 i 0,93.

Zarówno po zastosowaniu kryteriów subiektywnego oraz obiektywnego, a także po analizie jakości odtworzenia punktów, jak i zmiennych, wydaje się, że prawidłową decyzją jest wybór przestrzeni trójwymiarowej (trzech głównych składowych).

5. Wnioski

Po przeprowadzeniu analizy można przejść do weryfikacji zasadności postawionych wcześniej hipotez badawczych i podjąć próbę udzielenia odpowiedzi na pytania badawcze zawarte we wstępie.

Hipoteza dotycząca tego, iż w województwie mazowieckim jakość zarządzania budżetem jest najlepsza, została odrzucona. Okazuje się, iż choć województwo mazowieckie posiada największe środki, to jednak nie wykorzystuje ich najbar-

dziej efektywnie. Województwo to skończyło rok 2020 z najgorszym wynikiem finansowym i największą stratą spośród wszystkich województw w Polsce, a środki z UE stanowiły zaledwie 2% przychodów tego województwa. Większość środków z UE w 2020 roku nie było skierowanych do województwa stołecznego, tak jak podejrzewano, lecz największa ich część trafiła do województwa dolnośląskiego, w którym stanowiły one aż 6% wszystkich przychodów w tym roku. Województwami najlepiej wykorzystującymi fundusze są województwa podkarpackie i dolnośląskie, których zysk pod koniec roku 2020 był największy i wyniósł kolejno 218 503 307,02 i 176 279 027,88 zł. W kwestii rozporządzania środkami unijnymi okazuje się, że w większości trafiają one do regionów tzw. przejściowych, które mają potencjał na przekształcenie się w regiony lepiej rozwinięte. Oznacza to zatem, iż należy również odrzucić hipotezę, która stanowiła, iż środki unijne w województwie mazowieckim są największe, m.in. ze względu na znajdującą się tam stolicę Warszawę. Jako region silnie rozwinięty, województwo to nie potrzebuje większego wsparcia z UE, a środki europejskie kierowane są w większości do regionów słabo i średnio rozwiniętych.

Spodziewano się zauważyć bardzo silną zależność między przychodami a wydatkami, jednak korelacje między zmiennymi potrafiły się wahać od 0,04 do 0,84, nie jest możliwe zatem jednoznaczne stwierdzenie, że zależność między wydatkami a przychodami jest bardzo silna.

Jeśli chodzi o hipotezę, która dotyczy różnego stopnia wykorzystania środków między Polską „A” a „B”, to również należy ją odrzucić. Nie istnieje bezpośrednia zależność, która wskazywałaby na to, iż zachodnia część Polski ma do dyspozycji więcej środków czy też funduszy europejskich niż wschodnie województwa państwa. Zaobserwowano oczywiście różnice w wielkości przychodów czy też wydatków pomiędzy województwami w Rzeczypospolitej Polskiej, jednak nie są one uwarunkowane położeniem w zachodniej bądź wschodniej części kraju, lecz innymi czynnikami.

6. Zakończenie

Głównym celem niniejszej analizy było zbadanie zależności pomiędzy wybranymi zmiennymi dotyczącymi przychodów i wydatków województw. Przy użyciu analizy głównych składowych możliwa była redukcja wymiaru z dziewięciu badanych zmiennych do trzech głównych składowych. Przeprowadzona w tej pracy analiza umożliwiła identyfikację regionów (województw), w których zarządzanie budżetem wojewódzkim nie jest tak efektywne, jak powinno. Wśród takich regionów znalazło się, m.in. województwo lubelskie, które posiada największe spośród wszystkich województw w Polsce zadłużenie.

Przed przeprowadzeniem badań spodziewano się innych wyników i wniosków niż te, które otrzymano. W efekcie większość postawionych we wstępie hipotez została odrzucona, w tym hipoteza o najlepszej efektywności zarządzania budżetem

wojewódzkim w województwie mazowieckim czy też hipoteza dotycząca różnicy w wielkości dostępnych środków i funduszy pomiędzy tzw. Polską „A” a Polską „B”. Hipoteza, która okazała się być częściowo prawdziwa to ta, która mówi o istnieniu zależności pomiędzy badanymi zmiennymi, jednak te zależności nie były aż tak silne, jak autorka tego tekstu zakładała.

Literatura

- Bdl.stat.gov.pl. (b.d.). Bank Danych Lokalnych. Pobrano 29 grudnia 2022 z <https://bdl.stat.gov.pl/bdl/start>.
- Gasz, M. (2014). Strategia Europa 2020 – założenia i perspektywy realizacji. *Nierówności Społeczne a Wzrost Gospodarczy*, 38(2), 85-97.
- Główny Urząd Statystyczny. (b.d.). *Pojęcia stosowane w statystyce publicznej*. Pobrano 29 grudnia 2022 z <https://stat.gov.pl/metainformacje/slownik-pojec/pojecia-stosowane-w-statystyce-publicznej/2164,pojecie.html?contrast=default>
- Gov.pl. (b.d.). *Fundusze Europejskie 2014-2020*. Pobrano 22 kwietnia 2023 z <https://www.gov.pl/web/fundusze-regiony/dowiedz-sie-wiecej-o-funduszach-europejskich>
- Gov.pl. (2021). *Strategia Europa 2020*. Pobrano 14 lutego 2023 z <https://www.gov.pl/web/rozwoj-technologie/strategia-europa-2021>
- Górniak, J. (1998). Analiza czynnikowa i analiza głównych składowych. *ASK*, (7), 83-102.
- Górniak, J. (1999). *My i nasze pieniądze*. Wydawnictwo Aureus.
- Internetowy System Aktów Prawnych (ISAP). (b.d.). *Akty prawne według roczników*. Pobrano dnia 29 grudnia 2022 z <https://isap.sejm.gov.pl/isap.nsf/ByYear.xsp>
- Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition*. Springer Series in Statistics. Springer.
- Kim, J., i Mueller, C.W. (1978). Introduction to Factor Analysis: What it is and How to Do it. *Sage University Paper Series on Quantitative Applications in the Social Sciences*, (13). <https://doi.org/10.4135/9781412984652>
- Nagórek, M. (2023). *Jak sporządzić roczne sprawozdanie z wykonania budżetu JST*. Pobrano dnia 23 kwietnia 2023 z <https://serwisy.gazetaprawna.pl/samorzad/artykuly/8665129,jak-sporzadzic-roczne-sprawozdanie-z-wykonania-budzetu-jst.html>
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka. Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Portal Funduszy Europejskich. (2022, 12 stycznia). *Polska w czołówce państw UE pod względem inwestowania Funduszy Europejskich*. Pobrano 24 kwietnia 2023 z <https://www.funduszeuropejskie.gov.pl/strony/wiadomosci/polska-w-czolowce-panstw-ue-pod-wzglem-inwestowania-funduszy-europejskich/>
- Ręklewski, M. (2020). *Statystyka opisowa. Teoria i przykłady*. Wydawnictwo Państwowej Uczelni Zawodowej we Włocławku.
- Słownik Języka Polskiego. (b.d.). *Fundusz*. Pobrano 29 grudnia 2022 z <https://sjp.pwn.pl/sjp/fundusz;2558714.html>
- Statut Województwa Dolnośląskiego (Dziennik Urzędowy Województwa Dolnośląskiego z 2010 r. Nr 47, poz. 693)
- Ustawa z dnia 27 sierpnia 2009 r. o finansach publicznych (Dz. U. z 2009 r. Nr 157, poz. 1240)
- Ustawa z dnia 5 czerwca 1998 r. o samorządzie województwa (Dz. U. z 1998 r. Nr 91, poz. 576)

Analysis of the Quality of Budget Management and EU Funds in Voivodeships in Poland in 2020

Abstract: This study investigates the management of budgets and European funds in voivodships in Poland. The purpose of this paper is a multidimensional comparative analysis of the quality of budget management and EU funds in each of the sixteen local government units in Poland in the year 2020. The study analyses the role of expenditure and revenues with an emphasis on EU funds in voivodship budgets. Moreover, the aim is to study the relationship between the analysed variables. Several methods were used to carry out the analysis, including Pearson's correlation and principal component analysis. Based on the results, it was concluded that there are relationships between the variables under study and their structure was presented using three principal components. The voivodship that made the best use of its budget was Lower Silesia which received the largest amount of EU funds in the year 2020. On the other hand, the worst of all provinces in 2020 was the Masovian Voivodeship. There were some voivodeships in which the budget management did not turn out to be as effective as one might expect. Among such regions was Lublin Voivodeship, which has the highest debt in all of Poland.

Keywords: voivodship budget, management of budgets, PCA – Principal Components Analysis, correlation analysis, EU funds

Karol Noganowicz

e-mail: karol.noganowicz@gmail.com

Uniwersytet Ekonomiczny we Wrocławiu

Związek infrastruktury transportowej z rozwojem regionalnym

DOI: 10.15611/2024.53.6.10

JEL Classification: L91, O18, R11, R42

© 2024 Karol Noganowicz

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Noganowicz, K. (2024). Związek infrastruktury transportowej z rozwojem regionalnym. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 146-162). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Rozwój regionalny jest procesem zmian wpływających na rozwój gospodarczy kraju. Jakość wybranych elementów sfery publicznej często zależy od tego, w jakim stopniu rozwinięty jest ten obszar. W niniejszym artykule autor zbada związek pomiędzy poziomem rozwoju regionalnego a jakością infrastruktury transportowej i postara się ustalić stopień tego powiązania ze wskazaniem potencjalnych szans i zagrożeń. Opracowanie będzie miało charakter badawczy. Zostanie przeprowadzona analiza porównawcza województw w roku 2021 pomiędzy jakością infrastruktury transportowej a poziomem rozwoju regionalnego wykonana za pomocą metod Warda oraz TOPSIS. W podsumowaniu pracy zaprezentowane zostaną wnioski, które nie potwierdzą wyraźnej zależności pomiędzy dwoma wspomnianymi pojęciami.

Słowa kluczowe: infrastruktura transportowa, rozwój regionalny, metoda TOPSIS, metoda Warda

1. Wstęp

XXI wiek to okres, w którym w dalszym ciągu można zaobserwować dynamicznie zmieniający się świat na wielu płaszczyznach. Bardzo ważną kwestią jest zatem rozwój i dbanie o pomniejszanie nierówności między krajami. Każdy z nich rozwija się we własnym tempie, które jest uzależnione od zasobów naturalnych czy sytuacji ekonomicznej. Jednym z takich krajów jest Polska, dla której XXI wiek jest okresem przełomowym, gdyż to właśnie wtedy zasiłała szeregi najważniejszej organizacji na starym kontynencie – Unii Europejskiej. Członkostwo w UE otworzyło kraj na nowe możliwości i dało możliwość zbliżenia się poziomem gospodarczym do najlepiej rozwiniętych państw europejskich. Rozwój całego kraju jest jednak efektem rozwijają-

cych się w nich regionów, wśród których także można znaleźć nierówności. Sektorem, którego rozwój powinien iść w parze z rozwojem gospodarczym, jest transport.

Z biegiem czasu dla ludzi zmieniają się priorytety. Coraz bardziej liczy się dla nich wygoda oraz komfort podróżowania. Równie istotne dla nich jest sprawne przewożenie towarów w jak najkrótszym czasie. Na te potrzeby ludzkości musiał odpowiedzieć segment transportu, który zmienił się na przestrzeni lat i z każdą kolejną dekadą otwiera nowe możliwości, pozwalając dotrzeć do najbardziej odległych miejsc. Wszystko wskazuje na to, że trend jego dalszego rozwoju zostanie utrzymany, lecz trudno na ten moment stwierdzić, w którym miejscu możliwości jego rozwoju zostaną wyczerpane.

Infrastruktura transportowa jest bardzo ważnym zagadnieniem, jeżeli chodzi o poziom krajowy. Pojęcie to jest dość szerokie i w jego ramach występuje kilka różnych podziałów, m.in. na infrastrukturę liniową i punktową. Natomiast najważniejszym czynnikiem w prezentowanym artykule będą różnego rodzaju obiekty przyporządkowane do odpowiednich gałęzi transportu. Myśląc o infrastrukturze drogowej, można wyszczególnić sieć dróg, a w szczególności autostrady, mosty, wiadukty czy tunele. W przypadku infrastruktury kolejowej pojawia się sieć dróg szynowych razem z trakcją. Ważną rolę w całym transporcie odgrywają także porty zarówno lotnicze, morskie, jak i przeładunkowe. Nie wszystkie elementy infrastruktury są widoczne dla ludzkiego oka. Jednym z przykładów są korytarze powietrzne dla samolotów. Pojęcie infrastruktury transportowej ma także swoje dalsze części (bezpośrednio z nią powiązana jest infrastruktura transportowo-logistyczna), do których należą, m.in. infrastruktura magazynowa czy urządzenia manipulacyjne (Kukiełka i Woźniak, 2011). Odgrywa ona istotną rolę w funkcjonowaniu państwa i może przyczyniać się do wielu korzystnych zmian. Teza ta została już nieraz potwierdzona przez ekonomistów, którzy mierzyli jej wpływ na rozwój ekonomiczny. Jednym z autorów jest D. Aschauer, który w 1991 roku wykazał, że niedobór inwestycji infrastrukturalnych może być przyczyną gorszej wydajności gospodarczej USA, a także w innych krajach. Raporty Banku Światowego potwierdzają z kolei ich pozytywny wpływ na PKB państw (World Bank, 1994). Regularny rozwój tej części infrastruktury powoduje także zmniejszenie różnic w dochodach mieszkańców (Brandenburg i Sekuła, 2015). Należy się jednak zastanowić, jakie znaczenie ma jej rozwój w kontekście lokalnym. Jest to szansa na rozwój całego regionu, co z kolei ma wpływ gospodarczy na funkcjonowanie kraju, lecz warto także pochylić się nad potencjalnymi zagrożeniami – czy masowy napływ międzynarodowych firm budujących siedziby, np. w pobliżu węzłów autostradowych, nie przyczynia się do upadku wielu lokalnych mikroprzedsiębiorstw? Zagadnienie dotyczące wpływu infrastruktury transportowej na rozwój regionów jest uważane za złożone, gdyż trudno jednoznacznie ocenić jej wpływ. Temat tego powiązania wciąż nie jest wyczerpany i stanowi przedmiot badań ekonomistów.

Współczesne modele polityki rozwoju regionalnego bazują na wzorcach polityki aktywnej i pasywnej. W tym pierwszym pojęciu zawiera się termin interwen-

cjonizmu państwowego, który mówi o tym, że państwo powinno wspierać proces zacierania różnic pomiędzy regionami, gdyż nie znikną one same (Bustowska, 2020). Może to być, np. analiza mocnych stron regionu i odpowiednie wykorzystanie jego atutów lub budowanie spójnej sieci infrastruktury, by nie doszło do wykluczenia transportowego. Druga teoria z kolei zupełnie odrzuca rolę państwa w tym aspekcie. Według neoliberalistów, którzy opowiadają się za tą opcją, jedynie naturalne funkcjonowanie mechanizmów rynkowych prowadzi do równowagi (Bustowska, 2020). W związku z tym jedyną rolę, jaką państwo może przyjąć w tej sytuacji, jest stworzenie odpowiednich warunków do uczciwej konkurencji poszczególnych regionów, co przełoży się na wzrost gospodarczy całego kraju (Pietrzyk, 2001).

Warto także zwrócić uwagę na proces globalizacji, który w czasach współczesnych odgrywa niebagatelną rolę w funkcjonowaniu poszczególnych regionów. W tym kontekście należy rozpatrywać podnoszenie jakości infrastruktury transportowej, która z pewnością jest jedną z pożądanych cech dla państwa wysoko rozwiniętego. Tak jak zostało wcześniej wspomniane, lepsza infrastruktura transportowa jest szasną dla regionu, by na jego terenie zagościły duże, często międzynarodowe firmy posiadające swoje centra logistyczne oraz magazyny. Przy koncentracji tego zjawiska dochodzi do powiększania różnic w poziomie zamożności wśród mieszkańców, ale także następuje ujednoclenie wzorca kulturowego, który staje się uniwersalny. Jest to także zagrożenie dla mniejszych przedsiębiorców, gdyż ich przedsiębiorstwa zostają przejęte przez olbrzymie koncerny lub są zmuszane do zamknięcia działalności w związku z nieopłacalnym jej prowadzeniem.

W związku z wieloma inwestycjami w Polsce poprawiającymi jakość infrastruktury transportowej można dostrzec wzmożone zainteresowanie przedsiębiorstw do budowania swoich obiektów w pobliżu dróg ekspresowych i autostrad lub węzłów komunikacyjnych będących ich skrzyżowaniem. Bardzo dobrze to widać, np. w województwie łódzkim, które ma ogromne predyspozycje do bycia ważnym obszarem na mapie Polski w kontekście gospodarczym. Pierwszym atutem jest oczywiście lokalizacja, gdyż województwo łódzkie jest centralnym obszarem nie tylko Polski, ale także Europy Środkowo-Wschodniej. Odległości do granic są zbliżone, co dla dystrybutorów jest bardzo ważne, gdyż daje możliwości sprawnego przewożenia towarów. Infrastruktura drogowa jest w tym regionie bardzo dobrze rozwinięta, co jest jego drugą zaletą. To właśnie tam krzyżują się dwie najważniejsze autostrady w Polsce – A1, łącząca porty w Trójmieście z południową granicą kraju, oraz A2, będąca połączeniem między Niemcami a Europą Wschodnią. Do tego można wskazać drogę ekspresową S8, która łączy Wrocław i Warszawę, a uwzględniając ją w całej sieci polskich dróg, przechodzi ona wszcz przez cały kraj. Na rozwój tego regionu wpływa także największy na tym obszarze ośrodek miejski – Łódź, jedno z największych miast w Polsce. Działa tam także port lotniczy, który jest ważnym elementem nie tylko w transporcie osób, ale także towarów. Te wszystkie czynniki powodują, że na terenie tego województwa działa Tulipan Park Stryków, będący drugim co do wielkości centrum magazynowym w Polsce. Docelowo ma on posiadać aż 400 tys. m²

powierzchni magazynowej i produkcyjnej, co spowoduje, że będzie on największym takim obiektem w tej części Europy (Kozłowski, 2011). Dobra jakość sieci dróg w tym województwie powoduje, że patrząc na mapę tego regionu, można dostrzec wiele takich obiektów, które stawiają łódzkie, pozornie nieduże województwo, w gronie najważniejszych w kontekście rozwoju regionalnego i gospodarczego obszarów.

Innym regionem jest województwo śląskie, które także nie jest duże, biorąc pod uwagę powierzchnię, ale jest gęsto zaludnione i o wysokim poziomie urbanizacji. Te dwa czynniki „wymuszają” dobrze rozwiniętą sieć transportową na tym obszarze. Dobrze rozwinięta infrastruktura pozwala na jak najlepsze wykorzystanie złoża węgla w tym regionie, który jest następnie transportowany do innych miejsc w kraju, a także eksportowany za granicę. Optymalna sieć drogowa (m.in. skrzyżowanie wspomnianej wcześniej autostrady A1 z autostradą A4 łączącą granice ukraińską i niemiecką) jest również kuszącym miejscem zlokalizowania magazynów i produkcji dla różnych firm i przedsiębiorstw.

W procesie rozwoju regionalnego duże znaczenie odgrywa konkurencyjność, dzięki czemu regiony wzajemnie się napędzają do dalszego poprawiania ich ogólnego stanu. W tym kontekście konkurencyjność oznacza zdolność do przystosowania się do warunków, trendów i wymogów rynku, co skutkuje utrzymaniem swojej pozycji lub jej poprawą w rywalizacji z innymi regionami (Kaczyńska i Korycińska, 2014). Poziom takiej konkurencyjności jest składowym kilku czynników, a jego synonimem jest postęp rozwoju gospodarczego. Na pewno jednym z kluczowych czynników, by region stał się konkurencyjny, jest wspomniana infrastruktura transportowa. Jej zły stan nie zachęca inwestorów do budowania różnych lokacji na takim obszarze. Nawet w miejscach, w których ogólne tempo rozwoju gospodarczego jest wolniejsze, najszybciej rozwijają się tereny mające dobre skomunikowanie z głównymi ośrodkami w kraju oraz z centralnymi regionami UE (Kaczyńska i Korycińska, 2014). Podsumowując te rozważania, można wysunąć wniosek, że region z odpowiednią wielkościową i jakościową siecią transportową jest uważany za atrakcyjny w kontekście inwestycyjnym. Bardzo dobra infrastruktura podnosi tę atrakcyjność, a nawet jest uważana za jeden z kluczowych czynników w tym aspekcie, co zostało potwierdzone badaniami naukowymi i szeregiem ankiet wśród zagranicznych inwestorów (Kaczyńska i Korycińska, 2014).

2. Analiza zależności infrastruktury transportowej i poziomu rozwoju województw w 2021 roku

Celem przeprowadzonej analizy będzie odpowiedź na pytanie: Czy poziom rozwoju poszczególnych regionów ma związek z jakością transportu na tych obszarach? Analiza przyniesie także w konsekwencji odpowiedź, w których regionach sytuacja w tych dwóch aspektach jest najlepsza, a w których najgorsza.

2.1. Opis danych wykorzystanych w badaniu

Dane, na podstawie których przeprowadzone zostały niniejsze analizy, pochodzą z Banku Danych Lokalnych (<https://bdl.stat.gov.pl/bdl/start>) i dotyczą województw w Polsce w 2021 roku. Przed wykonaniem analiz część zmiennych została przekształcona tak, by uwzględniała populację oraz powierzchnię danego województwa. W badaniu zmienne zostały podzielone na dwie grupy: pierwsza z nich ma za zadanie zobrazować jakość rozwoju regionalnego. Zmienne należące do tej grupy przedstawiono oraz szczegółowo opisano poniżej:

- X1 – liczba umów zawartych w ramach Regionalnego Programu Operacyjnego na 10 tys. mieszkańców [szt.]
Zmienna ta określa stan zawartych umów przez dane województwo na dany dzień.
- X2 – wartość umów/decyzji o dofinansowanie ze środków unijnych na 1 mieszkańca [zł]
Łączna wartość pieniężna, którą dane województwo otrzymało w celach realizacji wybranych projektów i inwestycji przeliczona na 1 mieszkańca.
- X3 – nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca [zł]
Wydatki poniesione na innowacje produktów oraz procesów biznesowych poniesione w roku sprawozdawczym z uwzględnieniem populacji danego województwa.
- X4 – nakłady na B+R na 1 mieszkańca [zł]
Wartość środków poniesionych na działalność badania i rozwój (B+R) w roku sprawozdawczym. W łącznej wartości zawierają się zarówno nakłady bieżące, jak i inwestycyjne. Nie obejmują amortyzacji środków trwałych (Główny Urząd Statystyczny [GUS], 2023). Uwzględniona została populacja danego województwa.
- X5 – wartość PKB na 1 mieszkańca [zł]
Wartość Produktu Krajowego Brutto [PKB] w przeliczeniu na 1 mieszkańca danego województwa.
- X6 – stopa bezrobocia rejestrowanego [%]
Jest to stosunek liczby zarejestrowanych osób bezrobotnych do osób aktywnych zawodowo. Nie uwzględnia osób odbywających czynną służbę wojskową (GUS, 2023).
- X7 – odsetek osób z wyższym wykształceniem [%]
Odsetek osób, które ukończyły studia na uczelni wyższej do ogólnej liczby obywateli.
- X8 – średnie wynagrodzenie brutto [zł]
Kwota wyrażona w złotych, którą przeciętnie otrzymuje mieszkaniec danego regionu za miesiąc wykonywanej pracy.
- X9 – średni udział przedsiębiorstw innowacyjnych w ogólnej liczbie przedsiębiorstw [%]

Stosunek przedsiębiorstw o charakterze działalności innowacyjnej do ogólnej liczby przedsiębiorstw.

- X10 – liczba nowo utworzonych miejsc pracy na 1000 mieszkańców [szt.] łączna liczba miejsc pracy otwarta w danym roku kalendarzowym w poszczególnych województwach. Uwzględnia populację wybranego regionu.
- X11 – liczba studentów na 1000 mieszkańców [osoby]
Liczba osób studiujących na polskich uczelniach z polskim i zagranicznym obywatelstwem (GUS, 2023). Uwzględniona została populacja danego regionu.
Wszystkie wyżej wymienione zmienne są stymulantami z wyjątkiem zmiennej X6, która jest destymulantą.

Druża grupa zmiennych opisuje z kolei jakość transportu w poszczególnych województwach. Ich oznaczenia oraz interpretacje przedstawiono poniżej:

- Z1 – liczba kilometrów dróg dla rowerów na 10 tys. mieszkańców [km]
Liczba kilometrów ścieżek rowerowych, która przeznaczona jest do ruchu rowerów jednośladowych. Wymagane jest ich także odpowiednie oznaczenie znakami drogowymi. Długość ścieżek rowerowych leżących po obu stronach jezdni jest liczona oddzielnie (GUS, 2023).
- Z2 – liczba kilometrów dróg ekspresowych i autostrad na 1000 km² [km]
łączna liczba kilometrów dróg szybkiego ruchu w danym województwie, przeliczona na 1000 km², tak by uwzględniała powierzchnię tego regionu.
- Z3 – liczba mostów i wiaduktów na 10 tys. mieszkańców [szt.]
Liczba konstrukcji, które pozwalają na pokonanie przeszkód wodnych bądź lądowych przeliczona na populację poszczególnych regionów.
- Z4 – liczba licencji na taksówki na 10 tys. mieszkańców [szt.]
Liczba uprawnień do przewożenia pasażerów taksówką uwzględniająca populację danego regionu.
- Z5 – liczba autobusów przystosowanych do przewozu niepełnosprawnych na 10 tys. mieszkańców [szt.]
Liczba określająca tabor autobusowy dostosowany do potrzeb osób niepełnosprawnych, m.in. część niskopodłogowa, przeliczona w ten sposób, by uwzględniała populację danego województwa.
- Z6 – liczba przewozów pasażerów w milionach na 100 tys. mieszkańców [osoby]
łączna liczba osób, która skorzystała z transportu publicznego uwzględniająca populację danego regionu.
- Z7 – liczba linii krajowych autobusowych na 10 tys. mieszkańców [szt.]
łączna liczba regularnych połączeń komunikacji autobusowej pomiędzy dwoma lokalizacjami w obrębie granic krajowych uwzględniająca populację danego regionu.
- Z8 – liczba samochodów osobowych na 10 tys. mieszkańców [szt.]
Obejmuje ona zarejestrowane samochody osobowe w poszczególnych województwach, uwzględniając populację danego regionu.
- Z9 – długość linii kolejowych eksploatowanych na 100 km² [km]

Liczba kilometrów linii kolejowych (zarówno zelektryfikowanych, jak i nie), których stan pozwala na ruch kolejowy.

- Z10 – liczba pasażerów korzystająca z transportu lotniczego na 10 tys. mieszkańców [osoby]
Zmienna obrazująca ruch na polskich portach lotniczych – jest to łączna liczba osób, które skorzystały z samolotu, zarówno opuszczając dany port lotniczy, jak i do niego przylatując. Uwzględnia populację danego regionu.
- Z11 – liczba pasażerów korzystających z krajowych połączeń morskich na 10 tys. mieszkańców [osoby]
Łączna liczba osób, która skorzystała z transportu morskiego celem zmiany miejsca położenia w obrębie granic państwowych. Uwzględnia populację danego regionu.
- Z12 – liczba wypadków na 10 tys. mieszkańców [szt.]
Liczba wypadków na drogach publicznych, których skutkiem była śmierć bądź uszkodzenie ciała osób w nich uczestniczących. Zmienna ta mówi o poziomie bezpieczeństwa na drogach. Uwzględnia populację danego regionu.
- Z13 – cena benzyny 95-oktanowej [zł]
Średnia cena, jaką należy zapłacić za kupienie 1 litra benzyny 95-oktanowej.
- Z14 – średnia cena kursu samochodowego kat. B [zł]
Średnia cena, którą należy uiścić za wykupienie dostępu do kursu samochodowego (zajęcia teoretyczne i praktyczne) w celu przystąpienia do państwowego egzaminu na prawo jazdy.
- Z15 – cena biletu normalnego jednorazowego na przejazd autobusem [zł]
Średnia cena biletu uprawniającego jednorazowy przejazd środkiem komunikacji miejskiej w obrębie danego miasta, regionu.
- Z16 – cena za przejazd 5 km taksówką [zł]
Średnia cena za przejazd 5 km taksówką.
Zmienne od Z1 do Z11 są stymulantami, natomiast od Z12 do Z16 destymulantami.

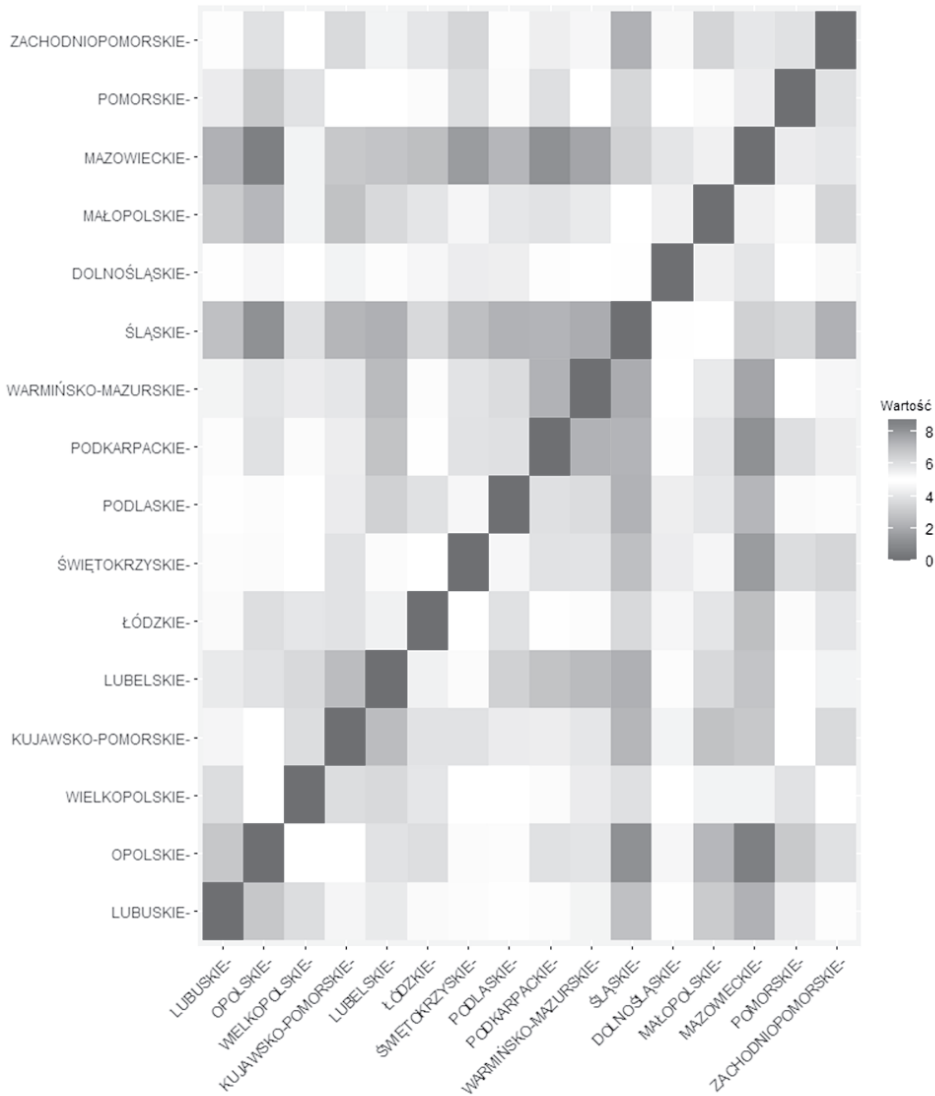
2.2. Klasyfikacja województw z wykorzystaniem hierarchicznej analizy skupień

Zweryfikowanie powiązania poziomu rozwoju regionalnego z jakością transportu wiąże się z przeprowadzeniem szeregu badań. W pierwszym z nich zdecydowano się na dokonanie grupowania województw osobno w każdym z analizowanych aspektów w odniesieniu do roku 2021. W tym celu przeprowadzono analizę, w której skorzystano z hierarchicznej metody aglomeracyjnej bazującej na analizie wariancji, czyli metodzie Warda. Pierwszym krokiem niezbędnym do jej przeprowadzenia było doprowadzenie zmiennych do porównywalności (jednakowa skala) i usunięcie jednostek. Zdecydowano się na przeprowadzenie standaryzacji. Następnie obliczono odległości euklidesowe pomiędzy wszystkimi obiektami (województwami), a wyniki tych obliczeń zaprezentowano w formie graficznej na rysunkach 1 i 2.



Rys. 1. Macierz odległości pomiędzy województwami na podstawie zmiennych związanych z rozwojem w 2021 roku

Źródło: opracowanie na podstawie badań własnych.

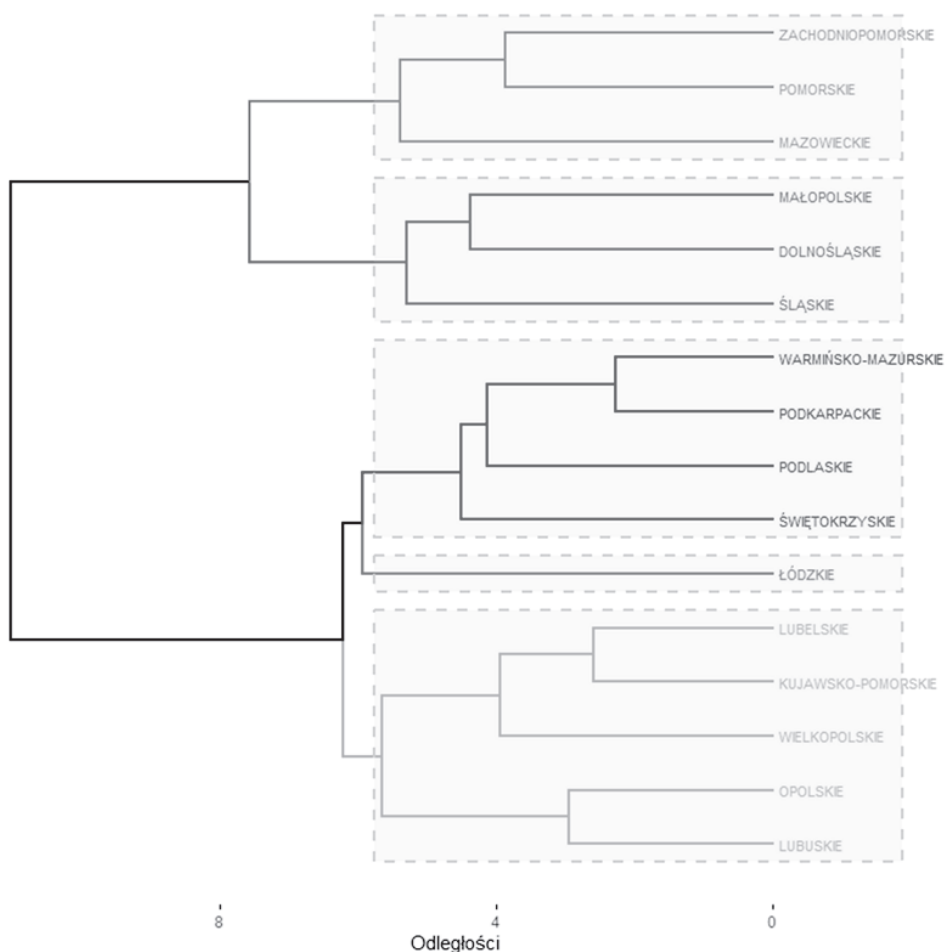


Rys. 2. Macierz odległości pomiędzy województwami na podstawie zmiennych związanych z transportem w 2021 roku

Źródło: opracowanie na podstawie badań własnych.

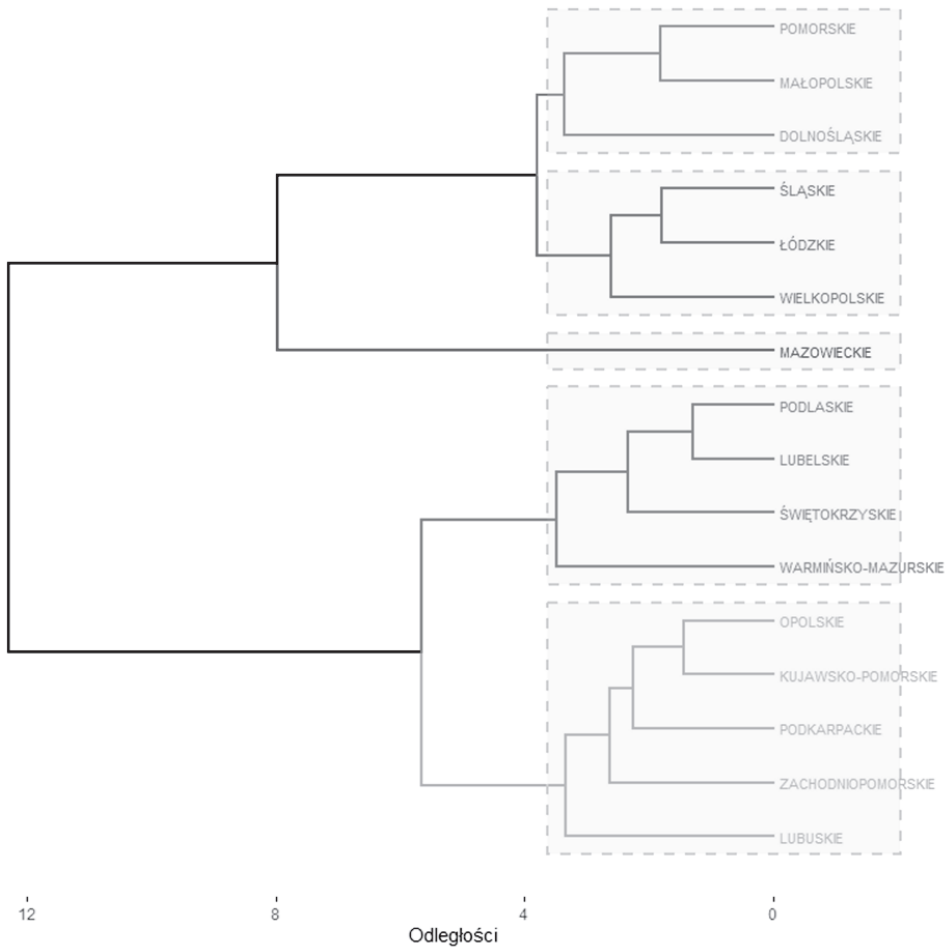
Istotną zależnością, którą można zauważyć na rysunku 1, są duże odległości euklidesowe pomiędzy województwem mazowieckim a pozostałymi regionami, co w konsekwencji może świadczyć o jego odrębności. Wysokie wartości można wskazać także dla województwa warmińsko-mazurskiego z województwami: wielkopolskim, lubuskim, dolnośląskim i małopolskim. W lewym, dolnym rogu przedstawiona

jest grupa województw o niewielkich odległościach, co może świadczyć o przynależności do jednej grupy. Z kolei analiza macierzy z rysunku 2 pozwala stwierdzić, że województwa mazowieckie i śląskie cechują się największymi odległościami od pozostałych regionów, co może sugerować, że pod względem jakości transportu różnią się najbardziej. W celu potwierdzenia hipotezy o odrębności województwa mazowieckiego oraz odpowiedzi na pytanie, które województwa są do siebie podobne w kwestii poziomu rozwoju regionalnego, została przeprowadzona klasyfikacja hierarchiczna za pomocą metody Warda, a jej wyniki zaprezentowano na rysunku 3.



Rys. 3. Dendrogram na podstawie zmiennych związanych z rozwojem w 2021 roku

Źródło: opracowanie na podstawie badań własnych.



Rys. 4. Dendrogram na podstawie zmiennych związanych z transportem w roku 2021

Źródło: opracowanie na podstawie badań własnych.

Przedstawiony na rysunku 3 dendrogram obrazuje podział województw na pięć grup. W pierwszym skupieniu znajdują się województwa pomorskie, małopolskie i dolnośląskie, natomiast w drugim śląskie, łódzkie i wielkopolskie. To co istotne, to fakt, iż w kolejnej fazie łączenia się klastrów łączą się one najszybciej, co oznacza najmniejsze różnice pomiędzy tymi grupami. Województwo mazowieckie zgodnie z przewidywaniami, tworzy grupę jednoelementową. Do kolejnej z kolei trafiły regiony ze wschodniej części Polski: podlaskie, lubelskie, warmińsko-mazurskie oraz świętokrzyskie, a do ostatniego skupienia należą opolskie, podkarpackie, lubuskie, kujawsko-pomorskie oraz zachodniopomorskie. Drugie grupowanie dotyczy jakości transportu w 2021 roku. Do pogrupowania województw ponownie wykorzystana

została klasyfikacja hierarchiczna z wykorzystaniem metody Warda, a liczba skupień ustalona została w taki sposób, by można było porównać grupy pomiędzy sobą.

Z analizy dendrogramu przedstawionego na rysunku 4 można wyciągnąć wniosek, iż pod względem jakości infrastruktury transportowej województwa różnią się istotnie, gdyż odległości pomiędzy nimi są większe. Podobnie jak poprzednio występuje tylko jedno skupienie jednoelementowe, a tworzy je tym razem województwo łódzkie. Z analizy dendrogramu można też wywnioskować, że do jednej grupy trafiły województwa leżące we wschodniej części Polski: warmińsko-mazurskie, podlaskie, podkarpackie oraz świętokrzyskie. Do innej grupy trafiło pomorskie oraz zachodniopomorskie. Ich wspólna obecność w jednym skupieniu nie dziwi, gdyż dzięki dostępowi do morza zyskują więcej możliwości transportowych. Razem z nimi do tej grupy należy województwo mazowieckie, lecz najpóźniej się z nimi połączyło, co jest widoczne na dendrogramie.

Analiza obu dendrogramów nie wskazuje na wyraźne podobieństwo pomiędzy składami poszczególnych grup, lecz nie pozwala na jednoznaczną ocenę zależności pomiędzy dwoma analizowanymi cechami, stąd zostanie przeprowadzone kolejne badanie.

2.3. Porządkowanie liniowe województw z wykorzystaniem metody TOPSIS

W celu stworzenia rankingów województw ze względu na rozwój regionalny, jak i infrastrukturę transportową, wykorzystano jedną z metod porządkowania liniowego – metodę TOPSIS stworzoną przez Ch.-L. Hwanga i K. Yoona (1981). Po określeniu każdej ze zmiennej, czy jest stymulantą, czy destymulantą, wyznaczone zostały współrzędne wzorca oraz antywzorca, czyli najlepszych i najgorszych znormalizowanych wartości rozpatrywanych zmiennych. Następnie obliczone zostały odległości pomiędzy poszczególnymi obiektami a wzorcem i antywzorcem. Na tej podstawie wyznaczone zostały wartości indeksu TOPSIS, które są efektem ilorazu odległości obiektów od antywzorca przez sumę odległości obiektów do wzorca i antywzorca. Wartości te zawierają się w przedziale $<0;1>$, gdzie im wyższa wartość, tym wyższa pozycja w rankingu danego województwa (Bąk, 2016).

Tabela 1. Wartości indeksu TOPSIS na podstawie zmiennych związanych z rozwojem w 2021 roku

| Wyszczególnienie | Województwo | Wartość indeksu TOPSIS |
|------------------|--------------|------------------------|
| 1 | 2 | 3 |
| 1 | Mazowieckie | 0,6832 |
| 2 | Dolnośląskie | 0,4980 |
| 3 | Małopolskie | 0,4843 |
| 4 | Lubelskie | 0,4007 |
| 5 | Podlaskie | 0,4000 |
| 6 | Pomorskie | 0,3945 |
| 7 | Łódzkie | 0,3804 |

Tabela 1., cd.

| 1 | 2 | 3 |
|----|---------------------|--------|
| 8 | Wielkopolskie | 0,3590 |
| 9 | Śląskie | 0,3324 |
| 10 | Warmińsko-mazurskie | 0,3305 |
| 11 | Świętokrzyskie | 0,3248 |
| 12 | Zachodniopomorskie | 0,2460 |
| 13 | Opolskie | 0,2319 |
| 14 | Kujawsko-pomorskie | 0,2294 |
| 15 | Lubuskie | 0,1957 |
| 16 | Podkarpackie | 0,1678 |

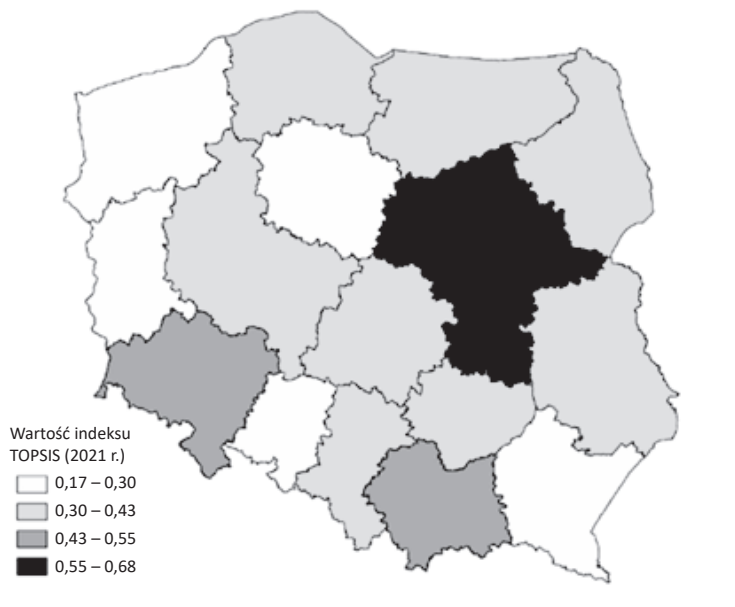
Źródło: opracowanie własne na podstawie (GUS, 2021).

Tabela 2. Wartości indeksu TOPSIS na podstawie zmiennych związanych z transportem w 2021 roku

| Wyszczególnienie | Województwo | Wartość indeksu TOPSIS |
|------------------|---------------------|------------------------|
| 1 | Pomorskie | 0,6355 |
| 2 | Mazowieckie | 0,4500 |
| 3 | Śląskie | 0,3638 |
| 4 | Dolnośląskie | 0,3270 |
| 5 | Zachodniopomorskie | 0,3215 |
| 6 | Małopolskie | 0,3137 |
| 7 | Lubuskie | 0,2894 |
| 8 | Kujawsko-pomorskie | 0,2808 |
| 9 | Opolskie | 0,2744 |
| 10 | Podlaskie | 0,2603 |
| 11 | Łódzkie | 0,2538 |
| 12 | Wielkopolskie | 0,2391 |
| 13 | Lubelskie | 0,2146 |
| 14 | Podkarpackie | 0,2132 |
| 15 | Warmińsko-mazurskie | 0,2103 |
| 16 | Świętokrzyskie | 0,1601 |

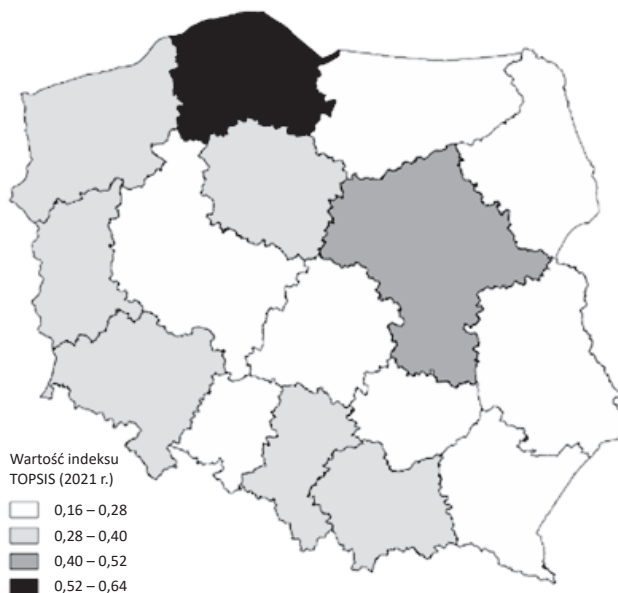
Źródło: opracowanie własne na podstawie (GUS, 2021).

Tabele 1 i 2 ukazują efekty przeprowadzenia porządkowania liniowego dla roku 2021 pod względem zarówno poziomu rozwoju, jak i transportu. Jak wynika z danych w tabeli 1, najlepiej rozwiniętym województwem okazało się, bez zaskoczenia, mazowieckie. Kolejne pozycje zajęły województwa dolnośląskie i małopolskie, które dzieliła bardzo niewielka różnica. Najgorzej rozwiniętym województwem okazało się podkarpackie z wartością indeksu ok. 0,17. Tabela 2 przedstawia z kolei ranking województw dotyczący poziomu infrastruktury transportowej. Liderem zestawienia okazało się województwo pomorskie a drugie miejsce zajęło mazowieckie, co potwierdza wysoki poziom rozwinięcia tego regionu. Ostatnie pozycje zajmują regiony ze wschodniej części kraju – lubelskie, podkarpackie, warmińsko-mazurskie oraz świętokrzyskie.



Rys. 5. Wartość indeksu TOPSIS na podstawie zmiennych związanych z rozwojem w 2021 roku

Źródło: opracowanie własne za pomocą programu QGIS.



Rys. 6. Wartość indeksu TOPSIS na podstawie zmiennych związanych z transportem w 2021 roku

Źródło: opracowanie własne za pomocą programu QGIS.

Województwa zostały podzielone na cztery grupy za pomocą metody równych przedziałów. Polega ona na obliczeniu różnicy pomiędzy największą a najmniejszą wartością, a następnie podzieleniu jej przez ustaloną liczbę grup. W ten sposób utworzona zostaje ustalona liczba przedziałów o takiej samej rozpiętości. Różnice pomiędzy przynależnością województw do poszczególnych grup są znaczące i trudno mówić o podobieństwie jakości infrastruktury transportowej i poziomu rozwoju regionalnego.

Tabela 3. Wyniki rangowania województw w 2021 roku z wykorzystaniem metody TOPSIS

| Województwo | Ranga (rozwój) | Ranga (transport) |
|---------------------|----------------|-------------------|
| Dolnośląskie | 2 | 4 |
| Kujawsko-pomorskie | 14 | 8 |
| Lubelskie | 4 | 13 |
| Lubuskie | 15 | 7 |
| Łódzkie | 7 | 11 |
| Małopolskie | 3 | 6 |
| Mazowieckie | 1 | 2 |
| Opolskie | 13 | 9 |
| Podkarpackie | 16 | 14 |
| Podlaskie | 5 | 10 |
| Pomorskie | 6 | 1 |
| Śląskie | 9 | 3 |
| Świętokrzyskie | 11 | 16 |
| Warmińsko-mazurskie | 10 | 15 |
| Wielkopolskie | 8 | 12 |
| Zachodniopomorskie | 12 | 5 |

Źródło: opracowanie własne.

Tabela 3 przedstawia rangi, jakie województwa osiągnęły w procesie porządkowania liniowego dla obu zestawów zmiennych w 2021 roku. W celu dokładnej oceny wpływu jednego sektora na drugi ponownie został obliczony współczynnik korelacji rang Spearmana. Niska wartość współczynnika wynosząca 0,37 nie potwierdza jednoznacznie zależności pomiędzy poziomem transportu a jakością rozwoju regionalnego.

Tabela 4. Wartości współczynnika korelacji Pearsona pomiędzy wartościami indeksu TOPSIS na podstawie zmiennych związanych z rozwojem a zmiennymi związanymi z transportem w 2021 roku

| | | | | | | | | |
|-----------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Zmienna | Z ₁ | Z ₂ | Z ₃ | Z ₄ | Z ₅ | Z ₆ | Z ₇ | Z ₈ |
| Korelacja | 0,28 | 0,03 | 0,09 | 0,54 | 0,67 | 0,68 | 0,67 | 0,15 |
| Zmienna | Z ₉ | Z ₁₀ | Z ₁₁ | Z ₁₂ | Z ₁₃ | Z ₁₄ | Z ₁₅ | Z ₁₆ |
| Korelacja | 0,08 | 0,78 | 0,07 | 0,12 | 0,03 | 0,47 | 0,55 | 0,24 |

Źródło: opracowanie własne.

W tabeli 4 przedstawiono wartości współczynnika korelacji Pearsona pomiędzy wartościami indeksu TOPSIS na podstawie wartości znormalizowanych zmiennych związanych z rozwojem a poszczególnymi wartościami znormalizowanych zmiennych związanych z transportem. Jeden z wniosków, jaki można wyciągnąć to taki, że w województwach o wyższym poziomie rozwoju więcej pasażerów korzysta z połączeń lotniczych. Współczynnik korelacji między wartością wskaźnika TOPSIS a tą zmienną wyniósł 0,78.

3. Zakończenie

Celem pracy było otrzymanie odpowiedzi na pytanie, czy – a jeśli tak, to w jakim stopniu – rozwój regionu powiązany jest z infrastrukturą transportową. W dążeniu do jego osiągnięcia autor postanowił zbadać związek tych dwóch obszarów na przykładzie województw w roku 2021. Wykorzystując metody statystyczne, takie jak analiza skupień metodą Warda oraz porządkowanie liniowe metodą TOPSIS, podjęte zostały próby pogrupowania województw, w których sytuacja pod względem rozwoju, a także transportu jest podobna, a także uzyskanie rezultatów, jeśli chodzi o ich pozycję na tle innych regionów.

Przeprowadzone analizy wykazały, że w 2021 roku w Polsce zjawiska rozwoju regionalnego oraz infrastruktury transportowej nie były od siebie silnie uzależnione. Porównując rankingi uzyskane na podstawie dwóch zestawów zmiennych (jeden dotyczący rozwoju, a drugi transportu), trudno znaleźć podobieństwa, a kolejność województw w rankingu różni się od siebie. Najlepiej rozwiniętym województwem w 2021 roku zostało województwo mazowieckie. Wynik ten nie dziwi, ponieważ na jego terenie znajduje się największe polskie miasto będące zarazem stolicą – Warszawa. Kolejne najwyżej sklasyfikowane województwa to dolnośląskie oraz małopolskie. Na ich terenie także działają duże ośrodki miejskie, odpowiednio Wrocław i Kraków, które pozytywnie działają na tempo wzrostu poziomu rozwoju tych regionów.

Najlepiej rozwiniętym województwem, jeśli chodzi o infrastrukturę transportową, okazało się pomorskie. Stanowi ono bardzo istotną rolę w handlu międzynarodowym, a dostęp do morza znacząco zwiększa jego atrakcyjność. Na jego terenie działa wiele portów, wśród nich te największe zlokalizowane w Trójmieście. Jest także bardzo dobrze połączone z pozostałymi częściami kraju, dzięki niezłe rozbudowanej sieci drogowej. Poza ważną rolę tego województwa w handlu jest to także często wybierany kierunek przez turystów, szczególnie w miesiącach letnich. Województwo pomorskie musi być przygotowane na tak duże zainteresowanie regionem, co zapewnia rozwój każdej z gałęzi transportu. Wśród dobrze rozwiniętych województw w tym kontekście znalazło się także województwo mazowieckie, na którego terenie leży największe lotnisko w Polsce. Mazowieckie ma szansę na dalszy rozwój, a jednym z potencjalnych elementów może być skuteczna budowa Centralnego Portu Komunikacyjnego.

Literatura

- Bąk, A. (2016). *Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza*. Wydawnictwo Uniwersytetu Ekonomicznego im. Oskara Langego we Wrocławiu.
- Brandenburg, H. i Sekuła, P. (2015). *Wpływ infrastruktury transportu drogowego na rozwój przedsiębiorczości*. Wydawnictwo Uniwersytetu w Białymstoku
- Busłowska, A. (2020). *Wpływ infrastruktury kolejowej na rozwój regionalny w ujęciu teorii myślenia sieciowego*. Wydawnictwo Uniwersytetu w Białymstoku.
- Główny Urząd Statystyczny [GUS]. (2023). Bank Danych Lokalnych. Pobrano 10 marca 2024 z [https:// bdl.stat.gov.pl/bdl/start](https://bdl.stat.gov.pl/bdl/start)
- Hwang, Ch.-L., i Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag.
- Kaczyńska, W. i Korycińska, K. (2014). *Wpływ infrastruktury transportu drogowego na rozwój regionu*. Uniwersytet Przyrodniczo-Humanistyczny w Siedlcach.
- Kozłowski, R. (2011). Wpływ infrastruktury transportu na rozwój klastrow na przykładzie regionu łódzkiego. *Acta Universitatis Lodzianensis. Folia Oeconomica*, (251).
- Kukiełka, L. i Woźniak, D. (2011). *Niektóre aspekty logistyki transportu*. Instytut Naukowo-Wydawniczy „SPATIUM” Sp. z o.o.
- Pietrzyk, I. (2001). *Polityka regionalna w Unii Europejskiej i regiony w państwach członkowskich*. Wydawnictwo Naukowe PWN.
- World Bank. (1994). *World Development Report 1994: Infrastructure for Development*. Oxford University Press.

The Impact of Transport Infrastructure on Regional Development

Abstract: Regional development is an element that affects the level of economic development of the country. The quality of selected elements of the public sphere often depends on how developed the country is. In this article, the author will consider whether the level of transport infrastructure is also an element influencing the level of development of the region, and if so, to what extent. The article will be divided into 2 chapters. First, the relationship between transport infrastructure and regional development will be described, along with an indication of potential opportunities and threats. The second chapter will be of a research nature, which will include a comparative analysis of voivodeships in 2021 between the quality of transport infrastructure and the level of development, performed using k-means and TOPSIS methods. In the final part, a summary will be presented, which will include conclusions from the conducted research.

Keywords: transport infrastructure, regional development, TOPSIS, Ward

Marta Sośnicka

e-mail: 182359@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Wpływ wybranych czynników społeczno-ekonomicznych na rynek książki w Polsce

DOI: 10.15611/2024.53.6.11

JEL Classification: C38, C83, D91, Z11

© 2024 Marta Sośnicka

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Sośnicka, M. (2024). Wpływ wybranych czynników społeczno-ekonomicznych na rynek książki w Polsce. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 163-178). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Celem artykułu jest zbadanie jakie czynniki społeczno-ekonomiczne wpływają na rynek książki w Polsce, na którym od kilku lat obserwowany jest kryzys. Badanie oparto na analizie dostępnej literatury, a także przeprowadzono ankietę internetową dotyczącą zachowań konsumenckich wśród osób, które lubią czytać. Zaprezentowano odpowiedzi respondentów, a następnie na ich podstawie przeprowadzono analizę z użyciem modelu regresji liniowej oraz modelu logitowego, zależności nieliniowych, a także wielowymiarowej analizy korespondencji wraz z analizą skupień metodą Warda. Badanie ankietowe wykazało, że przyczyną malejącego popytu na książki nie jest jedynie ich rosnąca cena, ale głównie tryb życia ludzi, którzy przez pracę i obowiązki domowe nie mają już czasu na czytanie.

Słowa kluczowe: rynek książki, czytelnictwo, książka, analiza zależności, analiza korespondencji, metoda Warda

1. Wstęp

Trudno byłoby wyobrazić sobie życie bez czytania, jest to podstawowa umiejętność, bez której człowiek nie byłby zdolny do normalnego funkcjonowania. Jednak dla niektórych czytanie może być czymś więcej – drogą do rozwoju myśli i umiejętności. Czytelnictwo według M. Skwarnickiego (1960, s. 252) „(...) jest procesem społecznym polegającym na zaspokojeniu potrzeb kulturalnych, naukowych i rozrywkowych ludzi poprzez przyswajanie sobie przekazywanego w piśmie dorobku myśli i ducha ludzkiego”.

Książka jest powszechnie znanym symbolem mądrości, jednak jej znaczenie słabnie na tle nowoczesnych technologii. Dzisiaj można znaleźć ją nie tylko w jej tra-

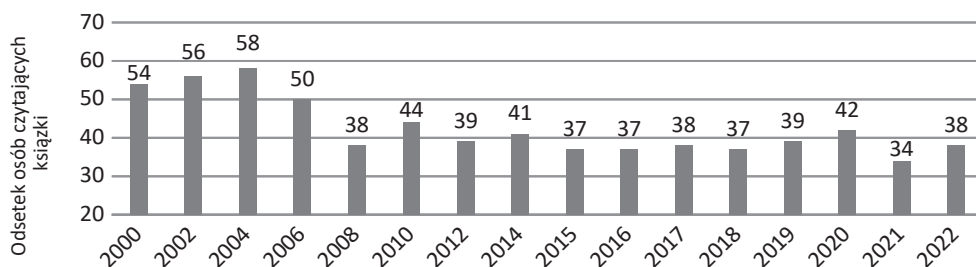
dycyjnej papierowej formie, ale również w wersji elektronicznej (e-book) i mówionej (audiobook). W ostatnich latach uległa zmianie nie tylko forma książki, ale także sposób życia ludzi. Rozwinął się rynek Video on Demand, czyli wideo na życzenie (w skrócie: VOD). Oglądanie filmów i seriali stało się jedną z form spędzenia wolnego czasu, którego przeważnie brakuje w dniu pełnym pracy i obowiązków domowych, co przekłada się na coraz rzadsze sięganie po lekturę.

Istnieje wiele czynników, które przyczyniają się do zamykania bibliotek publicznych i wydawnictw. Kryzys na rynku wydawniczym, zwanym także rynkiem książki lub księgarskim, trwa już od kilku lat. Można by sądzić, że książka, która jest obecna od wielu wieków, zajmuje bezpieczną pozycję i nie istnieje żadne ryzyko jej zniknięcia w przyszłości.

Niniejszy artykuł skupia się na zbadaniu, jakie czynniki wpływają na rynek wydawniczy. Przedstawiono obecną sytuację na rynku książki w Polsce oraz wyniki ankiety internetowej przeprowadzonej na przełomie roku 2022 i 2023 wśród osób, które deklarują, że lubią czytać. Celem badania ankietowego była analiza zachowań konsumenckich, dostępności finansowej i materialnej dotyczącej czytelnictwa oraz korzystania z usług VOD, a także zbadanie, czy istnieje zależność pomiędzy wydatkami na wskazane formy aktywności.

2. Rynek książki

Liczba osób czytających w Polsce z każdym rokiem maleje. Biblioteka Narodowa przeprowadza cyklicznie sondaż na ogólnopolskiej, reprezentatywnej próbie składającej się z Polaków powyżej 15. roku życia. Odsetek ankietowanych, którzy odpowiedzieli, że w ciągu ostatniego roku przeczytali przynajmniej fragment jednej książki papierowej lub elektronicznej wynosi 38%. Ze względu na reprezentatywność próby badawczej można stwierdzić, że stan czytelnictwa w Polsce, czyli odsetek osób czytających, wynosił 38% i utrzymuje się na tym poziomie od 2008 roku, w którym to odnotowano znaczny spadek w stosunku do poprzedniego badania (Zasacka i Chymkowski, 2023). Ze względu na szeroką definicję czytelnictwa faktyczny jego poziom może być przeszacowany i osób, które regularnie czytają, jest prawdopodobnie mniej. W 2020 roku, w którym rozpoczęła się pandemia koronawirusa, odnotowano najwyższy od 10 lat odsetek osób czytających, kształtujący się na poziomie 42%, natomiast w 2021 wskaźnik ten był najniższy od początku XXI w. i wynosił zaledwie 34%. Poziom czytelnictwa w Polsce na przestrzeni lat zaprezentowano na rysunku 1. Chwilowy wzrost w 2020 roku był spowodowany całkowitą zmianą sposobu funkcjonowania społeczeństwa. Towarzyszący pandemii kryzys gospodarczy, rosnąca liczba zwolnień pracowniczych oraz obostrzenia sprawiły, że ludzie cały czas przebywali w swoich domach zmuszeni do rezygnacji ze spotkań i niektórych aktywności. Znaleźli oni wtedy czas i motywację do sięgnięcia po książkę.



Rys. 1. Poziom czytelnictwa w Polsce w latach 2000-2022

Źródło: opracowanie na podstawie badań Biblioteki Narodowej.

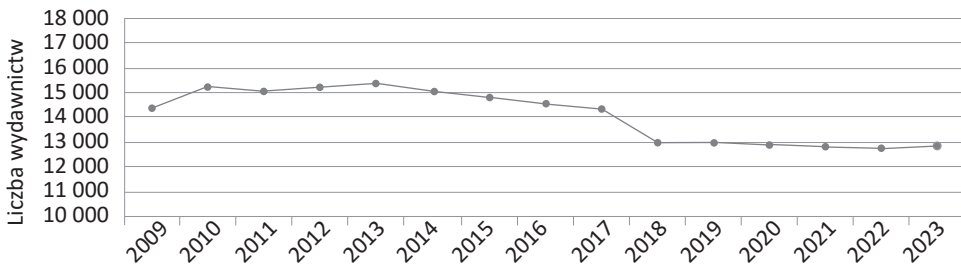
Badanie Biblioteki Narodowej w 2021 i 2022 roku wykazało, że istotnym czynnikiem wpływającym na czytelnictwo jest, m.in. sytuacja finansowa. W grupie osób, które nie mają problemów finansowych i są w stanie oszczędzać na przyszłość, 64% z nich przeczytało co najmniej jedną książkę. W grupie, której starcza pieniędzy jedynie na bieżące wydatki – 48%, wśród osób żyjących oszczędnie – 36%, bardzo oszczędnie – 28%, natomiast w grupie ludzi deklarujących, że pieniędzy starcza za ledwie na podstawowe potrzeby – tylko 17% przeczytało co najmniej jedną pozycję (Zasacka i Chymkowski, 2022).

Kolejnym ważnym aspektem czytelnictwa jest źródło, z którego pochodzą książki. Jedynie 13% ankietowanych w badaniu Biblioteki Narodowej wskazało, że korzysta z możliwości bezpłatnego wypożyczenia książek z biblioteki publicznej. Najczęściej czytelnicy kupują egzemplarze – 39% respondentów (Zasacka i Chymkowski, 2022).

Oprócz Biblioteki Narodowej analizą czytelnictwa zajmuje się również Polska Izba Książki współpracująca z portalem internetowym Lubimyczytac.pl. Instytucje te od 2016 roku przeprowadzają badanie ankietowe „Czytelnicy w sieci”, w którym każdy może wziąć udział. W 2022 roku respondenci najczęściej korzystali z księgarń internetowych (59% ankietowanych), sieciowych księgarń stacjonarnych (49%) oraz bibliotek (43%). Jedynie 10% odpowiadających przyznało, że głównym źródłem są niesieciowe księgarnie stacjonarne, 39% respondentów nie odwiedza ich, a 20% uważa ich istnienie za nieważne (Instytutksiązki.pl, 2022). Wyniki badania dowodzą, w jak ciężkiej sytuacji znalazły się lokalne księgarnie. Przyczyną tego stanu rzeczy jest dominowanie na rynku sieci sklepów, m.in. Empiku, oraz malejący popyt na książki.

Nie tylko księgarnie niesieciowe znajdują się w niekorzystnym położeniu, kryzys obejmuje także wydawnictwa. Według danych Głównego Urzędu Statystycznego [GUS] w 2017 roku funkcjonowało 14 375 przedsiębiorstw wydawniczych, natomiast w 2018 już tylko 13 004, co stanowi największy spadek w XXI wieku na poziomie 9,54%. Wysoka liczba wydawnictw na początku tego okresu była spowodowana zniesieniem cenzury w 1990 roku (Dobrołęcki i Waszczyk, 2015). Po

jej likwidacji ludzie poczuli wolność i potrzebę czytania, co wiązało się z zakupem nowych egzemplarzy. Przedsiębiorcy widząc rosnący popyt na to dobro, zakładali działalność wydawniczą i publikowali książki w ogromnych nakładach. W niektórych przypadkach produkcja przewyższała nawet trzykrotnie zapotrzebowanie na dany tytuł (Klukowski i Tobera, 2013). Na rysunku 2 można zauważyć tendencję malejącą, która utrzymywała się od 2013 do 2022 roku. W 2023 roku liczba wydawnictw wynosiła 12 850 i jest to o 85 więcej niż w poprzednim roku, co stanowi wzrost na poziomie 0,67%.



Rys. 2. Liczba przedsiębiorstw zajmująca się działalnością wydawniczą

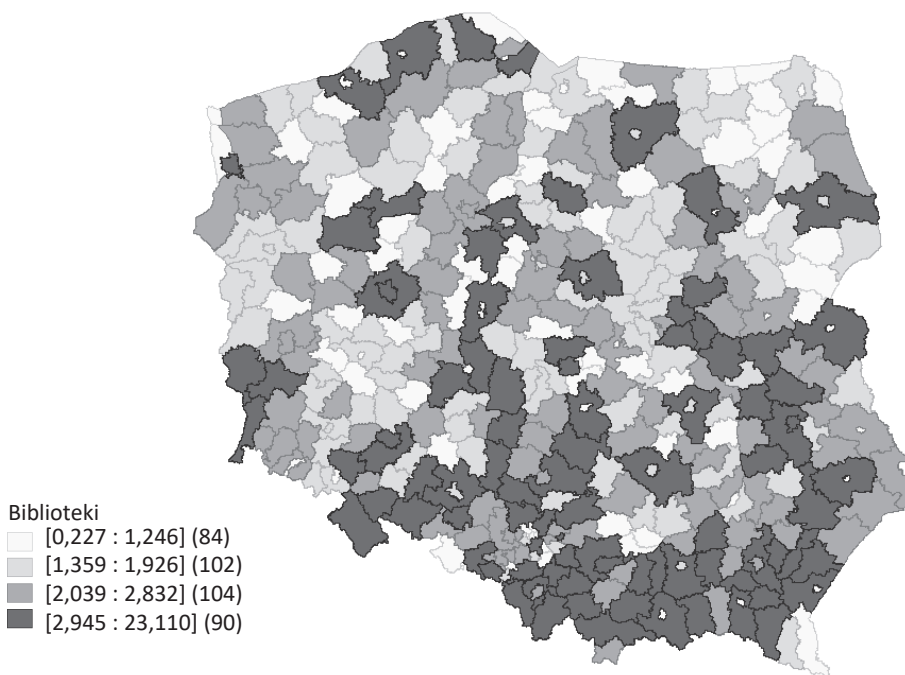
Źródło: opracowanie na podstawie danych GUS.

Malejąca liczba wydawnictw wiąże się ze spadającym poziomem czytelnictwa i z rosnącymi kosztami wydania pojedynczego egzemplarza. Ceny niektórych materiałów produkcyjnych w 2021 roku drastycznie wzrosły, w tym tektury introliigatorskiej o 100%, a papieru o prawie 40%. Ponadto wydawnictwa mają problem z dostępnością materiałów drukarskich oraz zatrudnieniem pracowników. Wszystkie te czynniki przekładają się na wzrost ceny książki dla docelowego klienta (Rynek-ksiazki.pl, 2022), co według badania przeprowadzonego przez A. Bernala (2014) skutecznie zniechęca do zakupu.

Bezpłatnym źródłem dostępu do książki jest biblioteka publiczna, jednak nie zawsze czytelnik znajdzie interesujący go tytuł, a ponadto biblioteki w Polsce nie są równomiernie rozmieszczone, zatem może to również ograniczać możliwości wypożyczenia. Na rysunku 3 zaprezentowano kartogram utworzony za pomocą metody kwantylowej dla czterech klas. Przedstawia on liczbę bibliotek i filii w przeliczeniu na 10 000 mieszkańców danego powiatu w 2021 roku na podstawie danych GUS.

Na rysunku 3 widać dysproporcję w rozmieszczeniu bibliotek w miastach na prawach powiatu oraz sąsiadujących z nimi powiatami niemiejskimi. Wynikać to może z istotnej różnicy gęstości zaludnienia między tymi rodzajami powiatów, ponadto biblioteki zlokalizowane w wielkomiejskich ośrodkach są przeważnie zaopatrzone w większe księgozbiory, a więc są dostępne dla większej liczby czytelników. Na podstawie utworzonego kartogramu można również dostrzec różnicę między częścią południową, w której powiaty przeważnie należą do najbardziej dostępnej klasy,

a częścią północną, w której liczba bibliotek jest niewystarczająca w stosunku do liczby mieszkańców. Ze względu na ograniczoną możliwość wypożyczania książek czytelnicy mieszkający na obszarze północnym mogą czytać mniej.



Rys. 3. Liczba bibliotek i filii na 10 000 mieszkańców powiatu w 2021 roku

Źródło: opracowanie na podstawie danych GUS.

Rynek książki, jak i jego uczestnicy zmagają się z wieloma trudnościami. Do najważniejszych należą rosnące koszty produkcji i deficyt materiałów drukarskich, niski dostęp do wypożyczalni w północnej części Polski, malejące zainteresowanie książką, brak czasu, a także rozwój konkurencyjnych rynków.

3. Wykorzystane metody badawcze

W celu zbadania zachowań konsumenckich czytelników posłużono się metodą ankietową. Postanowiono także zbadać, czy istnieje zależność między wydatkami na czytanie a oglądaniem filmów lub seriali jako substytucyjnymi formami spędzania czasu. Kwestionariusz został utworzony za pomocą Google Forms, a następnie udostępniony w mediach społecznościowych. Okres zbierania odpowiedzi trwał od listopada 2022 do marca 2023 roku. Badaną grupę stanowili czytelnicy w znaczeniu osób, które zadeklarowały, że lubią czytać książki. Respondenci wypełniali

kwestionariusz indywidualnie i bez pomocy ankietera. Próba nie jest reprezentatywna, gdyż czytelnicy nie mają równego dostępu do Internetu, a tym bardziej do mediów społecznościowych. Problematyczne jest również określenie rzeczywistej wielkości populacji osób lubiących czytać w Polsce. Uzyskano 933 anonimowych odpowiedzi, które zostały poddane analizie przy pomocy programów Microsoft Excel i Statistica.

Analiza odpowiedzi ankietowanych została podzielona na dwie części. W pierwszej z nich zbadano podstawowe charakterystyki respondentów, natomiast w drugiej części skupiono się na zbadaniu czynników wpływających na zakup książek. W tym celu użyto modelu regresji liniowej, modelu logitowego, analizy zależności nieliniowej oraz analizy korespondencji.

Celem wielowymiarowej liniowej funkcji regresji było określenie wpływu wybranych zmiennych niezależnych na zmienną zależną, większość z nich stanowiły zmienne binarne. Parametry oszacowano za pomocą klasycznej metody najmniejszych kwadratów, w której dąży się do zminimalizowania odchyień wartości empirycznych od teoretycznych (Kukuła i in., 2009). Oceniono dopasowanie funkcji za pomocą współczynnika determinacji oraz zbadano istotność modelu za pomocą testu F, wykorzystując iloraz średniego kwadratowego odchylenia regresyjnego i losowego (Aczel i Sounderpandian, 2018). Po oszacowaniu współczynników regresji sprawdzono istotność wpływu poszczególnych zmiennych za pomocą testu *t*-Studenta opartego na odchyleniach standardowych estymatorów parametrów strukturalnych (Piłatowska, 2008). Zmienne, które nie były istotne, nie zostały wyeliminowane z modelu, ponieważ celem budowania modelu nie była predykcja, ale zbadanie wpływu wybranych cech na zmienną zależną.

Większość zmiennych pochodzących z ankiety miała charakter niemierzalny, wykorzystano więc model regresji logistycznej do określenia prawdopodobieństwa zajścia danego zjawiska oraz czynników wpływających na nie. Parametry wyznaczono w oparciu o metodę największej wiarygodności, w której dąży się do zmaksymalizowania prawdopodobieństwa otrzymania wyników zgodnych z zaobserwowanymi. Otrzymane współczynniki informują o tym, czy wzrost wartości danej cechy wpływa na zwiększenie, czy zmniejszenie szansy wystąpienia badanego zjawiska. Model został zweryfikowany poprzez wskaźniki oparte na funkcji wiarygodności oraz statystykę χ^2 . Sprawdzono również istotność parametrów za pomocą testu *t*-Studenta i testu Warda. Dodatkowo przeprowadzono analizę trafności klasyfikacji, oparając się na wskaźniku globalnym i indywidualnym. Ponadto sprawdzono za pomocą miernika Huberty'ego, czy predykcja na podstawie modelu jest bardziej korzystna w stosunku do losowego przypisania (Grześkowiak, 2006).

Sprawdzono również, czy istnieją istotne zależności pomiędzy wydatkami na przypuszczalnie substytucyjne aktywności. W tym celu posłużono się testem niezależności χ^2 . Utworzono tablicę z liczebnościami empirycznymi i na jej podstawie oszacowano liczebności teoretyczne oraz obliczono statystykę χ^2 . Przyjęcie hipotezy zerowej świadczy o niezależności zmiennych (Rusnak, 2003). Kolejnym sposobem

sprawdzenia wzajemnego wpływu obu zmiennych było obliczenie współczynnika tau Kendalla. Jest to miara zależności oparta na różnicy prawdopodobieństwa wystąpienia jednoczesnych zmian wartości zmiennych w tym samym lub w przeciwnym kierunku, a im większa jego wartość bezwzględna, tym silniejsza jest korelacja między nimi (Malevergne i Sornette, 2006).

Aby znaleźć współwystępowanie kategorii, posłużono się wielowymiarową analizą korespondencji, w której dąży się do zredukowania liczby wymiarów. Współrzędne obliczane są na podstawie złożonej macierzy znaczników, analizę prowadzi się, opierając się na różnicach między położeniem punktów wskazujących na poszczególne cechy. Ich bliskie sąsiedztwo świadczy o jednoczesnym występowaniu cech. Liczbę wymiarów można wybrać na podstawie skumulowanego procentu bezwładności, dla których udział inercji jest wysoki, a zwiększenie wymiaru nie powoduje dużego wzrostu. Alternatywnym sposobem identyfikacji odpowiedniego wymiaru jest znalezienie momentu na wykresie wartości własnych, po którym nie występują gwałtowne spadki (Stanimir, 2012). Aby zwizualizować wyniki analizy korespondencji, skorzystano z analizy skupień, stosując metodę Warda, która łączy w grupy obiekty o podobnych wartościach zmiennych wynikające z odległości euklidesowych pomiędzy współrzędnymi. Po każdym połączeniu ponownie obliczane są odległości z uwzględnieniem nowo powstałych grup. Procedura powtarzana jest do momentu połączenia wszystkich obiektów w jedno skupienie (Balicki, 2013).

4. Wyniki badania ankietowego

4.1. Statystyki opisowe poszczególnych pytań kwestionariusza

Ankiętę internetową prowadzoną na przełomie 2022 i 2023 roku wypełniły 933 osoby. Składała się z 34 pytań. Wypełnić ją mogła każda osoba lubiąca książki i posiadająca dostęp do mediów społecznościowych, na których ankieta była udostępniona.

Grupę respondentów tworzyło 844 kobiet i 49 mężczyzn. 25% najmłodszych respondentów stanowiły osoby do 23. roku życia, natomiast 25% najstarszych przekroczyło 38 lat. Mediana wieku kształtowała się na poziomie 29,5 lat. Największą część respondentów stanowiły osoby pracujące (50,70%), studenci (19,40%), osoby pracujące i jednocześnie uczące się (15,86%), a także osoby niepracujące (8,47%) i uczniowie (5,57%). Najwięcej osób ukończyło studia magisterskie (38,16%) i posiadało średnie wykształcenie (35,16%). Najczęściej respondenci mieszkali na wsi (26,69%) lub w mieście powyżej 500 000 mieszkańców (24,97%). Ankietowani uważali swoją sytuację finansową przeważnie jako dobrą (43,84%) lub przeciętną (39,76%), natomiast w kwestii sytuacji rodzinnej były to głównie osoby będące w związku formalnym lub nieformalnym i nieposiadające dzieci (32,80%).

Osoby czytające lubią również oglądać filmy i seriale. Stanowią oni 90,53% ankietowanych, a wśród nich wyróżnia się grupa oglądająca tylko wybrane pozycje (44,31%). Najmniejszy odsetek stanowili ludzie, którzy starają się unikać tej czynno-

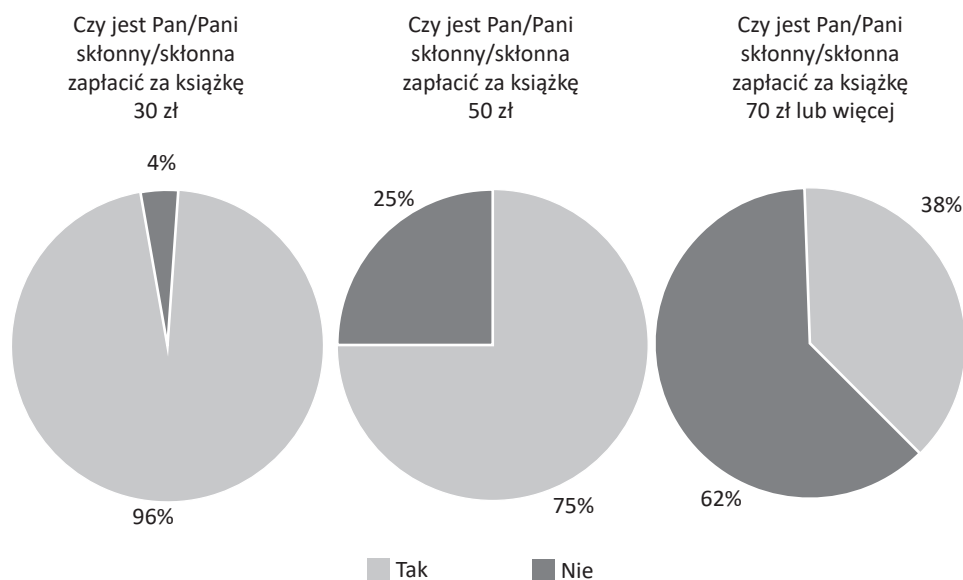
ści (1,82%). Osoby oglądające filmy i seriale przynajmniej raz w tygodniu stanowią 78,67% respondentów, a rzadziej niż raz w miesiącu – 5,79%. 70,28% ankietowanych odpowiedziało, że poświęca dokładnie tyle czasu, ile chce na oglądanie filmów i seriali, 23,61% chciałoby robić to częściej, a jedynie 6,33% wołałoby ograniczyć czas spędzany na tym. Jako najczęstszy powód ograniczania oglądania podawano brak czasu spowodowany obowiązkami domowymi (15,97%), pracą (12,33%), spędzaniem czasu z rodziną i przyjaciółmi (10,29%) oraz nauką (9,22%). Ograniczenia techniczne były powodem tylko dla 0,96% respondentów, natomiast finansowe dla 0,86%. W dzisiejszych czasach oglądanie stało się mobilne. Można to robić nie tylko w domu, korzystając z telewizora, ale wystarczy posiadać telefon z dostępem do Internetu i serwisów VOD. Wśród ankietowanych 89,48% korzysta właśnie z tego typu usług, z czego najczęściej respondenci korzystają z Netflixa (79,21%), HBO (37,83%) i Disney+ (34,73%). 64,95% respondentów płaci za dostęp do bazy filmów i seriali, najczęściej ok. 15-30 zł miesięcznie (17,70%), natomiast 25,50% ogląda je bezpłatnie.

Badanie ankietowe dotyczyło wyłącznie osób, które lubią książki. Dla 60,02% z nich czytanie jest jedną z pasji. Dla 24,65% respondentów jest to duża przyjemność, natomiast 15,33% lubi czytać tylko wybrane pozycje. Codziennie czyta 39,12% osób i jest to najczęściej wybierana odpowiedź. Przynajmniej średnio raz w tygodniu czyta 81,57% ankietowanych, a rzadziej niż raz w miesiącu – 5,68%. Zaledwie 0,64% osób przyznało, że chciałoby ograniczyć czas spędzony na tym, 26,37% uważa, że czyta wystarczająco, a 49,20% chciałoby przeznaczyć więcej czasu na książki. 62,40% respondentów wskazało jako przyczynę ograniczenia swoich chęci częstszego czytania obowiązki domowe, 54,70% – pracę, natomiast 40,52% – spędzanie czasu z rodziną i przyjaciółmi. Często wskazywano również na problemy z koncentracją, brak motywacji i zmęczenie. Do najrzadszych powodów należały problemy zdrowotne oraz brak dostępności materialnej lub ograniczenia finansowe, zatem głównym czynnikiem niesięgania po książkę jest brak czasu.

W ciągu roku ankietowani czytają średnio 31,36 książek, przy czym 50% z nich czyta więcej niż 21,17 tytułów. 25% ankietowanych nie czyta więcej niż 8,65 książek rocznie, a 25% najbardziej oddanych czytelników czyta powyżej 47,9 pozycji. Najczęściej respondenci odpowiadali, że czytają więcej niż 50 książek rocznie (19,82%), od 6 do 10 (12,97%) i od 3 do 5 (10,83%). Według badań Biblioteki Narodowej pandemia COVID-19 spowodowała tymczasowy wzrost stanu czytelnictwa. Dla większości respondentów (61,47%) jednak nie wywarła ona żadnego wpływu, natomiast dla 38,26% ten okres skłonił do sięgnięcia po książkę, z kolei 15,97% ankietowanych wciąż czytało więcej niż przed lockdownem.

Najistotniejsze w badaniu są preferencje zakupowe czytelników. 45,98% uczestników ankiety zdecydowanie woli kupować nowe książki ze sklepu, 27,22% wypożyczać z biblioteki, natomiast 10,50% woli kupić używane książki. Najczęściej ludzie przeznaczają w ciągu miesiąca średnio do 50 zł na zakup książek, jednak najwięcej – 29,26% badanych – nabywa książki okazjonalnie i ciężko jest im określić miesięczne

wydatki, a 12,43% nie kupuje ich wcale. Anketowani przeciętnie kupują prawie 20 książek w ciągu roku. Największa grupa osób odpowiedziała, że kupuje ich od 11 do 20 (23,3%). Mediana kształtuje się na poziomie 13,59 zakupionych książek na rok.



Rys. 4. Skłonność do zakupu książki w zależności od jej ceny

Źródło: opracowanie własne.

Cena pojedynczego nowego egzemplarza wynosi ok. 50 zł, a jej dalszy wzrost jest nieunikniony (Awolusi, 2023). Inflacja jest jedną z przyczyn ograniczania wydatków na dobra, które nie są niezbędne. 44,26% respondentów prawdopodobnie ograniczy wydatki związane z subskrypcją serwisów VOD oraz z zakupem książek. Wydawnictwa, księgarnie i czytelnicy od kilku lat usiłują wynegocjować z rządem przyznanie zerowego VAT-u, jednak nieskutecznie i ciągle obowiązuje 5% (Awolusi, 2023). Anketowanym zadano zatem pytanie o skłonność do zapłacenia za interesującą ich książkę daną cenę. Wyniki przedstawiono na rysunku 4. Gdyby egzemplarz książki kosztował poniżej średniej – 30 zł, 96,36% anketowanych zdecydowałoby się na jej zakup, gdyby 50 zł – już tylko 74,81% osób zgodziłoby się na to, natomiast 61,95% badanych zrezygnowałoby z zakupu, gdyby interesująca ich książka kosztowała 70 zł lub więcej. Wraz ze wzrostem ceny maleje chęć do zakupu nowej pozycji, zatem jeśli książka dalej będzie drożeć pogorszy to sytuację na rynku wydawniczym.

4.2. Badanie wpływu wybranych czynników na zachowania konsumentkie

Rynek książki znajduje się obecnie w kryzysowej sytuacji. Z jednej strony koszt wytworzenia pojedynczego egzemplarza wzrasta oraz potrzebne materiały są mniej

dostępne, co przekłada się na końcową cenę dla czytelnika. Z drugiej strony popyt spada ze względu na tryb życia społeczeństwa i spadający poziom czytelnictwa.

Postanowiono sprawdzić, jakie czynniki wpływają na liczbę kupowanych książek drukowanych. W tym celu zbudowano model liniowy, opierając się na odpowiedziach uzyskanych w ankiecie. Do analizy przyjęto zmienne binarne: trwająca nauka i praca, posiadanie wykształcenia średniego lub wyższego, mieszkanie w mieście, co najmniej dobra sytuacja finansowa, wychowywanie dzieci, bycie w związku formalnym lub nieformalnym, częste i bardzo częste oglądanie filmów lub seriali, korzystanie z serwisów VOD, opłacanie subskrypcji serwisów VOD oraz korzystanie z biblioteki. Zmiennymi ilościowymi w modelu były wiek i liczba czytanych książek w ciągu roku. Pozostawiono zmienne, które nie są istotne statystycznie na żadnym sensownym poziomie, jednak są ważne z punktu analizowanego zjawiska, ich współczynniki zawarto w tabeli 1.

Tabela 1. Współczynniki i wartości p-value dla zmiennych niezależnych w modelu liniowym objaśniającym liczbę kupowanych książek

| Nazwa zmiennej | Współczynnik | P-value | Nazwa zmiennej | Współczynnik | P-value |
|------------------------------|--------------|---------|---------------------------------------|--------------|---------|
| Wyraz wolny | 22,6953 | 0,0000 | Dzieci | 1,1712 | 0,5223 |
| Wiek | 0,0009 | 0,9924 | Związek | 1,7616 | 0,1963 |
| Trwająca nauka | 1,6821 | 0,3860 | Częste oglądanie filmów i seriali | 1,7638 | 0,1656 |
| Praca | 0,5723 | 0,7179 | Korzystanie z serwisów VOD | 4,4615 | 0,0457 |
| Wykształcenie <i>średnie</i> | 2,5323 | 0,3297 | Opłacanie usług VOD | 0,1217 | 0,0000 |
| Wykształcenie wyższe | 4,8857 | 0,0665 | Korzystanie z biblioteki | 6,3504 | 0,0000 |
| Mieszkanie w mieście | 2,2654 | 0,1055 | Liczba czytanych książek w ciągu roku | 0,4253 | 0,0000 |
| Dobra sytuacja finansowa | 0,7721 | 0,5359 | | | |

Źródło: opracowanie własne.

Największy korzystny wpływ na liczbę kupowanych książek ma liczba przeczytanych egzemplarzy. Przeczytanie o 10 książek więcej w ciągu roku powoduje średni wzrost zakupionych pozycji o ok. 4,25, zakładając niezmiennosc pozostałych czynników. Na wzrost liczby zakupionych książek wpływa również to, czy dana osoba opłaca subskrypcję serwisów VOD. Osoby takie są skłonne zakupić przeciętnie o 0,12 książek więcej, *ceteris paribus*. Jednakże osoby, które korzystają z takich usług, kupują średnio o 4,46 mniej egzemplarzy rocznie, a tym bardziej osoby, które deklarują częste lub bardzo częste oglądanie filmów i seriali. Przy założeniu niezmienności pozostałych zmiennych jest to o 1,76 książek mniej.

Najbardziej negatywnie wygląda korzystanie z biblioteki. Ankietowani mając możliwość wypożyczenia książki, kupują średnio o 6,35 egzemplarza mniej. Również wysokie oddziaływanie niesie posiadanie wyższego (-4,89) lub średniego (-2,53) wykształcenia i mieszkanie w mieście (-2,27). Wysoce prawdopodobne jest, że osoby mieszkające na wsi kupią więcej książek ze względu na utrudniony dostęp do biblioteki lub ograniczone zasoby księgozbiorów oraz prowadzą zupełnie inny tryb życia niż osoby mieszkające w większych miastach.

Chcąc potwierdzić, które ze zmiennych wpływają korzystnie na liczbę kupowanych książek, utworzono model logitowy, którego celem było określenie, jakie czynniki przyczyniają się do wyboru głównego źródła dostępu do książki. Zmienną zależną był zakup nowej książki w sklepie lub księgarni. Obliczono parametry, a ich istotność potwierdzono testem *t*-Studenta i Warda. Wyniki zaprezentowano w tabeli 2.

Tabela 2. Wyniki estymacji modelu logitowego objaśniającego zakup nowych książek

| Nazwa zmiennej | Ocena parametru β | Wartość p-value dla statystyki <i>t</i> -Studenta | Wartość p-value dla statystyki Walda |
|---|-------------------------|---|--------------------------------------|
| Wyraz wolny | 0,2270 | 0,0158 | 0,0156 |
| Dzieci | 0,2627 | 0,0017 | 0,0000 |
| Czytanie powyżej 20 książek rocznie | 0,5222 | 0,0000 | 0,0000 |
| Korzystanie z biblioteki | 0,7512 | 0,0000 | 0,0016 |
| Miesięczne wydatki na książki powyżej 50 zł | 0,9570 | 0,0000 | 0,0081 |
| Kupowanie e-booków | 0,2143 | 0,0082 | 0,0463 |
| Miesięczne wydatki na VOD powyżej 30 zł | 0,1660 | 0,0466 | 0,0000 |

Źródło: opracowanie własne.

Statystyka χ^2 jest równa 271,9568 o 6 stopniach swobody, wartość p-value jest w przybliżeniu równa 0, zatem oceny parametrów są łącznie istotne. Logarytm wiarygodności ma wartość 1015,42. Obliczono *likelihood ratio index* – wynosi on 0,2112, oraz korektę tego wskaźnika – R2CS jest równa 0,2528. Unormowany wskaźnik R2CU wynosi 0,3379, a więc model objaśnia w 33,79% przypisanie do grupy osób najczęściej pozyskujących swoje książki ze sklepów bądź nie. Jest to objaśnianie w umiarkowanym stopniu, jednak wystarczające do zastosowania tego modelu w niniejszym badaniu.

Iloraz szans obliczany jest przez podniesienie podstawy logarytmu naturalnego do potęgi równej ocenie parametru β , tzn. e^β . Prawdopodobieństwo zakupu nowej książki jest wyższe dla osób, których średnie miesięczne wydatki na książki wynoszą powyżej 50 zł miesięcznie oraz ponad 30 zł za usługi VOD. Czytelnicy, mający dzieci, korzystający z biblioteki, kupujący e-booki lub czytający powyżej 20 książek rocznie są bardziej skłonni do korzystania z tańszych źródeł książki.

Dla osób bezdzietnych prawdopodobieństwo częstszego zakupu nowych egzemplarzy jest prawie 1,7 razy większe niż dla rodziców przy założeniu, że wartości po-

zostałych czynników pozostaną niezmienione. Wśród osób wydających na książki ponad 50 zł miesięcznie szansa na to, że przeważnie kupują je w sklepie, jest prawie 6,8 razy większa, *ceteris paribus*. Dla czytelników czytających mniej niż 20 książek w ciągu roku prawdopodobieństwo, że nabywają głównie nowe egzemplarze, jest ponad 2,8 razy większe niż dla osób czytających więcej, zakładając stałość pozostałych zmiennych. Czytając wiele pozycji, zakup nowych może być nieopłacalny, więc osoby takie korzystają z biblioteki lub odkupują używane egzemplarze. Czytelnicy, którzy nie korzystają z wypożyczalni publicznych, są 4,5 razy bardziej skłonni kupować książki ze sklepów lub księgarni.

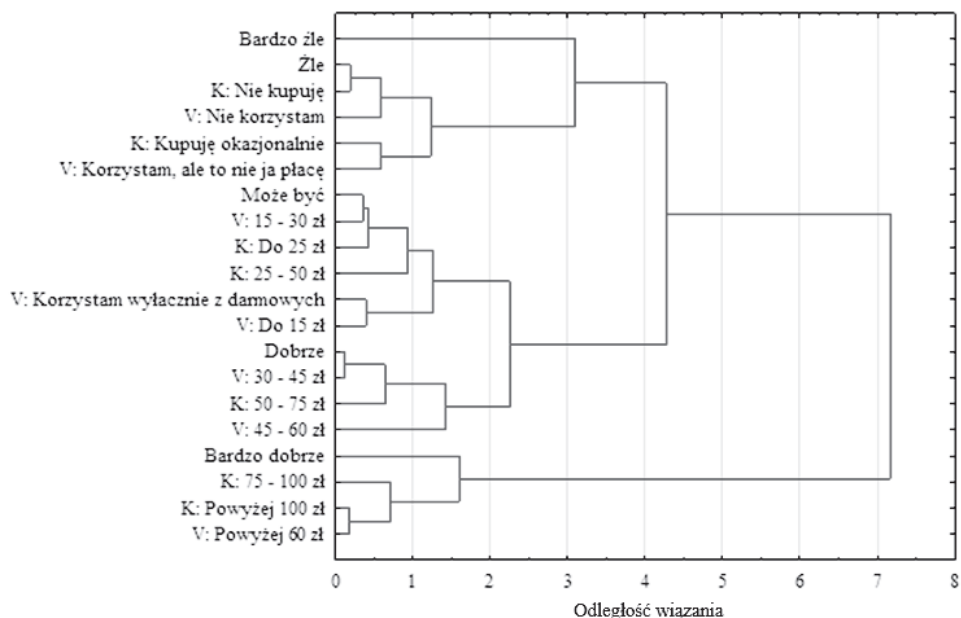
Postanowiono zbadać trafność zaklasyfikowania ankietowanych. Według ich odpowiedzi 45% dokonuje zakupu głównie w sklepach, zatem ustanowiono wartość progową równą 0,5 ze względu na podobny rozkład prawdopodobieństwa kupienia nowych książek i korzystania z alternatywnych źródeł. Poprawnie przypisano 296 osób, które kupują, czyli 69% dobrze zaklasyfikowanych, oraz 387 czytelników (76,79%), którzy nie kupują nowych książek, zatem całkowita trafność poprawnego przypisania przypadków stanowi 73,2%. Za pomocą miernika Huberty'ego zbadano, czy oszacowany model jest lepszy niż losowe przypisanie do grupy. Miernik wynosi 0,4641. Jest on większy od zera, więc użycie tego modelu logitowego niesie ze sobą większe korzyści niż losowa przynależność. Redukcja błędów jest równa 46,41%.

Średnia cena jednostkowa nowego egzemplarza wynosi ok. 50 zł, natomiast oferty za używaną sztukę wahają się między 5 a 15 zł. W przypadku pojedynczej miesięcznej subskrypcji VOD jest to koszt w wysokości 15-30 zł, aczkolwiek w niektórych przypadkach istnieje możliwość współdzielenia kont z innymi użytkownikami, przez co ostateczna kwota przypadająca na jedną osobę może być niższa (Krajewski, 2023). Najczęściej ankietowani wskazywali, że kupują książki okazjonalnie oraz korzystają z serwisów VOD, ale ktoś inny opłaca subskrypcję. Biorąc pod uwagę jedynie konkretne kwoty wskazane przez respondentów, dominują wydatki w okolicy średniej każdego z produktów, tzn. książki w przedziale 25-50 zł za egzemplarz oraz serwisy VOD 15-30 zł miesięcznie. Rozkład odpowiedzi przedstawiono w tabeli 3. Postanowiono sprawdzić, czy istnieje zależność między tymi wydatkami. Przeprowadzono test χ^2 dla zestawu odpowiedzi ankietowanych. Statystyka χ^2 wynosi 97,28 dla 42 stopni swobody, natomiast p-value jest bliskie 0, zatem można odrzucić hipotezę o niezależności wydatków. Biorąc pod uwagę jedynie grupę osób o liczebności 382, które wskazały konkretne wartości swoich wydatków, miesięczne koszty zakupu książek i subskrypcji są zależne. χ^2 wynosi 34,05 o 16 stopniach swobody, a p-value jest równe 0,0054. Współczynnik korelacji tau Kendalla wynosi 0,18, a p-value jest mniejsze od 0,05, co świadczy o słabym, chociaż istotnym wzajemnym wpływie wydatków. Im większe wydatki na książki, tym również większe na subskrypcję serwisów VOD.

Tabela 3. Rozkład odpowiedzi respondentów dotyczący miesięcznych opłat serwisów VOD i wydatków na książki

| Zakup książek | Subskrypcje serwisów VOD | | | | | | | |
|---------------------|--------------------------|-----------------------------|-----------------------------|----------|----------|----------|----------|-------------|
| | Nie korzystam | Korzystam tylko z darmowych | Korzystam i płaci ktoś inny | Do 15 zł | 15-30 zł | 30-45 zł | 45-60 zł | Ponad 60 zł |
| Nie kupuję | 15 | 9 | 25 | 14 | 20 | 15 | 11 | 7 |
| Kupuję okazjonalnie | 24 | 28 | 68 | 38 | 49 | 32 | 16 | 18 |
| Do 25 zł | 15 | 9 | 15 | 18 | 20 | 17 | 17 | 8 |
| 25-50 zł | 14 | 16 | 19 | 20 | 36 | 21 | 29 | 19 |
| 50-75 zł | 5 | 6 | 17 | 7 | 19 | 14 | 17 | 15 |
| 75-100 zł | 11 | 5 | 11 | 1 | 17 | 12 | 9 | 19 |
| Więcej niż 100 zł | 9 | 2 | 8 | 6 | 4 | 9 | 14 | 14 |

Źródło: opracowanie własne.

**Rys. 5.** Grupowanie zmiennych użytych w analizie korespondencji

Źródło: opracowanie własne.

Mogłoby się wydawać, że osoba, która uważa swoją sytuację finansową za dobrą lub bardzo dobrą, będzie wydawać więcej pieniędzy na rzeczy, które sprawiają jej przyjemność. Przeprowadzono wielowymiarową analizę korespondencji w celu

potwierdzenia, czy to przypuszczenie jest prawdziwe. Inercja całkowita wynosi 5,6667. Bazując na wykresie osypiska, wybrano trójwymiarową przestrzeń rzutowania, dla której skumulowany procent bezwładności wynosi 14,78%. Jest to słabe odwzorowanie, jednak ma ono charakter wyłącznie poglądowy, aby sprawdzić, czy osoby o lepszej sytuacji finansowej przeznaczają więcej pieniędzy na książki lub serwisy VOD. Biorąc pod uwagę otrzymane współrzędne z trójwymiarowego rozkładu rozrzutu, obliczono odległości między nimi i utworzono dendrogram, korzystając z metody Warda. Zaprezentowano go na rysunku 5.

Wyniki analizy potwierdzają założenie, że średnie wydatki na czytanie i oglądanie zwiększają się wraz z polepszeniem sytuacji finansowej. Osoby, które opisują swoją sytuację jako złą lub bardzo złą, najczęściej wstrzymują się od kupna książek oraz subskrypcji. Osoby mówiące co najmniej dobrze o swoich finansach mogą uważać takie koszty za niewielkie i warte wydania stosunkowo niskiej dla nich kwoty za wypełnienie wolnego czasu, które sprawi im przyjemność.

5. Zakończenie

Rynek książki nie jest stabilny. Wpływa na niego wiele czynników o charakterze globalnym, o czym świadczy, np. tymczasowy wzrost stanu czytelnictwa w trakcie pandemii koronawirusa, krajowym – starania o ustanowienie zerowego VAT-u na książki, oraz inflacja, a także powody indywidualne. Przez codzienne obowiązki, pracę i naukę coraz bardziej problematyczne staje się znalezienie czasu na czytanie. Nawet jeżeli się to uda, często zmęczenie zniechęca do sięgnięcia po lekturę, zatem przeważnie wybierane są czynności niewymagające większego wysiłku, jak oglądanie telewizji.

Sytuacja finansowa nie jest obecnie przeszkodą w czytaniu, co wykazało badanie ankietowe, gdyż osoby zainteresowane daną książką mają możliwość skorzystania z innych źródeł. Jednak za wysoka cena egzemplarzy jest jedną z przyczyn kryzysowej sytuacji wydawnictw oraz księgarni, gdyż spada liczba bezpośrednio zakupionych książek. Do pozostałych powodów należy dostępność i cena materiałów oraz zagrożenie monopolem przez funkcjonowanie wielkich przedsiębiorstw, które są konkurencją dla mniejszych księgarni niesieciowych.

Wyniki badania ankietowego wykazały, że im lepsza sytuacja finansowa, tym wyższe wydatki na książki. Jednakże takie osoby przeważnie kupują mniej egzemplarzy. Ponadto im większe opłaty za usługi VOD, tym większe miesięczne wydatki na książki wśród osób lubiących czytać. Istotnym wynikiem sondażu jest również to, że dla osób czytających więcej niż 20 książek rocznie prawdopodobieństwo zakupu nowej książki z księgarni lub sklepu jest niższe niż dla osoby, która czyta mniej.

Najważniejszą przeszkodą, która doprowadza do upadku rynku książki, jest brak czasu obecnych konsumentów i potencjalnych czytelników. Tryb życia, jaki prowadzą ludzie w dzisiejszych czasach, skutecznie odciąga od lektury, przez co maleją szanse na ponowny rozkwit rynku i przywrócenie znaczenia książki.

Literatura

- Aczel, A. i Sounderpandian, J. (2019). *Statystyka w zarządzaniu*. Wydawnictwo Naukowe PWN.
- Awolusi, D. (2023). *Nie będzie tańszych książek w 2023 r. Rząd nie zgodził się na książki bez VAT*. Granice.pl. Pobrano 4 kwietnia 2024 z <https://www.granice.pl/news/zerowy-vat-na-ksiazk-czy-beda-ksiazki-bez-vat/19906>
- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Bernal, A. (2014). Przerzucalność podatku od towarów i usług na rynku książek. *Ruch prawniczy, Ekonomiczny I Socjologiczny*, 76(4), 297-315. <https://doi.org/10.14746/rpeis.2014.76.4.20>
- Dobrołęcki, P. i Waszczyk, P. (2015). *Jubilat stawia na rozwój*. Rynek-książki.pl. Pobrano 2 kwietnia 2024 z <https://rynek-ksiazki.pl/czasopisma/jubilat-stawia-na-rozwoj/>
- Grześkowiak, A. (2006). Regresja logistyczna, s. 208-231, 236-247. W: A. Stanimir (red.), *Analiza danych marketingowych Problemy, metody, przykłady*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego
- Instytutksiążki.pl. (2022, 27 maja). *Opublikowano wyniki badania „Czytelnicy w sieci”*. Pobrano 1 kwietnia 2024 z <https://instytutksiazki.pl/aktualnosci,2,opublikowano-wyniki-badania--%E2%80%9E-czytelnicy-w-sieci%E2%80%9D,7648.html>
- Klukowski, B. i Tobera, M. (2013). *W tym niezwykłym czasie. Początki transformacji polskiego rynku książki (1989-1995)*. Wydawnictwo Akademickie Sedno.
- Krajewski, R. (2023). *Ile zapłacimy za platformy streamingowe w 2023 roku? Wszystkie subskrypcje to koszt kilku tysięcy złotych*. Gram.pl. Pobrano 5 kwietnia 2024 z <https://www.gram.pl/news/ile-zaplacimy-za-platformy-streamingowe-w-2023-roku-wszystkie-subskrypcje-to-koszt-kilku-tysiecy-zlotych>
- Kukuła, K., Goryl, A., Jędrzejczyk, Z., Osiewalski, J. i Walkosz, A. (2009). *Wprowadzenie do ekonometrii*. Wydawnictwo Naukowe PWN.
- Malevergne, Y., i Sornette, D. (2006). *Extreme Financial Risks: From Dependence to Risk Management*. Springer-Verlag.
- Piątowska, M. (2008). *Repetitorium ze statystyki*. Wydawnictwo Naukowe PWN.
- Rusnak, Z. (2003). Testowanie hipotez nieparametrycznych, s. 291-308. W: S. Ostasiewicz, Z. Rusnak, i U. Siedlecka, *Statystyka. Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Rynek-książki.pl. (2022, 19 stycznia). *Drastycznie wzrosły koszty produkcji książek*. Pobrano 2 kwietnia 2024 z <https://rynek-ksiazki.pl/czasopisma/drastycznie-wzrosly-koszty-produkcji-ksiazek/>
- Skwarnicki, M. (1960). Czytelnictwo i metody jego badania. W: S. Tazbir (red.), *Problemy czytelnictwa w wielkich miastach. Krajowa Konferencja Bibliotekarska w Warszawie (2-4 czerwca 1958). Referaty – przebieg obrad*, 241-253. Stowarzyszenie Bibliotekarzy Polskich.
- Stanimir, A. (2012). Różne techniki prezentacji powiązań kategorii zmiennych niemetrycznych *Ekonometria = Econometrics*, 3(37), 9-25.
- Stat.gov.pl. (2024, 25 marca). *Sytuacja społeczno-gospodarcza kraju w lutym 2024 r*. Pobrano z <https://stat.gov.pl/obszary-tematyczne/inne-opracowania/informacje-o-sytuacji-spolesczno-gospodarczej/sytuacja-spolesczno-gospodarcza-kraju-w-lutym-2024-r-,1,142.html>
- Zasacka, Z. i Chymkowski, R. (2023). *Stan czytelnictwa książek w Polsce w 2022 roku*. Biblioteka Narodowa.
- Zasacka, Z. i Chymkowski, R. (2022). *Stan czytelnictwa książek w Polsce w drugim roku pandemii (2021-2022)*. Biblioteka Narodowa.

The Impact of Selected Socio-Economic Factors on the Book Market in Poland

Abstract: The main aim of this article is to investigate the influence of socio-economics factors on the book market in Poland, that has been experiencing a crisis for several years. The research was based on the analysis of available literature and also a web-based survey was conducted. The purpose of the survey was to explore readers' consumer behaviour. The survey results were presented and analysed using regression and logit model, analysis of non-linear dependencies, discrimination and correspondence analysis with Ward's clustering. A declining books' demand is caused not only by the cost of production but mostly by people's lifestyle, because of work and housework there is no more time for reading.

Keywords: book market, reading, book, dependency analysis, correspondence analysis, Ward's method

Julia Wobalis

e-mail: 182566@student.ue.wroc.pl

ORCID: 0009-0003-2373-3733

Uniwersytet Ekonomiczny we Wrocławiu

Analiza jakości systemów ochrony zdrowia w państwach Unii Europejskiej

DOI: 10.15611/2024.53.6.12

JEL Classification: I10, C38

© 2024 Julia Wobalis

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

Cytuj jako: Wobalis, J. (2024) Analiza jakości systemów ochrony zdrowia w państwach Unii Europejskiej. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 179-194). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Streszczenie: Dostęp do wysokiej jakości opieki zdrowotnej jest prawem każdego obywatela Unii Europejskiej. Istnieją jednak różnice między krajami w zakresie jakości systemów ochrony zdrowia a ich finansowaniem. Działania Unii Europejskiej umożliwiają koordynację współpracy państw, monitorowanie efektów i przyspieszenie poprawy jakości opieki zdrowotnej. W artykule porównano systemy opieki zdrowotnej w krajach członkowskich UE z wykorzystaniem metod porządkowania liniowego. Za pomocą analizy skupień pogrupowano je w klasy na podstawie podobieństw w wybranych zmiennych dotyczących jakości systemu ochrony zdrowia. W celu lepszego zrozumienia tematu, w pracy zawarto również opis wskaźników wykorzystywanych do oceny jakości usług ochrony zdrowia i wnioski pochodzące z badań wykorzystujących te wskaźniki.

Słowa kluczowe: porządkowanie liniowe, analiza skupień, ochrona zdrowia

1. Wstęp

Systemy opieki zdrowotnej odgrywają istotną rolę w zapobieganiu chorobom, leczeniu, utrzymaniu zdrowia populacji i poprawie jakości życia w społeczeństwach. Dostęp do niech jest uważany za podstawowe prawo człowieka zgodnie z Kartą Praw Podstawowych Unii Europejskiej, znacząco wpływając na rozwój społeczno-gospodarczy. Istnieją jednak różnice między państwami członkowskimi Unii Europejskiej w zakresie alokacji środków na opiekę zdrowotną i organizacji systemów, co prowadzi do rozbieżności w jakości usług, a w konsekwencji wpływa na dobrobyt obywateli.

Niniejszy artykuł ma na celu ocenę i porównanie jakości systemów opieki zdrowotnej w państwach członkowskich UE przy użyciu technik wielowymiarowej analizy porównawczej. Podjęto w nim próbę zidentyfikowania krajów wykazujących zarówno wyższą, jak i niższą jakość opieki zdrowotnej, z wykorzystaniem porządkowania liniowego, a także zrobiono grupowanie krajów o podobnych wartościach zmiennych za pomocą analizy skupień. Dostrzegając podobieństwa i rozbieżności między systemami opieki zdrowotnej, możliwe jest wskazanie obszarów wymagających większej uwagi i poprawy.

2. Wskaźniki jakości opieki zdrowotnej w literaturze

Analiza jakości opieki zdrowotnej w krajach UE ma istotne znaczenie we współczesnej Europie, kształtowanej przez różne czynniki. Wraz ze starzeniem się społeczeństw i presją na doskonalenie systemów zabezpieczenia społecznego znaczenie usług zdrowotnych staje się coraz bardziej wyraźne. Przewidywany wzrost w tym sektorze, napędzany zmianami demograficznymi i postępem technologicznym, dodatkowo podkreśla potrzebę kontroli i poprawy jakości usług. Komisja Europejska dostrzega ogromny wpływ opieki zdrowotnej na codzienne życie obywateli oraz konieczność zajęcia się niedociągnięciami jakościowymi w tym obszarze dla zapewnienia dobrobytu społeczeństw w całej Europie (Dubois i in., 2019).

Istnieją różne metody wyodrębniania wskaźników jakości opieki zdrowotnej. Jedną z opcji jest podział na wskaźniki bazujące na proporcjach lub na liczbach. Wskaźniki oparte na proporcjach są bardziej powszechne i mają jasno określony licznik i mianownik (np. odsetek pacjentów, którzy otrzymują antybiotykoterapię przed operacją). Natomiast wskaźniki oparte na liczbach koncentrują się na poszczególnych niepożądanych zdarzeniach, z których każde powinno być przeanalizowane w celu poprawy jakości. Inną klasyfikacją jest podział na wskaźniki ogólne, czyli obejmujące wszystkich pacjentów, i wskaźniki specyficzne dla określonych diagnoz. Oprócz tego wskaźniki mogą być klasyfikowane na podstawie wymiarów jakości oraz funkcji opieki zdrowotnej, które oceniają (Busse i in., 2019). Najczęściej stosowanym sposobem klasyfikacji wskaźników jest triada Donabediana wskaźników struktury, procesu i wyniku, która bada wpływ struktur opieki zdrowotnej na zdrowie pacjentów. Wskaźniki struktury oceniają warunki opieki świadczonej przez dostawców, wskaźniki procesu oceniają działania podczas świadczenia usług, a wskaźniki wyników dostarczają informacji na temat wyników zdrowotnych pacjentów (McCullough i in., 2022).

Można wyróżnić kilka zestawów wskaźników, które wykorzystywane są do badania i porównywania jakości ochrony zdrowia w Unii Europejskiej. Jednym z nich jest European Core Health Indicators [ECHI] – lista 88 wskaźników podzielonych na 5 kategorii. Kategoria „stan zdrowia” opisuje śmiertelność, oczekiwaną długość życia oraz zapadalność na choroby i urazy. „Uwarunkowania zdrowotne” to grupa

związana z nawykami wpływającymi na zdrowie, takimi jak dieta i uzależnienia. Pozostałe badane kategorie to „usługi zdrowotne”, „promocja zdrowia” oraz „sytuacja demograficzna i społeczno-ekonomiczna” (Ec.europa.eu, b.d.). Na wskaźnikach ECHI bazuje coroczny raport OECD „Health at a Glance: Europe”, który opisuje stan opieki zdrowotnej w krajach Europy.

Kolejną miarą jakości ochrony zdrowia w 35 europejskich krajach jest Euro Health Consumer Index [EHCI]. Jest to grupa 46 wskaźników podzielonych na 6 kategorii: „prawa pacjenta i informacje”, „dostęp do opieki”, „wyniki leczenia”, „zakres i zasięg usług”, „profilaktyka” i „stosowanie farmaceutyków”. W celu uwydatnienia znaczenia niektórych obszarów ochrony zdrowia kategoriom nadaje się wagi, a następnie oblicza końcowy wynik dla danego kraju, zawierający się w przedziale od 333 do 1000 (Health Consumer Powerhouse, 2019). W ostatnich kilku edycjach pomiaru tym indeksem jakości ochrony zdrowia najwyższe miejsca w rankingu krajów były zajmowane prawie wyłącznie przez Holandię, Szwajcarię i Norwegię. W ostatnim wydaniu EHCI, w 2018 roku, pierwsze miejsce osiągnęła Szwajcaria z wynikiem 893 punktów na 1000 możliwych do uzyskania. Kolejne miejsca zajęły Holandia i Norwegia (Hampel i in., 2021).

Istotnym źródłem informacji o jakości usług związanych z ochroną zdrowia są również wyniki European Quality of Life Surveys [EQLS]. Jest to badanie mierzące obiektywne warunki życia w krajach Europy oraz związane z nimi odczucia obywateli. Jedną z kategorii kwestionariusza dotyczy zdrowia oraz opieki zdrowotnej i porusza kwestie związane z poziomem zdrowia, ograniczeniami związanymi z chorobami i niepełnosprawnością, częstością korzystania z opieki zdrowotnej, oczekiwaniem na usługi, a także jakością leczenia ambulatoryjnego i szpitalnego (Eurofound.europa.eu, 2016). Wyniki badania wskazują, że najczęściej korzysta się z podstawowej opieki zdrowotnej, podczas gdy opieka szpitalna i specjalistyczna jest mniej powszechna. Wśród państw Europy występują jednak rozbieżności dotyczące korzystania z usług medycznych – państwa, takie jak Austria i Niemcy wykazują wyższy odsetek wykorzystania wszystkich usług wymienionych w badaniu w porównaniu z Grecją czy Rumunią. Zestawienie wyników ankiety z kolejnych lat pokazuje, że średnia ocena usług opieki zdrowotnej w UE rośnie, przy czym w większości państw członkowskich usługi podstawowej opieki zdrowotnej otrzymują generalnie wyższe oceny niż usługi szpitalne, co podkreśla ich znaczenie. W szczególności w Luksemburgu, Rumunii i Włoszech odnotowuje się wyższe oceny usług podstawowej opieki zdrowotnej w porównaniu z opieką szpitalną lub specjalistyczną. Istnieją jednak wyjątki: Wielka Brytania, Portugalia i Słowacja wykazują odwrotny wzorzec. Mimo iż ogólny średni poziom zadowolenia z usług medycznych rośnie, istnieją rozbieżności między państwami członkowskimi UE, w których odnotowano różne poziomy tego wskaźnika. W szczególności Grecja, Łotwa, Cypr, Polska i Słowacja wykazują niższy poziom satysfakcji, a znaczny odsetek populacji w tych krajach ocenia usługi opieki zdrowotnej nisko (Dubois i in., 2019).

3. Metody badawcze i opis danych

Wielowymiarowa analiza porównawcza obejmuje szereg technik statystycznych mających na celu badania zjawisk opisywanych przez liczne zmienne. W jej ramach można wyróżnić m.in. metody taksonomiczne, czyli metody grupowania (dyskryminacyjne i klasyfikacyjne), oraz metody porządkowania liniowego (Dziechciarz, 2003).

Aby skutecznie stosować wielowymiarową analizę porównawczą, konieczne jest znormalizowanie zmiennych w celu zapewnienia jednolitych jednostek miary i wielkości. Jedną z metod normalizacji jest standaryzacja zapewniająca średnią arytmetyczną równą 0 i odchylenie standardowe równe 1. Unitaryzacja natomiast przekształca wartości zmiennych na przedział wartości zmiennych w przedziale [0;1] (Stanimir, 2006). Niektóre metody analizy porównawczej, w szczególności techniki porządkowania liniowego, wymagają ujednoczenia charakteru zmiennych. Do transformacji charakteru zmiennej można wykorzystać, m.in. formułę ilorazową lub różnicową (Gatnar i Walesiak, 2004).

3.1. Metody porządkowania liniowego

Jedną z najprostszych metod porządkowania liniowego jest metoda sum standaryzowanych. Aby zastosować tę metodę, zmienne muszą być standaryzowane i mieć charakter stymulant. Wszystkim cechom można przypisać jednakową wagność lub nadać im wagi w celu uwydatnienia ich wpływu na badane zjawisko. Dla każdego obiektu należy obliczyć sumę wartości zestandaryzowanych zmiennych cząstkowych stanowiącą zmienną syntetyczną (Dziechciarz, 2003):

$$p_i = \sum_{j=1}^m z_{ij}, \quad (1)$$

gdzie: z_{ij} – wartość i -tego obiektu j -tej zmiennej znormalizowanej.

Na podstawie otrzymanych sum można wykonać ranking obiektów, lecz należy zauważyć, że wartości te są nieunormowane, czyli wykraczają poza przedział [0;1]. Ponieważ przyjmuje się, że wszystkie zmienne są stymulantami, obiekty, które uzyskały wysoką wartość p_i są oceniane najlepiej.

Metoda wzorca rozwoju jest jedną z metod wzorcowych porządkowania liniowego. Do jej wykonania niezbędna jest normalizacja zmiennych i sprowadzenie nominant do postaci stymulanty lub destymulanty. Następnie należy wyznaczyć wzorzec (z_0) i antywzorzec ($z_{0'}$) rozwoju, będące abstrakcyjnymi obiektami o odpowiednio najlepszych i najgorszych wartościach dla każdej cechy (Stanimir, 2006):

$$z_0 = [z_{01} z_{02} \dots z_{0j} \dots z_{0m}], \quad (2)$$

$$\text{gdzie: } z_{0j} = \begin{cases} \max_i \{z_{ij}\} & \text{gdy zmienna } Z_j \text{ jest stymulantą} \\ \min_i \{z_{ij}\} & \text{gdy zmienna } Z_j \text{ jest destymulantą} \end{cases}$$

$$z_{-0} = [z_{-01} \ z_{-02} \dots \ z_{-0j} \dots \ z_{-0m}], \quad (3)$$

$$\text{gdzie: } z_{-0j} = \begin{cases} \max_i \{z_{ij}\} & \text{gdy zmienna } Z_j \text{ jest stymulantą} \\ \min_i \{z_{ij}\} & \text{gdy zmienna } Z_j \text{ jest destymulantą} \end{cases}$$

Kolejnym etapem jest obliczenie odległości każdego z obiektów od wzorca, n.p. z użyciem odległości euklidesowej, w celu zbadania podobieństwa do najlepszego obiektu. Następnie korzystając z odległości między wzorcem i antywzorcem, wyznacza się wartości zmiennej agregatowej (Stanimir, 2006):

$$m_i = 1 - \frac{d_{i0}}{d_0} \quad (i = 1, 2, \dots, n) \quad (4)$$

gdzie: m_i – miara rozwoju dla i -tego obiektu, d_{i0} – odległość i -tego obiektu od wzorca, d_0 – odległość między wzorcem rozwoju i antywzorcem.

Wartości zmiennej agregatowej zawierają się w przedziale od 0 do 1. Ponieważ antywzorec przyjmuje wartość 0, a wzorec 1, wobec tego obiekty z wyższą miarą rozwoju charakteryzują się wyższym poziomem badanego zjawiska złożonego.

Metoda TOPSIS stworzona przez Ch.-L. Hwanga i K. Yoona również zalicza się do wzorcowych metod porządkowania liniowego. Do jej przeprowadzenia zmienne powinny mieć charakter stymulant lub destymulant oraz być znormalizowane przekształceniem ilorazowym. Podobnie jak w metodzie wzorca rozwoju ustala się wzorec i antywzorec (Bąk, 2016):

$$z_0^+ = [z_{01} \ z_{02} \dots \ z_{0j} \dots \ z_{0m}], \quad (5)$$

$$\text{gdzie: } z_{0j}^+ = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych o charakterze stymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych o charakterze destymulant} \end{cases}$$

$$z_0^- = [z_{-01} \ z_{-02} \dots \ z_{-0j} \dots \ z_{-0m}], \quad (6)$$

$$\text{gdzie: } z_{0j}^- = \begin{cases} \max_i \{z_{ij}\} & \text{dla zmiennych o charakterze destymulant} \\ \min_i \{z_{ij}\} & \text{dla zmiennych o charakterze stymulant} \end{cases}$$

Następnie należy wyznaczyć odległość każdego obiektu zarówno od obiektu wzorcowego, jak i od antywzorca. Najczęściej stosowana jest odległość euklidesowa. Kolejnym etapem jest obliczenie zmiennej syntetycznej (Bąk, 2016):

$$q_i = \frac{d_{i0}^-}{d_{i0}^+ + d_{i0}^-} \quad (7)$$

gdzie: d_{i0}^+ – odległość obiektu i od wzorca, d_{i0}^- – odległość obiektu i od antywzorca.

Wartości zmiennej agregatowej należą do przedziału $[0;1]$. Im wyższą wartość miernika syntetycznego uzyskał obiekt, tym wyższy jest poziom zjawiska złożonego dla tego obiektu.

Dodatkowo istnieje możliwość segmentacji badanych obiektów na klasy w zależności od wartości wskaźnika syntetycznego. Mimo iż liczbę klas powinno się dostosować do liczebności zbiorowości, często sugeruje się podział na cztery klasy z wykorzystaniem średniej i odchylenia standardowego wskaźnika syntetycznego obiektów (Kisielińska i in., 2021):

Klasa I: $q_i \geq \bar{q} + s_q$,

Klasa II: $\bar{q} \leq q_i < \bar{q} + s_q$,

Klasa III: $\bar{q} - s_q \leq q_i < \bar{q}$,

Klasa IV: $q_i < \bar{q} - s_q$.

Wyniki dwóch metod porządkowania liniowego można porównać poprzez obliczenie współczynnika korelacji rang Spearmana, który umożliwia określenie zgodności uporządkowań obiektów, szczególnie gdy liczba obiektów jest ograniczona lub można im przypisać rangi. Podczas rangowania każdemu obiektowi przyporządkowuje się kolejne liczby całkowite, zaczynając od 1. Jeśli natomiast obiekty mają identyczne wartości zmiennej, należy im przypisać średnią arytmetyczną kolejnych rang (Ostasiewicz i in., 2006).

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (8)$$

gdzie: d_i – różnica między rangami odpowiadających sobie wartości zmiennej syntetycznej uzyskanych z dwóch metod porządkowania liniowego ($i = 1, 2, \dots, n$).

3.2. Analiza skupień

Analiza skupień jest jednym z działań wielowymiarowej analizy porównawczej i stanowi zbiór metod umożliwiających podział zestawu obiektów na wewnętrznie homogeniczne grupy, które jednocześnie różnią się od siebie nawzajem (Balicki, 2013). W ramach pojęcia metod grupowania można wyróżnić, m.in. metody hierarchiczne, które dzielą się na metody podziałowe i aglomeracyjne. Metody podziałowe (degglomeracyjne) polegają na dzieleniu jednego skupienia na kolejne, coraz mniej liczne grupy. W metodach aglomeracyjnych grupy zawierające początkowo po jednym elemencie następnie są łączone aż do momentu otrzymania jednego skupienia zawierającego wszystkie elementy (Frątczak i in., 2009). Grupowanie w ramach me-

tod aglomeracyjnych polega na znalezieniu najbardziej podobnych klas o najmniejszej odległości i połączeniu ich w jedną, nową klasę. Następnie należy ponownie obliczyć odległości między nową klasą a pozostałymi klasami za pomocą wybranej metody. Kroki te powtarza się do momentu, aż powstanie jedno skupienie zawierające wszystkie obiekty (Gatnar i Walesiak, 2004).

Mechanizm stosowany do obliczania odległości między klasami różni się w zależności od techniki grupowania. Można wyróżnić 7 najczęściej stosowanych metod aglomeracji: metoda najbliższego sąsiada, metoda najdalszego sąsiada, metoda mediany, metoda środka ciężkości nazywana inaczej metodą centroidalną, metoda średniej odległości wewnątrz skupień, metoda średniej odległości między skupieniami i metoda Warda (Frątczak i in., 2009).

W badaniu została wykorzystana metoda Warda, która stosuje „kryterium minimalnego wzrostu łącznej wewnątrzgrupowej sumy kwadratów odchyień wszystkich zmiennych dla każdego obiektu od ich średnich grupowych” (Balicki, 2013, s. 276). Aby uzyskać końcową strukturę skupień obiektów, należy przerwać wybrane połączenie na dendrogramie prezentującym kolejność łączenia grup. Decyzja dotycząca miejsca przecięcia wiązania jest zależna od badacza. Jedną z metod wyznaczania przerywanego wiązania jest skorzystanie z wykresu przebiegu aglomeracji w celu znalezienia dłuższego wiązania sugerującego niską homogeniczność łączonych klas (Panek, 2009). Inną metodą jest obliczenie kilku wersji indeksu sylwetkowego dla różnej liczby klas i wybór wariantu o najwyższej wartości (Dudek, 2020).

3.3. Opis danych

W analizie wykorzystano 9 zmiennych opisujących poziom zdrowia, jakość życia oraz funkcjonowanie systemów ochrony zdrowia w państwach członkowskich UE (tab. 1).

Tabela 1. Analizowane zmienne cząstkowe dotyczące zdrowia, jakości życia i ochrony zdrowia

| Zmienna | Nazwa | Opis |
|---------|--|---|
| 1 | 2 | 3 |
| X_1 | Liczba łóżek szpitalnych na 1000 osób | Łóżka znajdujące się w szpitalach publicznych, prywatnych, ogólnych i specjalistycznych oraz ośrodkach rehabilitacyjnych (Worldbank.org, 2022a). |
| X_2 | Liczba lekarzy na 1000 osób | Lekarze ogólni oraz specjaliści (Worldbank.org, 2022d). Z powodu braku danych z roku 2017 w niektórych krajach przyjęto dane z roku 2018 dla Bułgarii, z roku 2016 dla Chorwacji i Cypru oraz dane z roku 2015 dla Malty. |
| X_3 | Liczba pielęgniarek i położnych na 1000 osób | Pielęgniarki zawodowe, położne zawodowe, pielęgniarki pomocnicze, położne pomocnicze, pielęgniarki zarejestrowane, położne zarejestrowane i inny personel, taki jak pielęgniarki dentystyczne i pielęgniarki podstawowej opieki zdrowotnej (Worldbank.org, 2022c). Z powodu braku danych z 2017 roku w niektórych krajach dla Cypru oraz Malty przyjęto dane z roku 2018, dla Chorwacji z roku 2015, a dla Rumunii z roku 2016. |

Tabela 1., cd.

| 1 | 2 | 3 |
|-------|---|--|
| X_4 | Umieralność z powodu chorób układu krążenia, raka, cukrzycy lub chorób układu oddechowego w wieku dokładnie od 30 do 70 lat | Odsetek osób w wieku 30 lat, które przed ukończeniem 70. roku życia umarłyby z powodu choroby układu krążenia, nowotworu, cukrzycy lub przewlekłych chorób układu oddechowego. Przyjmuje się aktualne wskaźniki umieralności dla każdego wieku oraz brak innych przyczyn zgonu (Worldbank.org, 2022b). |
| X_5 | Lata życia w zdrowiu w chwili urodzenia | Oczekiwana liczba lat życia bez poważnych lub umiarkowanych problemów zdrowotnych u osoby w momencie urodzenia (Ec.europa.eu, 2020) |
| X_6 | Wydatki na ochronę zdrowia (% PKB) | Wydatki państw na ochronę zdrowia przedstawione jako odsetek Produktu Krajowego Brutto. Wskaźnik ten bierze pod uwagę rządowe oraz obowiązkowe mechanizmy finansowania (OECD, 2022). |
| X_7 | Niezaspokojone potrzeby z powodu listy oczekujących | Odsetek osób, które w ciągu ostatniego roku zgłosiły opóźnienie otrzymania opieki zdrowotnej z powodu długiej listy oczekujących (Ec.europa.eu, 2021). Z powodu braku danych dla wszystkich państw wykorzystano dane z roku 2019. |
| X_8 | Zgony niemowląt (na 1000 żywych urodzeń) | Liczba zgonów niemowląt (poniżej 1. roku życia) w przeliczeniu na 1000 żywych urodzeń (Ec.europa.eu, 2024). |
| X_9 | Umieralność przedwczesna (na 100 000 osób) | Liczba osób, których zgonów można uniknąć za pomocą interwencji medycznej, w tym wczesnego rozpoznania choroby i leczenia, w przeliczeniu na 100 000 osób (Ec.europa.eu, 2023b). |

Źródło: opracowanie własne.

Badanie obejmowało 28 państw członkowskich, w tym Zjednoczone Królestwo. Dane w większości pochodzące z 2017 roku zostały pozyskane ze stron internetowych Europejskiego Urzędu Statystycznego, Organizacji Współpracy Gospodarczej i Rozwoju (OECD) oraz Banku Światowego. Braki danych uzupełniano danymi z najbliższego dostępnego okresu. Za stymulanty uznano zmienne: X_1 , X_2 , X_3 , X_5 oraz X_6 , ponieważ ich wyższe wartości świadczą o wysokim poziomie jakości ochrony zdrowia. Natomiast zmienne: X_4 , X_7 , X_8 i X_9 są destymulantami, co oznacza, że wysokie wartości wskazują na niski poziom jakości ochrony zdrowia.

4. Wyniki analizy

4.1. Porządkowanie liniowe

W celu określenia, które kraje cechują się najwyższą jakością ochrony zdrowia ze względu na wszystkie zmienne jednocześnie, najpierw zamieniono destymulanty na stymulanty z użyciem formuły ilorazowej, macierz stymulant zestandaryzowano. Następnie zastosowano 3 metody porządkowania liniowego i w tabeli 2 porównano powstałe rankingi. Wyniki otrzymane metodą sum standaryzowanych oznaczono, metodą wzorca rozwoju – m , a metodą TOPSIS – q_i .

Tabela 2. Wyniki porządkowania liniowego

| Państwo | p_i | Państwo | p_i | Państwo | p_i |
|------------------|--------|------------------|-------|------------------|-------|
| Francja | 6,122 | Francja | 0,459 | Szwecja | 0,536 |
| Szwecja | 6,003 | Niemcy | 0,457 | Finlandia | 0,535 |
| Niemcy | 5,557 | Finlandia | 0,456 | Francja | 0,534 |
| Finlandia | 5,068 | Belgia | 0,442 | Niemcy | 0,525 |
| Włochy | 4,260 | Szwecja | 0,439 | Włochy | 0,502 |
| Cypr | 4,138 | Włochy | 0,395 | Cypr | 0,498 |
| Belgia | 4,033 | Irlandia | 0,384 | Belgia | 0,485 |
| Austria | 2,575 | Austria | 0,384 | Irlandia | 0,453 |
| Hiszpania | 1,988 | Czechy | 0,371 | Austria | 0,453 |
| Irlandia | 1,936 | Hiszpania | 0,363 | Hiszpania | 0,445 |
| Holandia | 1,351 | Holandia | 0,359 | Holandia | 0,428 |
| Zjedn. Królestwo | 0,571 | Zjedn. Królestwo | 0,357 | Zjedn. Królestwo | 0,418 |
| Dania | 0,155 | Cypr | 0,340 | Malta | 0,409 |
| Malta | 0,099 | Słowenia | 0,337 | Dania | 0,405 |
| Czechy | 0,061 | Luksemburg | 0,335 | Luksemburg | 0,400 |
| Luksemburg | 0,047 | Dania | 0,333 | Czechy | 0,398 |
| Słowenia | -0,452 | Portugalia | 0,332 | Grecja | 0,395 |
| Grecja | -0,456 | Grecja | 0,328 | Słowenia | 0,389 |
| Portugalia | -0,845 | Malta | 0,302 | Portugalia | 0,378 |
| Bułgaria | -2,921 | Litwa | 0,276 | Bułgaria | 0,358 |
| Litwa | -3,432 | Estonia | 0,259 | Rumunia | 0,342 |
| Rumunia | -3,707 | Węgry | 0,242 | Litwa | 0,324 |
| Estonia | -4,106 | Słowacja | 0,241 | Węgry | 0,303 |
| Węgry | -4,245 | Bułgaria | 0,238 | Estonia | 0,300 |
| Słowacja | -4,397 | Chorwacja | 0,218 | Polska | 0,282 |
| Polska | -4,972 | Rumunia | 0,215 | Słowacja | 0,281 |
| Chorwacja | -5,271 | Polska | 0,209 | Chorwacja | 0,255 |
| Łotwa | -9,158 | Łotwa | 0,113 | Łotwa | 0,180 |

Źródło: opracowanie własne.

Według ustalonych kryteriów, im wyższa wartość wskaźnika rozwoju, tym lepsza opieka zdrowotna w danym kraju. Ranking utworzony metodą sum standaryzowanych wskazuje, że Francja wykazuje się najlepszą jakością systemu ochrony zdrowia, a następnie plasują się Szwecja, Niemcy i Finlandia. Natomiast Słowacja, Polska, Chorwacja i Łotwa osiągnęły najniższe wskaźniki rozwoju. Warto zaznaczyć, że wartość syntetyczna dla Łotwy jest znacząco niższa w porównaniu z innymi wymienionymi państwami.

Uzeregowanie metodą wzorca rozwoju sugeruje, że Francja, Niemcy, Finlandia i Belgia posiadają systemy ochrony zdrowia o najwyższej jakości. Wśród czterech krajów o najgorszej jakości opieki zdrowotnej znalazły się Chorwacja, Rumunia, Polska i Łotwa.

Natomiast ranking stworzony z wykorzystaniem metody TOPSIS typuje Szwecję jako kraj o najlepszej ochronie zdrowia. Kolejno plasują się Finlandia, Francja i Niemcy. Najgorzej ocenione kraje to Polska, Słowacja, Chorwacja i Łotwa.

Łączna ocena za pomocą trzech metod pozwala stwierdzić, że kraje, w których jakość opieki zdrowotnej jest najlepsza w porównaniu do pozostałych państw Unii Europejskiej to Francja, Szwecja, Niemcy, Finlandia oraz Belgia. Najgorzej sklasyfikowana została Łotwa, Chorwacja, Polska, Słowacja i Rumunia, co świadczy o najniższej jakości ochrony zdrowia w tych państwach.

W tabeli 3 przedstawiono podział obiektów na 4 klasy według zmiennej syntetycznej. Klasa pierwsza oznacza obiekty o najlepszej jakości opieki zdrowotnej, natomiast klasa czwarta – obiekty o najgorszej jakości opieki zdrowotnej. W rankingach utworzonych metodami porządkowania liniowego można zaobserwować podobieństwa. Wszystkie stosowane metody przypisały Francję, Finlandię, Niemcy i Szwecję do klasy 1, natomiast Belgia, Cypr i Włochy zostały umieszczone w klasie 1 przez dwie spośród trzech użytych metod. W klasie drugiej znalazły się Austria, Dania, Irlandia, Holandia, Hiszpania i Zjednoczone Królestwo. Czechy, Luksemburg, Malta, Portugalia i Słowenia zajmują miejsca zarówno w klasie 2, jak i 3j, w zależności od metody klasyfikacji. Do klasy 3 przez wszystkie metody zostały przypisane Grecja i Litwa. Kraje, takie jak Bułgaria, Estonia, Węgry i Rumunia, plasują się między klasą 3 a czwartą, a Chorwacja, Łotwa, Polska i Słowacja zostały sklasyfikowane do klasy 4 przez wszystkie metody.

W celu sprawdzenia ogólnego podobieństwa poszczególnych par rankingów obliczono współczynnik Spearmana. Wszystkie metody wykazały wysokie podobieństwo w porządkowaniu obiektów, jednak wyniki uzyskane metodą sum standaryzowanych oraz metodą TOPSIS są skorelowane nasilniej (0,99), co sugeruje, że obie te metody przypisały wielu krajom identyczne lub bardzo zbliżone pozycje. Rankingi stworzone za pomocą metody sum standaryzowanych oraz metody wzorca rozwoju są skorelowane na poziomie 0,94. W przypadku porównania wyników uzyskanych metodą wzorca rozwoju i metodą TOPSIS wskaźnik korelacji Spearmana wyniósł 0,93.

Tabela 3. Przypisanie obiektów do klas według wartości zmiennej syntetycznej

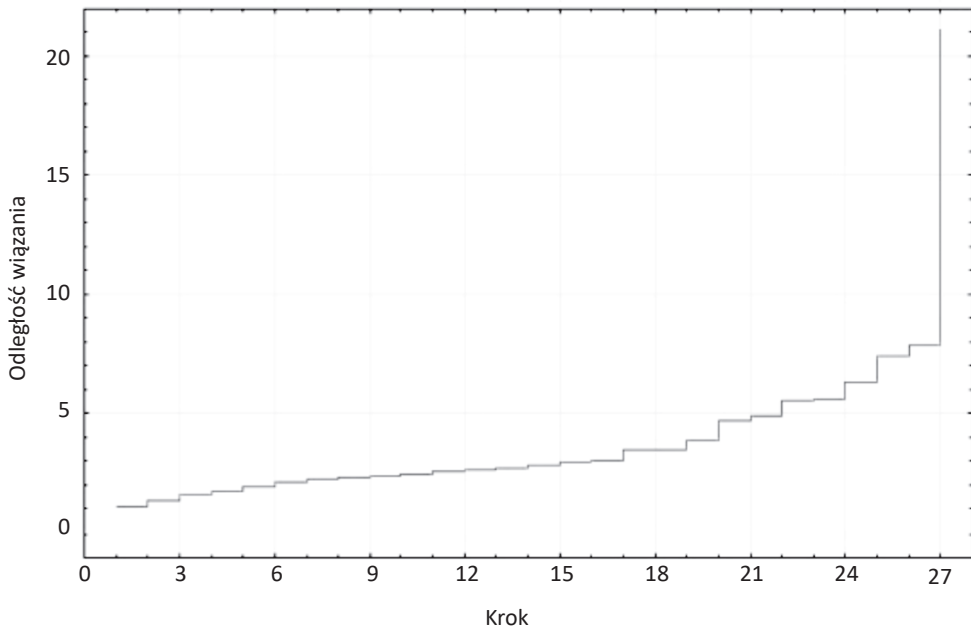
| Państwo | Metoda sum standaryzowanych | Metoda wzorca rozwoju | Metoda TOPSIS |
|------------------|-----------------------------|-----------------------|---------------|
| Austria | 2 | 2 | 2 |
| Belgia | 1 | 1 | 2 |
| Bułgaria | 3 | 4 | 3 |
| Chorwacja | 4 | 4 | 4 |
| Cypr | 1 | 2 | 1 |
| Czechy | 2 | 2 | 3 |
| Dania | 2 | 2 | 2 |
| Estonia | 4 | 3 | 4 |
| Finlandia | 1 | 1 | 1 |
| Francja | 1 | 1 | 1 |
| Niemcy | 1 | 1 | 1 |
| Grecja | 3 | 3 | 3 |
| Węgry | 4 | 3 | 4 |
| Irlandia | 2 | 2 | 2 |
| Włochy | 1 | 2 | 1 |
| Łotwa | 4 | 4 | 4 |
| Litwa | 3 | 3 | 3 |
| Luksemburg | 2 | 2 | 3 |
| Malta | 2 | 3 | 2 |
| Holandia | 2 | 2 | 2 |
| Polska | 4 | 4 | 4 |
| Portugalia | 3 | 2 | 3 |
| Rumunia | 3 | 4 | 3 |
| Słowacja | 4 | 4 | 4 |
| Słowenia | 3 | 2 | 3 |
| Hiszpania | 2 | 2 | 2 |
| Szwecja | 1 | 1 | 1 |
| Zjedn. Królestwo | 2 | 2 | 2 |

Źródło: opracowanie własne.

4.2. Analiza skupień

Do obliczenia macierzy odległości między obiektami wykorzystano odległość euklidesową, natomiast grupowanie przeprowadzono za pomocą metody Warda. Na rysunku 1 przedstawiającym odległość wiązania względem kroków aglomeracji można zauważyć, że wiązania numer 17 i 20 mają większe odległości od pozostałych wiązań, co sygnalizuje utratę homogeniczności skupień. Przerwanie wiązania 17 skutkowałoby utworzeniem 12 klas, natomiast przerwanie wiązania 20 spowodowałoby powstanie 9 klas. W celu porównania jakości klasyfikacji dla obu wymienio-

nych wariantów obliczono indeks sylwetkowy. Wartość indeksu dla klasyfikacji z 12 grupami wynosi w przybliżeniu 0,54 i jest większa od indeksu dla klasyfikacji z 9 grupami wynoszącego około 0,50. Oznacza to, że biorąc pod uwagę jedynie wskazania wykresu odległości wiązań oraz indeksu sylwetkowego, należy przeciąć 17 wiązanie i utworzyć 12 skupień.



Rys. 1. Wykres odległości wiązania względem etapu wiązania

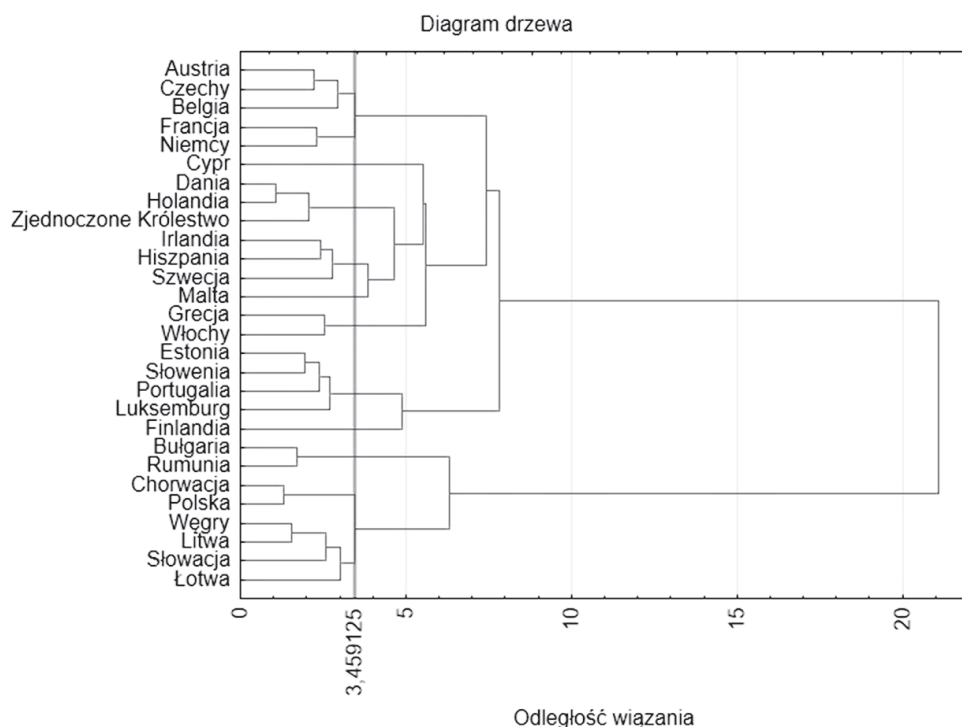
Źródło: opracowanie własne.

Rysunek 2 przedstawia otrzymany diagram drzewa, na którym znajdują się utworzone klasy. Linia zaznaczona jest miejsce przecięcia wiązania 17 o długości 3,459 łączącego Austrię, Czechy, Belgię, Francję i Niemcy.

Powstały następujące skupienia:

- skupienie 1: Austria, Czechy, Belgia,
- skupienie 2: Francja, Niemcy,
- skupienie 3: Cypr,
- skupienie 4: Dania, Holandia, Zjednoczone Królestwo,
- skupienie 5: Irlandia, Hiszpania, Szwecja,
- skupienie 6: Malta,
- skupienie 7: Grecja, Włochy,
- skupienie 8: Estonia, Słowenia, Portugalia, Luksemburg,
- skupienie 9: Finlandia,

- skupienie 10: Bułgaria, Rumunia,
- skupienie 11: Chorwacja, Polska,
- skupienie 12: Węgry, Litwa, Słowacja, Łotwa.



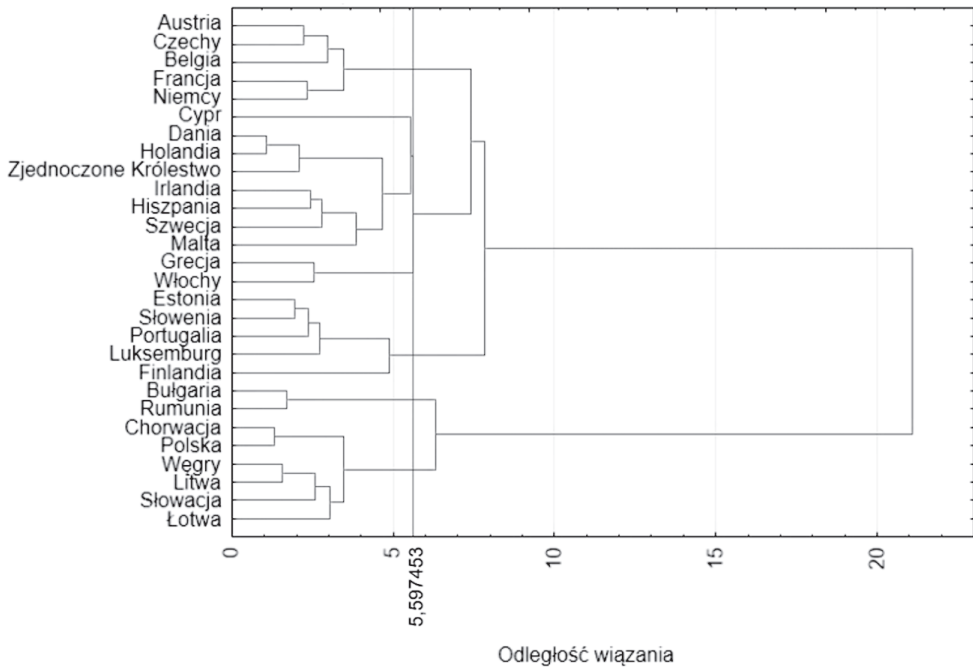
Rys. 2. Dendrogram z 12 klasami

Źródło: opracowanie własne.

Stworzenie 12 klas na podstawie 28 obiektów może prowadzić do zbyt szczegółowego podziału, co potwierdza obecność 3 klas jednoelementowych. Podjęto próbę zmniejszenia liczby grup przez poprowadzenie linii cięcia na dendrogramie w taki sposób, aby utworzonych zostało 6 klas.

Pierwsza grupa krajów obejmuje Austrię, Czechy, Belgię, Francję i Niemcy – są to kraje Europy Zachodniej cechujące się wysoką jakością opieki zdrowotnej. Następnie, druga grupa – obejmująca kraje śródziemnomorskie oraz część Europy Zachodniej i Północnej składa się z Cypru, Danii, Holandii, Zjednoczonego Królestwa, Irlandii, Hiszpanii, Szwecji i Malty. W podziale stworzonym według zmiennej syntetycznej kraje te najczęściej przypisywane są do klasy 1 lub 2. Grupę 3 stanowią Włochy i Grecja, reprezentujące Europę Południową z charakterystycznym klimatem śródziemnomorskim. Grupę 4 tworzą Estonia, Słowenia, Portugalia, Luksemburg i Finlandia. Jest to grupa o największym zróżnicowaniu, ponieważ kraje te zostały

przypisane do wszystkich 4 klas według zmiennej syntetycznej. Klaster 5 tworzą Bułgaria i Rumunia, które dołączyły do UE w 2007 roku i charakteryzują się podobną jakością opieki zdrowotnej. Ostatnia, klasa 6, obejmuje Chorwację, Polskę, Węgry, Litwę, Słowację i Łotwę, czyli państwa głównie położone w Europie Centralnej i Wschodniej. W większości przypadków kraje te znajdują się w klasie 4, co oznacza najniższą jakość opieki zdrowotnej w porównaniu z innymi państwami UE.



Rys. 3. Dendrogram z 6 klasami

Źródło: opracowanie własne.

Wprowadzenie subiektywnego podziału na mniejszą liczbę klas niż sugeruje wykres odległości wiązania względem etapu wiązania, powoduje powstanie skupień zawierających kraje, które osiągały podobne wyniki według metod porządkowania liniowego.

5. Zakończenie

Kraje, które zajęły wysokie miejsca w rankingach stworzonych metodami porządkowania liniowego, takie jak Francja, Szwecja, Niemcy lub Finlandia, mogą stanowić punkt odniesienia dla krajów o niskiej jakości ochrony zdrowia, np. Polski, Chorwacji i Łotwy. Analiza skupień wyodrębniła grupy krajów o podobnych systemach opieki

zdrowotnej, co sugeruje podobne wyzwania i potrzeby rozwiązań w celu poprawy jakości ochrony zdrowia. Kraje o złej opiece zdrowotnej charakteryzują się wysokimi wskaźnikami zachorowalności i umieralności na choroby niezakaźne przed 70. rokiem życia oraz krótkim życiem w zdrowiu, co wskazuje na problemy z profilaktyką, diagnostyką i leczeniem. Głównymi przyczynami niskiej jakości opieki zdrowotnej w tych krajach wydają się być niedofinansowanie i braki kadrowe wśród personelu medycznego.

Literatura

- Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*. Wydawnictwo Uniwersytetu Gdańskiego.
- Bąk, A. (2016). Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (426), 22-31.
- Busse, R., Klazinga, N., Panteli, D., i Quentin, W. (red.). (2019). *Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies*. World Health Organization. Pobrano z <https://iris.who.int/bitstream/handle/10665/327356/9789289051750-eng.pdf?sequence=1>
- Dubois, H., Leončikas, T., i Molinuevo, D. (2019, 11 lipca). *Quality of Health and Care Services in the EU*. Publications Office of the European Union. Pobrano z <https://www.eurofound.europa.eu/en/publications/2019/quality-health-and-care-services-eu>
- Dudek, A. (2020). Silhouette Index as Clustering Evaluation Tool. W: *Studies in Classification, Data Analysis, and Knowledge Organization*, 19-33. https://doi.org/10.1007/978-3-030-52348-0_2
- Dziechciarz, J. (red.). (2003). *Ekonometria: metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
- Ec.europa.eu. (b.d.). *ECHI – European Core Health Indicators*. Pobrano z https://health.ec.europa.eu/indicators-and-data/european-core-health-indicators-echi/echi-european-core-health-indicators_en
- Ec.europa.eu. (2024). *Infant Deaths Occurring in the Country by Cause*. Pobrano z [https://ec.europa.eu/eurostat/databrowser/view/HLTH_CD_INFO\\$DEFAULTVIEW/default/table](https://ec.europa.eu/eurostat/databrowser/view/HLTH_CD_INFO$DEFAULTVIEW/default/table)
- Ec.europa.eu. (2023). *Treatable and Preventable Mortality of Residents by Cause and Sex*. Pobrano z https://ec.europa.eu/eurostat/databrowser/view/HLTH_CD_APR__custom_4399314/default/table
- Ec.europa.eu. (2021). *Self-reported Unmet Needs for Health Care by Sex, Age, Specific Reasons and Educational Attainment Level*. Pobrano z https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_UN1E/default/table?lang=en
- Ec.europa.eu. (2020). *Healthy Life Years at Birth by Sex*. Pobrano z <https://ec.europa.eu/eurostat/databrowser/view/tps00150/default/table?lang=en>
- Eurofound.europa.eu. (2016). *European Quality of Life Survey 2016*. Pobrano z <https://www.eurofound.europa.eu/en/surveys/european-quality-life-surveys-eqls/european-quality-life-survey-2016>
- Frątczak, E., Gołata, E., Klimanek, T., Ptak-Chmielewska, A. i Pęczkowski, M. (2009). *Wielowymiarowa analiza statystyczna. Teoria – przykłady zastosowań z systemem SAS*. Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- Gatnar, E. i Walesiak, M. (2004). *Metody statystycznej analizy danych w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu we Wrocławiu.

- Hampel, K., Ucieklak-Jeż, P., i Bem, A. (2021). Health System Responsiveness in the Light of the Euro Health Consumer Index. *European Research Studies Journal*, 24(4B), 659-667. <https://doi.org/10.35808/ersj/2696>
- Healthpowerhouse.com. (2019). Pobrano z <https://healthpowerhouse.com/publications/>
- Kisielinińska, J. i in. (2021). *Wielowymiarowa analiza danych w ekonomice rolnictwa*. Wydawnictwo SGGW.
- McCullough, K., Andrew, L., Genoni, A., Dunham, M., Whitehead, L., i Porock, D. (2022). An Examination of Primary Health Care Nursing Service Evaluation Using the Donabedian Model: A Systematic Review. *Research in Nursing and Health*, 48(1), 159-176. <https://doi.org/10.1002/nur.22291>
- OECD. (2022). *Health spending*. Pobrano z <https://www.oecd.org/en/data/indicators/health-spending.html?oecdcontrol-00b22b2429-var3=2017>
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka. Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Panek, T. (2009). *Statystyczne metody wielowymiarowej analizy porównawczej*. Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- Stanimir, A. (red.). (2006). *Analiza danych marketingowych: problemy, metody, przykłady*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Worldbank.org. (2022a). *Hospital Beds (per 1,000 People)*. Pobrano z https://data.worldbank.org/indicator/SH.MED.BEDS.ZS?end=2017&name_desc=false&start=1960&view=map
- Worldbank.org. (2022b). *Mortality from CVD, Cancer, Diabetes or CRD between Exact Ages 30 and 70 (%)*. Pobrano z https://data.worldbank.org/indicator/SH.DYN.NCOM.ZS?name_desc=false
- Worldbank.org. (2022c). *Nurses and Midwives (per 1,000 People)*. Pobrano z https://data.worldbank.org/indicator/SH.MED.CMHW.P3?end=2017&name_desc=false&start=1960&view=map
- Worldbank.org. (2022d). *Physicians (per 1,000 People)*. Pobrano z https://data.worldbank.org/indicator/SH.MED.PHYS.ZS?end=2017&name_desc=false&start=1960&view=map

Analysis of the Quality of Health Care Systems in European Union Countries

Abstract: Access to quality health care is the right of every citizen of the European Union, but there are differences between countries in the quality of health care systems and their financing. The European Union's actions make it possible to coordinate the cooperation of countries, monitor the effects and accelerate the improvement of health care quality. The article compares the quality of health care systems in EU member states using linear ordering methods. Using cluster analysis, they were grouped into classes based on similarities in selected health system quality variables. In order to better understand the topic, the paper also includes a description of the indicators used to assess the quality of health care services and conclusions drawn from studies using these indicators.

Keywords: linear ordering, cluster analysis, health care