



## Natalia Adamczyk

Ocena atrakcyjności miejsca zamieszkania  
w przekroju powiatów województwa  
kujawsko-pomorskiego w 2020 roku

Assessment of the residential  
attractiveness across counties in the  
kujawsko-pomorskie voivodeship in 2020

Praca licencjacka

Promotor: dr hab. Marcin Szymkowiak, prof. UEP

Data przyjęcia:

Podpis promotora

Kierunek: Informatyka i ekonometria

Specjalizacja: Analityka gospodarcza

Poznań 2023



# Spis treści

<b>Wstęp</b>	<b>1</b>
<b>1 Preferencje nabywców przy wyborze miejsca zamieszkania – aspekty teoretyczne</b>	<b>2</b>
1.1 Determinanty atrakcyjności miejsca zamieszkania . . . . .	2
1.1.1 Rynek pierwotny i wtórny . . . . .	2
1.1.2 Rynek pracy . . . . .	4
1.1.3 Sytuacja społeczno–kulturowa obszaru . . . . .	6
1.1.4 Opieka medyczna . . . . .	7
1.1.5 Środowisko . . . . .	8
1.1.6 Bezpieczeństwo . . . . .	8
1.1.7 Transport . . . . .	9
1.2 Charakterystyka województwa kujawsko-pomorskiego . . . . .	9
1.3 Źródła danych i wybór zmiennych . . . . .	11
<b>2 Analiza porównawcza – ujęcie teoretyczne</b>	<b>15</b>
2.1 Metody porządkowania liniowego . . . . .	15
2.1.1 Metoda standaryzowanych sum . . . . .	17
2.1.2 Metoda TOPSIS . . . . .	18
2.1.3 Metoda wzorcowa Hellwiga . . . . .	20
2.2 Metody analizy skupień . . . . .	21
2.2.1 Metoda k–średnich . . . . .	22
2.2.2 Metoda łokcia . . . . .	24
2.2.3 Metoda Warda . . . . .	25

<b>3 Atrakcyjności miejsca zamieszkania w przekroju powiatów województwa kujawsko-pomorskiego</b>	<b>27</b>
3.1 Charakterystyka danych . . . . .	27
3.2 Rankingi porządkowania liniowego . . . . .	30
3.3 Grupowanie powiatów z wykorzystaniem analizy skupień . . . . .	36
3.4 Interpretacja wyników analizy taksonomicznej . . . . .	40
<b>Zakończenie</b>	<b>44</b>
<b>Spis tabel</b>	<b>47</b>
<b>Spis rysункów</b>	<b>48</b>
<b>Kody języka R</b>	<b>49</b>

# Wstęp

Decyzja o miejscu zamieszkania należy do najważniejszych wyborów jakich musimy dokonać wkraczając w życie dorosłe, ponieważ nasz wybór będzie definiował satysfakcję z prowadzonego życia codziennego. Jednak niezależnie od obszaru, każda lokalizacja będzie wiązała się z pewnymi wadami, a ich znaczenie będzie zależało od indywidualnych preferencji. Dlatego rozpoczynając poszukiwania warto zastanowić się nad swoimi priorytetami i potrzebami, aby podjąć świadomą decyzję.

W niniejszej pracy zostanie zbadana atrakcyjność zamieszkania powiatów w województwie kujawsko–pomorskim w 2020 r. W tym celu wykorzystano 3 metody porządkowania liniowego oraz 2 metody analizy skupień, które pozwalają uwzględnić wiele aspektów problemu związanego z właściwą oceną atrakcyjności obszaru.

Praca została podzielona na trzy rozdziały. W rozdziale pierwszym omówiono determinanty wpływające na decyzję osoby zmieniającej miejsce zamieszkania, opisano województwo kujawsko-pomorskiego oraz przedstawiono wstępную charakterystykę danych. Rozdział drugi zawiera szczegółowe wyjaśnienie procedur wykorzystanych metod porządkowania liniowego, jak i metod analizy skupień oraz opis sposobów interpretacji uzyskanych wyników, dzięki zaaplikowaniu każdej z metod. W trzecim rozdziale została przedstawiona procedura obróbki danych, omówione zostały otrzymane rankingi powiatów oraz grupy, a także dokonano porównania wyników otrzymanych z użyciem różnych metod. Do zaaplikowania wykorzystywanych metod został użyty język R, a kody implementujące metody można znaleźć w rozdziale 3.4 lub na stronie [github.com](https://github.com).

Do napisania pracy wykorzystano podręczniki i artykuły o różnorodnej tematyce m.in. książki omawiające zaawansowane zagadnienia statystyczne oraz publikacje poruszające kwestie ekonomiczne. Zdecydowana większość danych została pobrana z Banku Danych Lokalnych, którym zarządza Główny Urząd Statyczny. Wyjątkiem są zmienne, których źródłem jest [waszaedukacja.pl](https://waszaedukacja.pl).

# Rozdział 1

## Preferencje nabywców przy wyborze miejsca zamieszkania – aspekty teoretyczne

### 1.1 Determinanty atrakcyjności miejsca zamieszkania

#### 1.1.1 Rynek pierwotny i wtórny

Handel mieszkaniami odbywa się na tzw. rynku nieruchomości, który w literaturze zdefiniowany jest jako ogół czynności między uczestnikami rynku prowadzących do wykreowania popytu oraz podaży na nieruchomości. Rynek ten posiada trzy podstawowe funkcje (Jędrzyński, 2020):

- funkcję wymiany,
- funkcję informacyjną,
- funkcję korygowania wykorzystania przestrzeni.

Pierwsza funkcja pozwala na przeniesienie własności z właściciela na uczestnika rynku zgłaszającego popyt za ustaloną cenę. Cenę tą jesteśmy w stanie oszacować, dzięki funkcji informacyjnej rynku, pozwalającej na określenie wartości sprzedawanego lokalu przy pomocy dostępnych danych np. cen lokali o podobnych cechach do sprzedawanego. Ostatnia funkcja – funkcja korygowania wykorzystania przestrzeni – nakłania właści-

cieli do maksymalizacji wartości posiadanych nieruchomości m.in. poprzez wznoszenie budynków w sytuacji, gdy wartość samego gruntu spada.

Nabycia nieruchomości możemy dokonać na jednym z dwóch rodzajów rynku: rynku pierwotnym, bądź rynku wtórnym.

Rynek pierwotny charakteryzuje się ogłoszeniami nieruchomości, które zostały wystawione po raz pierwszy. Zazwyczaj są to nieruchomości sprzedawane w stanie deweloperskim, co wiąże się z szeregiem korzyści tj. minimalna ilość usterek przy odbiorze nieruchomości. Co więcej, jeżeli zauważymy pewne niedociągnięcia ze strony dewelopera<sup>1</sup> to mamy prawo do tzw. rękojmi, która upoważnia nas do zażądania naprawy, wymiany lub obniżenia ceny lokalu, a w niektórych przypadkach odstąpienia od umowy i wystąpienia o zwrot zapłaconej ceny. Kolejną zaletą rynku pierwotnego jest swoboda aranżacyjna, dzięki której możemy dopasować każdy aspekt nieruchomości do naszych indywidualnych preferencji. Możemy wybrać kolor farby, rodzaj paneli, ale również możemy wprowadzać bardziej zaawansowane zmiany, tj. rozmieszczenie gniazdek kontaktowych czy zmiana rozkładu ścian w mieszkaniu. Dzięki temu, możemy stworzyć mieszkanie idealnie dopasowane do naszych potrzeb i gustów, co często jest niemożliwe w przypadku zakupu gotowego lokalu na rynku wtórnym.

Ostatnim plusem rynku pierwotnego, jaki chciałabym omówić w tej pracy jest wysoki standard nowo-wybudowanych budynków oferujących coraz ciekawsze udogodnienia tj. smart home. Ponadto, coraz częściej deweloperzy stawiają na energooszczędność i tworzą osiedla z ekologicznymi rozwiązaniami, aby przyszłościowo zakup nie był przyjemny tylko dla klienta, ale również dla jego portfela. Niestety z im nowszymi lokalami mamy do czynienia, tym znajdują się one dalej od centralnych dzielnic miasta co wiąże się z dłuższymi dojazdami do pracy czy szkoły. Dodatkowo potencjalni inwestorzy muszą pamiętać, że decyzje o kupnie nieruchomości podejmuje się na wiele miesięcy przed odbiorem i wykończeniem mieszkania, więc zakup na rynku pierwotnym będzie się wiązać z m.in. dodatkowymi kosztami wynajmu.

W przypadku rynku wtórnego mamy do czynienia z mieszkaniami/budynkami, które są w posiadaniu osoby fizycznej – właściciela, który nie jest deweloperem. Oferty te bardzo często znajdują się bliżej centrum niż nowo powstałe osiedla. Z tego powodu cieszą się one podobnie wysokim zainteresowaniem co oferty z rynku pierwotnego. Bu-

---

<sup>1</sup>Deweloper to przedsiębiorca, który zajmuje się realizacją inwestycji budowlanych.

dynki w centralnych dzielnicach wyróżniają się pięknymi zdobieniami, których brakuje na nowo wybudowanych osiedlach. Niestety zakup mieszkania na rynku wtórnym wiąże się z pewnym zużyciem budynku, co może prowadzić do późniejszych problemów z ogrzewaniem czy hydrauliką. Wadą używanych mieszkań jest również konieczność posiadania pewnego kapitału zarezerwowanego na remont lokalu. Co prawda koszty remontu są rekompensowane niższymi cenami sprzedaży używanych lokali, ale często nie jest to wystarczająca rekompensata, aby zakup na rynku wtórnym był bardziej opłacalny od zakupu nowego mieszkania. Jeżeli jednak zależy nam na lokalu, do którego możemy się wprowadzić jak najszybciej to powinniśmy zacząć swoje poszukiwania właśnie od rynku wtórnego (Jędrzyński, 2020).

Jak widać oferty na obu rynkach posiadają zalety jak i wady, zatem wybór między nowością a wygodą zależy wyłącznie od indywidualnych preferencji. Niemniej jednak niezależnie od cech lokali, rynek nieruchomości ma znaczący wpływ na atrakcyjność zamieszkania regionu, a w szczególności możemy wyróżnić zniechęcający wpływ wysokich cen mieszkań na skłonność konsumentów do przeprowadzki do danego regionu. W niniejszej pracy będziemy brać pod uwagę jedynie ceny ofert rynku wtórnego z powodu niekompletnych danych ofert rynku pierwotnego.

### 1.1.2 Rynek pracy

Rynek pracy możemy zdefiniować jako ogólną czynności pomiędzy pracodawcami a osobami zatrudnionymi lub osobami bezrobotnymi<sup>2</sup>, które kształtują popyt<sup>3</sup> oraz podaż<sup>4</sup> pracy (Mortimer-Szymczak, 1995).

Jednym z ważniejszych czynników makroekonomicznych wpływających na sytuację panującą na rynku pracy jest **bezrobocie**, które możemy podzielić m.in. według przyczyn powstania. Przyjmując ten rodzaj podziału rozróżniamy (Nowosielska, 1995):

- bezrobocie przymusowe, czyli bezrobocie, w którym osoby aktywnie poszukujące pracy są zmuszone do pozostania bez pracy z powodu braku alternatywnych możliwości zatrudnienia,

---

<sup>2</sup>Osoby bezrobotne to osoby, które nie mają pracy ale jej aktywnie poszukują.

<sup>3</sup>Popyt na pracę to zapotrzebowanie pracodawców na pracę.

<sup>4</sup>Podaż pracy to liczba osób zdolnych do pracy.

- bezrobocie dobrowolne pojawia się, w momencie zbyt niskiego wynagrodzenia na rynku pracy,
- bezrobocie sezonowe wynikające z pory roku np. na północy Polski bezrobocie może wzrosnąć w okresach powakacyjnych,
- bezrobocie strukturalne spowodowane złym dopasowaniem dostępności pracy do zapotrzebowania na nią,
- bezrobocie frykcyjne (krótkotrwałe) wynikające z poszukiwania pierwszej lub kolejnej pracy,
- bezrobocie technologiczne spowodowane zastąpieniem pracowników maszynami,
- bezrobocie koniunkturalne będące skutkiem aktualnej sytuacji gospodarczej w danym regionie np. rezultat spowolnienia gospodarczego lub recesji.

Miernikiem nasilenia bezrobocia jest tzw. **stopa bezrobocia**, czyli stosunek osób niezatrudnionych, ale zarejestrowanych jako osoby poszukujące pracy i liczby osób w wieku produkcyjnym w danym regionie.

Jednak stopa bezrobocia nie jest wystarczającym miernikiem oceniającym sytuację na rynku pracy. Kolejnym parametrem, którego powinniśmy użyć do takiej oceny jest **średnie wynagrodzenia** pracowników w danym regionie, którego wysoka wartość przyciągnie bardziej wykwalifikowanych pracowników ceniących sobie wyższy poziom życia. Warto również wziąć pod uwagę liczbę wakatów, czyli liczbę wolnych miejsc pracy, których wyższa wartość będzie świadczyła o lepszych możliwościach znalezienia pracy w danym regionie.

Sytuacja na rynku pracy danego regionu jest silnie powiązana z jego atrakcyjnością zamieszkania. Udowodnić to można rozpatrując następujące sytuacje. Jeśli w danym regionie będzie panowała wysoka stopa bezrobocia, a liczba nowych ofert pracy będzie niska, to mieszkańcy zaczną szukać miejsca zatrudnienia w innych miastach, a co za tym idzie zaczną również myśleć o przeprowadzce do miejsca znajdującego się bliżej nowemu miejscu zatrudnienia. Z kolei, jeśli w powiecie istnieją dobre możliwości zatrudnienia oraz poziom średniego wynagrodzenia jest zadowalający, to przyciągnie to nowych mieszkańców, którzy chcą osiedlić się w okolicy charakteryzującej się wyższym poziomem życia oraz łatwiejszym procesem znalezienia pracy.

### **1.1.3 Sytuacja społeczno-kulturowa obszaru**

Sytuacja społeczno-kulturowa konkretnego obszaru to zbiór różnych aspektów życia codziennego odzwierciedlający charakter regionu. Aspekty te obejmują m.in. strukturę demograficzną obszaru, poziom wykształcenia obywateli, dostępność do placówek kulturowych, sportowych i turystycznych oraz dostępność ośrodków opiekuńczych.

Zgodnie z artykułem 35 Prawa Oświatowego (Olszewski i Pilich, 2022) każdy obywatel Rzeczypospolitej Polskiej ma obowiązek podjęcia nauki w roku (kalendarzowym), w którym kończy 7 lat, a zakończyć ją może dopiero po osiągnięciu pełnoletniości. Proces edukacji może odbywać się w placówce publicznej, prywatnej lub w zaciszu domowym, obejmując wychowanie przedszkolne, szkołę podstawową i szkołę średnią. Jednakże, w sytuacji, gdy rodzic (rodzice) pragnie kontynuować swoją karierę zawodową krótko po narodzinach niemowlęcia, może on skorzystać z dodatkowej opcji oddania swojego dziecka pod opiekę wykwalifikowanych pracowników, np. w żłobku. Ośrodki te zapewniają doskonałe warunki dla najmłodszych, by nabywały one umiejętności społeczne i poznawcze. Co więcej, w przypadku korzystania z tej formy opieki, sytuacja majątkowa rodziców nie odgrywa roli, dzięki czemu każdy podopieczny ma zapewniony równy dostęp do rozwoju.

Jako weryfikację zgromadzonej wiedzy, na zakończenie szkoły średniej absolwenci zdają egzamin maturalny z 3 obowiązkowych przedmiotów (matematyki, języka polskiego i języka obcego), a także z maksymalnie 6 przedmiotów dodatkowych na poziomie rozszerzonym. Wynik z egzaminu maturalnego decyduje o dalszej karierze naukowej absolwenta, co ma wpływ na jakość życia oraz możliwości znalezienia pracy po ukończeniu szkoły wyższej.

Dlatego też, osoby zmieniające miejsce zamieszkania będą chciały zmaksymalizować efekty swojej edukacji, bądź edukacji swoich dzieci, wybierając miejscowości z najwyższymi wynikami egzaminów. W związku z tym maturzyści będą przenosić się w okolice wymarzonej uczelni, rodzice w okolice renomowanych szkół podstawowych i licealnych, a osoby starsze prawdopodobnie podążą za swoimi dziećmi lub wnukami, jeżeli w ogóle zdecydują się na przeniesienie.

Ze względu na brak dostępności informacji o średnich wynikach egzaminu maturalnego w danym powiecie, w analizie zostały uwzględnione wyniki egzaminu ósmoklasisty. Dodatkowo, w związku ze wzrostem popularności języka angielskiego jako sposobu

komunikacji w pracy, turystyce i Internecie, lekcje języka obcego zostały również uznane jako kryterium oceniające jakość kształcenia w wybieranej placówce edukacyjnej.

Kolejnym istotnym elementem wpływającym na sytuację społeczno–kulturową regionu jest aspekt kulturowy, czyli dostępność do ośrodków tj. muzea, teatry czy świątlice. Wspomniane placówki oferują mieszkańcom możliwość uczestniczenia w różnego rodzaju wydarzeniach kulturowych, które nie tylko wzbogacają życie kulturowe mieszkańców, ale również zwiększą atrakcyjność turystyczną regionu. Dzięki temu, obszary o szerszej ofercie kulturowej silniej przyciągają nowych mieszkańców.

#### **1.1.4 Opieka medyczna**

Zgodnie z art. 68 Konstytucji Rzeczypospolitej Polskiej (Tuleja i in., 2019) każdy obywatel ma prawo do ochrony zdrowia, a świadczenia zdrowotne są udzielane przez Państwo zgodnie z normą konstytucji mówiącej o równości wobec prawa. Wyjątkiem są osoby niepełnoletnie, kobiety ciężarne, osoby niepełnosprawne oraz seniorzy, którzy mają prawo do szczególnej opieki.

W Polsce opieka medyczna jest przyznawana na podstawie posiadanych ubezpieczeń. Wyróżnia się dwa rodzaje ubezpieczenia: obowiązkowe i dobrowolne. Ubezpieczenie obowiązkowe przysługuje każdej osobie spełniającej warunki określone w Ustawie z dnia 27 sierpnia 2004 r. o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych. Zapewnia ona m.in. konsultacje lekarskie, refundacje recept oraz pobyt w szpitalu. Jednak ze względu na coraz dłuższe okresy oczekiwania na wizytę w publicznych placówkach medycznych, coraz większa część społeczeństwa decyduje się na zakup dodatkowych ubezpieczeń lub na wizytę w prywatnym gabinecie lekarskim. Przykładem dobrowolnego ubezpieczenia zdrowotnego jest np. polisa zapewniająca zabezpieczenie finansowe w sytuacji utraty wynagrodzenia z powodu długotrwałego pobytu w szpitalu na skutek choroby czy nieszczerśliwego wypadku.

Niezależnie od posiadanych świadczeń zdrowotnych dostępność opieki medycznej oraz oferty aptek stanowi ważną determinantę wpływającą na jakość życia mieszkańców danego obszaru. Szybki i łatwy dostęp do konsultacji lekarskich oraz lekarstw w razie choroby jest niezwykle ważny dla zdrowia i poczucia bezpieczeństwa społeczności, szczególnie jest to ważne dla osób starszych. Dlatego brak wysoko rozwiniętej infrastruktury medycznej wpłynie negatywnie na subiektywną ocenę atrakcyjności obszaru.

### **1.1.5 Środowisko**

Stan środowiska, w którym żyjemy ma znaczący wpływ na nasilenie występowania chorób na danym obszarze oraz stan zdrowia obywateli.

Jednakże rozwój społeczeństwa przyczynił się do znaczającej degradacji środowiska na przestrzeni ostatnich lat m.in. poprzez nieodpowiedzialne wykorzystanie zasobów naturalnych oraz brak uwzględniania konsekwencji podjętych działań stawiamy dzisiaj czoło wielu globalnym zagrożeniom (Ginsbert-Gebert, 1984).

Możemy wyróżnić kilka rodzajów pogarszającego się stanu środowiska: zanieczyszczenie powietrza, zanieczyszczenie wody, zanieczyszczenie gleby, zanieczyszczenie krajobrazu, zanieczyszczenie światłem i hałasem oraz skażenie promieniotwórcze. Przykładowym scenariuszem ilustrującym negatywny wpływ zanieczyszczeń na nasze zdrowie jest spożywanie warzyw uprawianych w toksycznej glebie, które skutkuje przedostaniem się toksyn do naszego organizmu, a globalnym przykładem mogą być skutki związane z zjawiskami tj. smog czy dziura ozonowa jakimi jest m.in. zwiększone prawdopodobieństwa zachorowania na choroby nowotworowe tj. czerniak (Gniadek i Marcisz, 2014).

Wspólnie z wieloma krajami europejskimi Polska stara się przeciwdziałać postępującej degradacji środowiska poprzez recykling, promowanie odnawialnych źródeł energii oraz zmniejszenie emisji dwutlenku węgla do atmosfery. Przez nagłośnienie szkodliwych efektów charakteryzujących obszary ozłym stanie środowiska społeczeństwo zaczęło preferować dzielnice znajdujące się w pobliżu terenów zielonych, daleko od fabryk czy autostrad. Dlatego też, obszary charakteryzujące się wieloma terenami zielonymi będą atrakcyjniejsze dla osób ceniących sobie dobry stan środowiska.

### **1.1.6 Bezpieczeństwo**

Bezpieczeństwo możemy zdefiniować jako stan braku zagrożenia dla naszego życia lub zdrowia. Sprzyja ono nie tylko dobremu samopoczuciu obywateli, ale również rozwojowi gospodarczemu obszaru. W związku z tym, powstał art. 5 Konstytucji Rzeczypospolitej Polskiej (Tuleja i in., 2019) zapewniający obywateli, że Państwo będzie strzegło niepodległości swojego terytorium oraz będzie stało na straży wolności i praw człowieka. W tym celu powołano różne rodzaje służb porządkowych, które specjalizują się

w zapobieganiu lub likwidacji konkretnych zagrożeń (Bogacka i Siniecka, 2016).

Wśród wielu aspektów wpływających na poczucie bezpieczeństwa, a jednocześnie wpływających na decyzję o wyborze nowego miejsca zamieszkania jest efektywność jednostek policyjnych w minimalizacji aktywności przestępcości, bądź ich skuteczność w znajdowaniu przestępcołów odpowiedzialnych za popełnioną zbrodnię (Krztoń, 2017). Wyższa efektywność służb porządkowych bezpośrednio wpłynie na wzrost zainteresowania zamieszkaniem na danym obszarze.

### **1.1.7 Transport**

Jakość transportu jest również ważnym aspektem wpływającym na decyzję o wyborze miejsca zamieszkania. Dostępność różnych form komunikacji tj. komunikacja miejska czy drogi rowerowe, umożliwia mieszkańców swobodne poruszanie się po mieście pozytywnie wpływając na jakość prowadzonego przez nich życia. Co więcej, zgodnie z przeprowadzonymi badaniami, wygoda transportowa jest jednym z kluczowych czynników wpływających na prawdopodobieństwa zatrudnienia. Jeżeli pracownik posiada stałe miejsce zatrudnienia oraz mieszka daleko od przedsiębiorstwa (bez możliwości wygodnego dojazdu do pracy) to zacznie rozważać przeprowadzkę w celu zminimalizowania czasu potrzebnego na dotarcie do pracy.

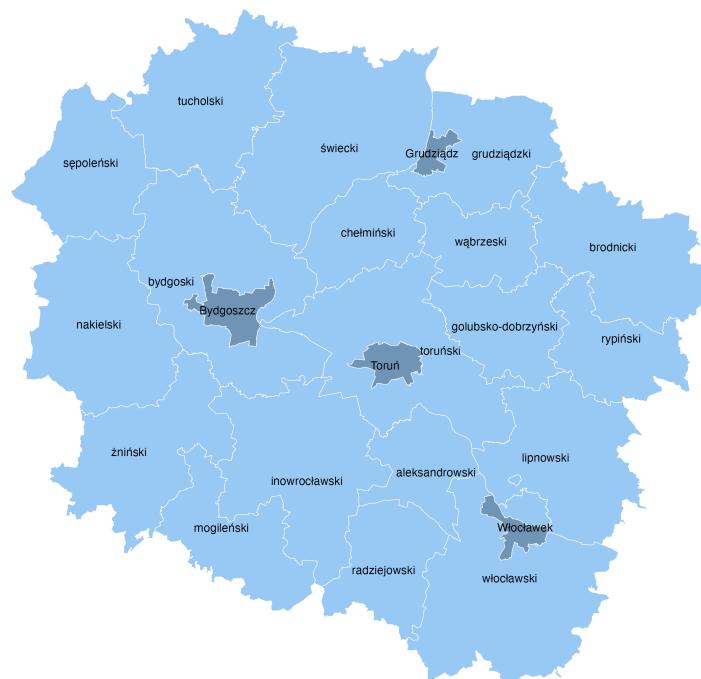
Podsumowując, osoby mające dostęp do szybkich i wygodnych środków transportu, będą bardziej pozytywnie nastawione do życia w danej miejscowości m.in. ze względu na mniejsze koszty transportu oraz oszczędność czasu. Dlatego też, rozwinięte formy transportu publicznego oraz infrastruktury drogowej stanowią ważne kryterium brane pod uwagę przez przyszłych mieszkańców.

## **1.2 Charakterystyka województwa kujawsko-pomorskiego**

Województwo kujawsko–pomorskie położone jest w środkowo-północnej części Polski i graniczy z 5 województwami: łódzkim, mazowieckim, pomorskim, warmińsko-mazurskim oraz wielkopolskim. Obszar województwa wynosi ok. 18 000 km<sup>2</sup>, w którego skład wchodzą 3 makroregiony: Pojezierze Południowo-pomorskie, Pojezierze

Chełmińsko-Dobrzyńskie i Pojezierze Wielkopolskie. Na terenie województwa znajdują się m. in. złoża soli kamiennej oraz punkty wydobycia wapienia i margli. Z budową geologiczną obszaru wiążą się również złoża węgla brunatnego oraz zasoby kruszywa, piasków kwarcowych, kredy jeziornej i torfu.

Teren województwa podzielony jest na 19 powiatów oraz 4 miasta na prawach powiatu widocznych na rys. 1.1. Podobnie do województwa lubuskiego posiada ono 2 stolice: stolicę administracyjną w Bydgoszczy i stolicę samorządową w Toruniu („Województwo kujawsko-pomorskie”, 2023).



**Rysunek 1.1. Podział administracyjny województwa kujawsko-pomorskiego**

**Źródło:** Opracowanie własne

Większość obszaru województwa kujawsko-pomorskiego stanowią tereny oddane do użytku rolnego, dlatego gleby są jednym z najważniejszych zasobów naturalnych regionu. Pamiętając, że obszar województwa stanowi zaledwie 5,57% terenu kraju uwagę przyciągają następujące wskaźniki dla 2020 r. (Urząd Statystyczny, 2021):

- na 100 tys. ton zebranego zboża w kraju aż 9,3% zostało zebrane na terenie województwa kujawsko-pomorskiego,

- na 100 tys. ton zebranych ziemniaków w kraju aż 8,5% zostało zebrane na terenie województwa kujawsko–pomorskiego,
- na 100 tys. ton zebranych warzyw gruntowych w kraju aż 18,2% zostało zebrane na terenie województwa kujawsko–pomorskiego,
- z 1 ha plonów zboża w województwie uzyskano 54,4 dt<sup>5</sup>, gdzie z takiej samej powierzchni w Polsce uzyskano 47,9 dt.

W 2020 r. województwo było zamieszkiwane przez około 2 mln mieszkańców, wśród których ok. 60% było w wieku produkcyjnym. Najwięcej osób zatrudnionych było w przetwórstwie przemysłowym, gdzie przeciętne wynagrodzenie brutto wyniosło 4565,42 zł. Jeżeli chodzi o ogólne przeciętne wynagrodzenie to wzrosło ono z 2019 r. na 2020 r. o ok. 300 zł i wyniosło 4601,76 zł. Pomimo wzrostu było ono nadal niższe od krajowego wynagrodzenia o ponad 600 zł. Innym ciekawym wskaźnikiem opisującym sytuację dochodową obywateli jest przeciętny miesięczny dochód rozporządzalny na 1 osobę w gospodarstwie domowym<sup>6</sup>, który wyniósł zaledwie 1845,34 zł, z czego ponad 63% musiało zostać przeznaczone na miesięczne wydatki przypadające na 1 osobę (w gospodarstwie domowym).

Ostatnim wskaźnikiem omówionym w tej sekcji jest stopa bezrobocia rejestrowanego w województwie, która wyniosła 9,0% i była o 2,7 pkt. proc. wyższa od krajowej stopy bezrobocia rejestrowanego. Powiaty charakteryzujące się najwyższymi stopami bezrobocia rejestrowanego znajdowały się na południowych obrzeżach województwa.

### **1.3 Źródła danych i wybór zmiennych**

Najbardziej logicznym kryterium decydującym o przyszłym miejscu zamieszkania jest szacowany wydatek, który wiąże się z kupnem nieruchomości w wybranej miejscowości. Jednak przy coraz dłuższych poszukiwaniach, nasza lista wymagań co do miejsca zamieszkania najprawdopodobniej ulegnie wydłużeniu. Dlatego warto stworzyć listę wymagań, zdefiniować istotność każdego z nich oraz wykreślić te najmniej istotne.

---

<sup>5</sup>decytona (dt) = 0,1 tony

<sup>6</sup>Średnia liczba osób w gospodarstwie domowym została oszacowana na poziomie 2,61 w 2020 r.

Na podstawie subiektywnej oceny znaczenia wskaźników dostępnych w Internecie dotyczących atrakcyjności zamieszkania obszaru, wybrano najbardziej istotne determinanty, a następnie przypisano je do jednej z 7 kategorii:

1. Sytuacja na rynku nieruchomości:

- **wskaźnik cen lokali** to średnia cena lokali mieszkalnych sprzedanych w ramach transakcji rynkowych na rynku wtórnego. Jego wysoka wartość zniechęca potencjalnych inwestorów do kupna, dlatego jest to destymulanta,
- **wskaźnik powierzchni mieszkania** to przeciętna powierzchnia użytkowa mieszkania przypadająca na 1 osobę [ $m^2$ ]. Im wyższa wartość wskaźnika, tym wygodniej będzie się żyło przyszłym lokatorom w mieszkaniu, dlatego wskaźnik jest stymulanta,
- **wskaźnik mieszkań** to liczba mieszkań przypadających na 1000 mieszkańców. Im niższa wartość wskaźnika tym trudniej zamieszkać w danym powiecie, dlatego jest to stymulanta.

2. Sytuacja na rynku pracy:

- **wskaźnik ofert pracy** to liczba ofert pracy, które pojawiły się w ciągu analizowanego roku w przeliczeniu na 1000 osób. Im wyższa jego wartość tym łatwiej mieszkańcom znaleźć zatrudnienie, dlatego jest to stymulanta,
- **wskaźnik bezrobocia** to stopa bezrobocia rejestrowanego danego powiatu; jego wzrost nie jest korzystny dla społeczeństwa, ponieważ świadczy o problemach gospodarki powiatu, więc wskaźnik jest destymulanta,
- **wskaźnik dochodu** to przeciętne miesięczne wynagrodzenia brutto mieszkańców obszaru; wraz z jego spadkiem komfort życia ludności spada, więc będzie to stymulanta.

3. Sytuacja społeczno-kulturowa:

- **wskaźnik kultury** to liczba centrów, domów i ośrodków kultury, klubów i świetlic przypadających na 1000 osób – im wyższa jego wartość tym więcej zajęć dodatkowych będzie oferował powiat, co jest atrakcyjne z punktu widzenia potencjalnych mieszkańców miejscowości, dlatego wskaźnik jest stymulantą,

- **wskaźnik rozrywki** to liczba imprez w przeliczeniu na 1000 osób w danym powiecie – wysoka wartość wskaźnika wskazuje na powiat pełen weekendowych możliwości, co będzie przyciągało nowych lokatorów. Wskaźnik jest stymulantą,
- **wskaźnik edukacji podstawowej** to średnia przeciętnych wyników egzaminu ósmoklasisty z języka polskiego, języka angielskiego i matematyki – oczywiście jego wysoki poziom oznacza sukces nauczania dzieci w danej szkole, więc będzie to stymulanta,
- **wskaźnik nauczania języków obcych w szkole podstawowej** to udział uczniów uczących się języka angielskiego w ogólnej liczbie uczniów w szkole podstawowej – jego niski poziom nie jest korzystny, ze względu na wysoką współpracę z firmami zagranicznymi polskich przedsiębiorstw; z tego powodu w analizie wskaźnik będzie traktowany jako stymulanta,
- **wskaźnik nauczania języków obcych w szkole licealnej** to udział uczniów uczących się języka angielskiego w ogólnej liczbie uczniów w szkole licealnej; jego interpretacja jest analogiczna do interpretacji wskaźnika nauczania języków obcych w szkole podstawowej,
- **wskaźnik opieki nad dziećmi** to liczba miejsc w żłobkach i klubach dziecięcych na 1000 dzieci w wieku do lat 3. Im wyższa jego wartość tym łatwiej znaleźć miejsce dla dziecka w żłobku lub klubie dziecięcym, dlatego jest to stymulanta.

#### 4. Opieka medyczna mieszkańców:

- **wskaźnik dostępności lekarstw** to liczba ogólnodostępnych aptek przypadających na 1 tys. osób; jego wysoka wartość sugeruje lepszą dostępność potrzebnych leków w najbliższej aptece – jest to stymulanta,
- **wskaźnik medycyny** to liczba lekarzy pracujących na 1 tys. ludności; jego wysoka wartość oznacza krótsze okresy oczekiwania na konsultację, co zaczęci osoby do wyboru danego obszaru jako swojego miejsca zamieszkania – stymulanta.

#### 5. Stan środowiska naturalnego:

- **wskaźnik zieleni** to udział lasów w ogólnej powierzchni powiatu; wskaźnik jest stymulantą, ponieważ jego wysoka wartość sprzyja zdrowemu trybowi życia obywateli.

## 6. Poziom bezpieczeństwa regionu:

- **wskaźnik wykrywalności sprawców przestępstw** stwierdzonych przez Policję będzie informował nas o prawdopodobieństwie znalezienia osób odpowiedzialnych za szkodę. Jego niska wartość jest niekorzystna dla obszaru, ponieważ będzie świadczyła o słabej skuteczności działań policji, dlatego wskaźnik będzie stymulantą,
- **wskaźnik przestępstw** to liczba przestępstw stwierdzonych przez Policję w przeliczeniu na 1000 mieszkańców to istotny parametr, ponieważ informuje o częstotliwości zagrożeń wynikających z działań ludzkich. Im wyższa jego wartość tym obywatele mogą czuć mniej bezpiecznie w swoim powiecie, dlatego wskaźnik jest destymulanta,
- **wskaźnik miejscowych zagrożeń** na 1000 ludności, jest destymulanta, ponieważ wraz z jego wzrostem wzrośnie prawdopodobieństwo wypadku/nieszczęścia w danym regionie przez wzrośnie również prawdopodobieństwo odpływu ludności z danego obszaru.

## 7. Transport:

- **wskaźnik przystanków komunikacji miejskiej** to liczba przystanków autobusowych (z trolejbusowymi) i tramwajowych; wskaźnik ten będzie stymulantą, ponieważ wraz z jego wzrostem poruszanie się po terenie powiatu będzie łatwiejsze i szybsze.

Szersza charakterystyka danych, na których opiera się niniejsza praca zostanie przedstawiona w rozdziale 3.1 z uwzględnieniem źródeł danych oraz transformacji.

## Rozdział 2

# Analiza porównawcza – ujęcie teoretyczne

Taksonomia to dziedzina nauki zajmująca się zasadami i procedurami porządkowania, grupowania oraz dyskryminacji obiektów. Umożliwia ona analizę obiektów (zjawisk) złożonych, czyli takich, których ostateczna ocena zależy od wielu kryteriów (Bąk i in., 2015).

### 2.1 Metody porządkowania liniowego

Metody porządkowania liniowego to jedne z najpopularniejszych metod takonomicznych stosowanych w różnych dziedzinach naukowych tj. ekonomia czy nauki społeczne. Swoją popularność zawdzięczają prostocie interpretacji wyników oraz szybkości przeprowadzanych obliczeń. Ich celem jest uporządkowanie zbioru obiektów według wygenerowanej **miary syntetycznej** na podstawie wybranych kryteriów, która pozwala uwzględnić wiele cech obiektu jednocześnie. Takie podejście pozwala otrzymać wyniki bliższe rzeczywistości (Bąk i in., 2015).

W zależności od pożądanych wartości cechy wyróżniamy trzy rodzaje kryteriów: **stymulantę**, **destymulantę** oraz **nominantę**. Możemy zakwalifikować cechę jako **stymulantę** jeżeli preferujemy jej wysokie wartości bez określonego maksimum. W przeciwej sytuacji, czyli gdy wzrost wartości cechy wpływa negatywnie na pozycję obiektu w rankingu to klasyfikujemy cechę jako **destymulantę**. Przykładem takiego kryterium będzie np. długość kolejki w sklepie czy ilość korków w mieście. Natomiast **nominanta** to taka

cecha, która wpływa pozytywnie na pozycję obiektu w rankingu wtedy i tylko wtedy, gdy jej wartość zawiera się w pewnym liczbowym przedziale zamkniętym (Kozioł, 2013). Omówiona klasyfikacja cech jest niezwykle ważna, ponieważ pozwala na lepsze zrozumienie wpływu poszczególnych cech na ostateczny ranking obiektów.

Po poznaniu powyższych definicji możemy przejść do omówienia ogólnego algorytmu stosowania metod porządkowania liniowego (Bąk i in., 2018):

1. Zdefiniowanie problemu badawczego.
2. Wybór kryteriów, określenie ich rodzaju oraz przypisanie im wag.
3. Wybór metody porządkowania w zależności od charakterystyki wybranych kryteriów.
4. Ocena obiektów i utworzenie rankingu.
5. Analiza wyników i wyciągnięcie wniosków.

Na zakończenie warto wiedzieć, że uporządkowane obiekty dowolną metodą porządkowania liniowego spełniają następujące założenia:

- każdy obiekt ma przynajmniej jednego sąsiada i nie więcej niż dwóch sąsiadów<sup>1</sup>,
- jeżeli sąsiadem  $i$ -tego obiektu jest  $k$ -ty obiekt, to jednocześnie sąsiadem  $k$ -tego obiektu jest  $i$ -ty obiekt,
- dokładnie dwa obiekty mają tylko jednego sąsiada.

W niniejszej pracy do analizy atrakcyjności zamieszkania powiatów województwa kujawsko–pomorskiego wykorzystamy następujące metody: metodę standaryzowanych sum, metodę wzorca rozwoju Z. Hellwiga oraz metodę TOPSIS. Pierwsza z wymienionych metod należy do kategorii **metod bezwzorcowych**, a kolejne dwie należą do tzw. **metod wzorcowych**. Bezwzorcowe metody porządkowania liniowego polegają na sortowaniu elementów bez wykorzystywania jakichkolwiek wcześniej ustalonych wzorców, bądź reguł. W przeciwieństwie do wzorcowych metod porządkowania liniowego, które uzależniają pozycję obiektów w rankingu od wcześniej ustalonych wzorców (a czasem również antywzorców).

---

<sup>1</sup>Dwa różne obiekty nie mogą uzyskać równej miary syntetycznej.

### 2.1.1 Metoda standaryzowanych sum

Metoda standaryzowanych sum jest przykładem bezwzorcowej metody porządkowania liniowego. Procedurę rozpoczynamy od przekształcenia wartości cech do postaci standaryzowanych wartości o charakterze stymulant. W sytuacji, gdy w zbiorze znajduje się kryterium o charakterze destymulanty należy ją sprowadzić do postaci stymulanty używając odpowiedniej formuły (Dziechciarz i Anna, 2003):

$$x_{ij} = \frac{1}{x_{ij}^d}. \quad (2.1)$$

Natomiast w przypadku, gdy zmienna ma charakter nominanty należy skorzystać ze wzoru:

$$x_{ij} = \begin{cases} 0 & \text{dla } x_{ij}^n \in \langle A, B \rangle, \\ -|A - x_{ij}^n| & \text{dla } x_{ij}^n < A, \\ -|B - x_{ij}^n| & \text{dla } x_{ij}^n > B, \end{cases} \quad (2.2)$$

gdzie

$x_{ij}^n$  – wartość zmiennej o charakterze nominanty przed transformacją,

$x_{ij}$  – wartość nominaty po transformacji w postać stymulanty,

$A$  – dolna granica przedziału, do którego należy oryginalna wartość nominanty,

$B$  – górna granica przedziału, do którego należy oryginalna wartość nominanty.

Procedurę metody standaryzowanych sum możemy opisać w 3 krokach (Dziechciarz, 1986):

1. Standaryzuj wartości zmiennych:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_{ij}}, \quad (2.3)$$

2. Oblicz sumę wartości zmiennych dla każdego obiektu używając wzoru<sup>2</sup> :

$$p_i = \sum_{j=1}^m z_{ij} \quad \text{dla } j = 1, 2, \dots, m, \quad (2.4)$$

gdzie  $m$  – liczba zmiennych.

---

<sup>2</sup>W metodzie standaryzowanych sum zakładamy, że każda zmienna ma jednakowy wpływ na zjawisko – wektor wag jest jednostkowy.

3. Oblicz miarę rozwoju dla każdego obiektu<sup>3</sup>:

$$m_i = \frac{p_i - p_{-0}}{p_0 - p_{-0}} \quad \text{dla } i = 1, 2, \dots, n, \quad (2.5)$$

gdzie  $p_0 = \sum_{j=1}^m z_{0j} \cdot w_j$ ,  $p_{-0} = \sum_{j=1}^m z_{-0j} \cdot w_j$ ,  
 $z_{0j}$  – wartości wzorca ( $z_{0j} = \max_i z_{ij}$ ),  
 $z_{-0j}$  – wartości antywzorca ( $z_{-0j} = \min_i z_{ij}$ ).

Ze względu na fakt, iż opisana metoda zakłada, że wszystkie zmienne są stymulantami to możemy stwierdzić, że im wartość  $p_i$  jest wyższa, tym  $i$ -ty obiekt będzie uważany za lepszy. Fakt ten pozwala uporządkować obiekty w rankingu na podstawie samej wartości sum  $p_i$ . Obliczenie opcjonalnej miary  $m_i$  pozwala nam uzyskać unormowane wartości na przedziale  $\langle 0, 1 \rangle$  i ułatwia interpretacje otrzymanych wyników. Im wartość  $m_i$  jest bliższa jedynce, tym lepiej to świadczy o  $i$ -tym obiekcie.

### 2.1.2 Metoda TOPSIS

Metoda TOPSIS (ang. The Technique for Order of Preference by Similarity to Ideal Solution) to jedna z metod wzorcowych wykorzystywanych w wielokryterialnej analizie decyzyjnej, której wynikiem jest ranking możliwych wyborów uszeregowanych według rosnącej wartości miary syntetycznej. Technika została opracowana przez Ching–Lai Hwang i Yoon w 1981 r., a następnie udoskonalona przez Yoon w 1987 r. oraz Hwang, Lai i Liu w 1993 r. Celem TOPSIS jest znalezienie rozwiązania będącego jak najbliżej rozwiązania idealnego, jednocześnie znajdującego się jak najdalej od rozwiązania antyidealnego (Hwang i Yoon, 1981).

Metoda wykorzystując funkcję agregującą cząstkowe informacje na temat poszczególnych kryteriów porównuje zestaw obiektów poprzez określenie wag dla każdego kryterium (cechy), normalizację wartości dla każdego kryterium (uzależniając sposób normalizacji od faktu czy preferujemy wysokie czy niskie wartości cechy) oraz obliczenie odległości geometrycznej (**miary syntetycznej**) między możliwym wyborem a rozwiązaniem idealnym, a następnie odległości od rozwiązania antyidealnego. Jako rozwiązanie idealne rozumiemy teoretyczny obiekt, który jest doskonałym spełnieniem każ-

---

<sup>3</sup>Krok 3 nie jest obowiązkowy, ale wskazany do wykonania.

dego z naszych kryteriów. Rozwiążanie antyidealne to przeciwnieństwo rozwiązania idealnego. Warto pamiętać przy wyborze kryteriów, aby ich wartości były monotoniczne.

Procedura postępowania zgodnie z metodą TOPSIS wygląda w następujący sposób (Hwang i Yoon, 1981):

1. Znormalizowanie danych każdego kryterium za pomocą wzoru:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}, \quad (2.6)$$

gdzie  $n$  – liczba obiektów w zestawie,

$k$  – liczba kryteriów.

2. Uwzględnienie przypisanych wag poszczególnym kryteriom:

$$v_{ij} = w_j z_{ij}. \quad (2.7)$$

3. Znalezienie rozwiązania idealnego oraz rozwiązania antyidealnego:

- wektor rozwiązania idealnego tworzymy przy pomocy wzoru:

$$z^+ = (z_1^+, z_2^+, \dots, z_k^+) = \begin{cases} \max\{z_{ij}\} & \text{dla kryterium stymulanty} \\ \min\{z_{ij}\} & \text{dla kryterium destymulanty} \end{cases}, \quad (2.8)$$

- wektor wartości rozwiązania antyidealnego tworzymy przy pomocy wzoru:

$$z^- = (z_1^-, z_2^-, \dots, z_k^-) = \begin{cases} \min\{z_{ij}\} & \text{dla kryterium stymulanty} \\ \max\{z_{ij}\} & \text{dla kryterium destymulanty} \end{cases}. \quad (2.9)$$

4. Obliczenie następujących odległości:

- odległości euklidesowych od rozwiązania idealnego:

$$d_i^+ = \sqrt{\sum_{j=1}^k (v_{ij} - z_j^+)^2}, \quad \text{dla } i = 1, 2, \dots, n \text{ oraz } j = 1, 2, \dots, k, \quad (2.10)$$

- odległości euklidesowych od rozwiązania antyidealnego:

$$d_i^- = \sqrt{\sum_{j=1}^k (v_{ij} - z_j^-)}, \text{ dla } i = 1, 2, \dots, n \text{ oraz } j = 1, 2, \dots, k. \quad (2.11)$$

5. Obliczenie miary syntetycznej określającej podobieństwo obiektów do rozwiązania idealnego i antyidealnego:

$$M_i = \frac{d_i^+}{d_i^+ + d_i^-} \quad \text{dla } i = 1, 2, \dots, n. \quad (2.12)$$

6. Uporządkowanie obiektów według rosnącej wartości  $M_i$ .

Im wyżej znajduje się obiekt w utworzonym rankingu tym jest on lepszy względem rozważanych kryteriów.

### 2.1.3 Metoda wzorcowa Hellwiga

Metoda wzorca rozwoju Hellwiga podobnie jak metoda TOPSIS należy do metod wzorcowych, ale w odróżnieniu od drugiej metody, metoda Helwiga bierze pod uwagę jedynie podobieństwo obiektu do rozwiązania idealnego.

Miarę rozwoju Hellwiga potrzebnego do uporządkowania obiektów możemy otrzymać w następujący sposób (Bąk, 2018):

1. Przeprowadź standaryzację zmiennych z wykorzystaniem następującego wzoru:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (2.13)$$

gdzie:

$x_{ij}$  – obserwacja  $j$ -tej zmiennej dla obiektu  $i$ ,

$\bar{x}_j$  – średnia arytmetyczna wartości  $j$ -tej zmiennej,

$s_j$  – odchylenie standardowe wartości  $j$ -tej zmiennej.

2. Stwórz wzorzec dla każdej zmiennej:

$$z_j^+ = \begin{cases} \max\{z_{ij}\} & \text{dla zmiennej stymulanty} \\ \min\{z_{ij}\} & \text{dla zmiennej destymulanty} \end{cases} \quad \text{dla } j = 1, 2, \dots, m, \quad (2.14)$$

gdzie  $m$  to liczba zmiennych.

3. Oblicz odległości obiektów od wzorca (rozwiązania idealnego):

$$d_i = \sqrt{\sum_{j=1}^m (z_{ij} - z_j^+)^2}. \quad (2.15)$$

4. Stwórz miarę syntetyczną z wykorzystaniem wzoru:

$$h_i = 1 - \frac{d_i}{d}, \quad \text{dla } i = 1, 2, \dots, n, \quad (2.16)$$

gdzie  $n$  to liczba obiektów,

$$d = \bar{d} + 2s_d, \quad \bar{d} = \frac{1}{n} \sum_i^n d_i, \quad s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \bar{d})^2}.$$

Obiekt jest uważany za lepszy jeżeli wartość jego miernika syntetycznego jest wyższa od miernika innego obiektu.

## 2.2 Metody analizy skupień

Analiza skupień to grupa metod analizy danych należących do kategorii algorytmów **uczenia nienadzorowanego** mająca na celu znalezienie takich grup, w których wartości cech obiektów są do siebie jak najbardziej podobne. Rozróżniamy dwa rodzaje metod grupowania: **metody hierarchiczne** i **metody niehierarchiczne**. Najważniejszą różnicą pomiędzy tymi dwoma technikami jest fakt, że hierarchiczne metody grupowania tworzą drzewiaste struktury, które przedstawiają kolejność grupowania obiektów, a niehierarchiczne metody jedynie dzielą zbiór danych na określoną z góry liczbę skupień.

Metody analizy skupień są często używane w sytuacjach, w których mamy do czynienia z ogromnymi zbiorami danych, które nie są oznaczone. Zbiory te charakteryzują się brakiem etykiet identyfikujących cechy lub klasyfikacje obiektów(Géron, 2022).

Algorytmy analizy skupień możemy wykorzystać m.in.:

- do wykrycia anomalii np. w transakcjach płatniczych kartą kredytową w celu zidentyfikowania kradzieży karty,
- do zapoznania się z podstawową strukturą zbioru danych np. analizując sprzedaż sklepu A możemy użyć grupowania do rozróżnienia różnych segmentów klientów i poznania ich nawyków zakupowych,
- do stworzenia systemu rekomendacji podobnego do oferowanych systemów przez m.in. Netflix i Amazon.

W niniejszej pracy zostaną wykorzystane dwie metody grupowania: **metoda  $k$ -średnich** oraz **metoda Warda**. Dodatkowo liczba wykrywanych skupień zostanie ustalona przy pomocy **metody łokcia**.

### 2.2.1 Metoda $k$ -średnich

Metoda  $k$ -średnich<sup>4</sup> to jedna z wielu niehierarchicznych metod analizy skupień, która pozwala na grupowanie zbioru danych na podstawie podobieństwa między obiekty. Podobieństwo szacujemy używając pojęcia wariancji, przy czym im mniejsza jej wartość tym obiekty należące do tego samego skupienia są bardziej do siebie podobne.

Główną ideą metody jest przypisanie każdego obiektu do najbliższego skupienia z  $k$  możliwych opcji, gdzie  $k$  oznacza liczbę skupień, które chcemy zidentyfikować. Odbywa się to w następujący sposób:

1. Rozpoczynamy od podjęcia decyzji o wartości  $k$ .
2. Przeprowadzamy standaryzację zmiennych<sup>5</sup> z wykorzystaniem następującego wzoru:
$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}. \quad (2.17)$$
3. Losujemy  $k$  obiektów, które stają się początkowymi skupieniami.
4. Obliczamy środki ciężkości istniejących skupień (Gambin, 2012):

<sup>4</sup>Metoda  $k$ -średnich jest również nazywana metodą centroidów.

<sup>5</sup>Standaryzacja jest potrzebna, ponieważ algorytm nie radzi sobie z rozrzedzonym zbiorem danych.

$$\hat{\mu}_j = \frac{\sum_{i=1}^n z_i}{n}, \quad (2.18)$$

gdzie:  $\hat{\mu}_j$  – środek ciężkości  $j$ -tego skupienia,

$n$  – liczba obiektów w  $j$ -tym skupieniu.

5. Obliczamy odległości między jednym z nieprzypisanych do żadnego skupienia obiektu i każdym ze środków ciężkości wykorzystując jedną z dostępnych metryk, a następnie przypisujemy rozważany obiekt do najbliższego skupienia. Przykładowe metryki<sup>6</sup>, z których możemy skorzystać to:

- odległość Euklidesowej:  $d_{ij} = \sqrt{\sum_{k=1}^n (i_k - \hat{\mu}_{jk})^2}$ ,
- odległość Czebyszewa:  $d_{ij} = \max_{k=1}^n |i_k - \hat{\mu}_{jk}|$ ,
- odległość miejska:  $d_{ij} = \sum_{k=1}^n |i_k - \hat{\mu}_{jk}|$ ,
- odległość Euklidesowa do kwadratu:  $d_{ij} = \sum_{k=1}^n (i_k - \hat{\mu}_{jk})^2$ .

Kroki 4 i 5 są powtarzane do momentu, aż wszystkie obiekty zostaną przypisane do pewnego skupienia (lub do moment osiągnięcia maksymalnej liczby iteracji algorytmu).

6. Obliczamy nowe środki ciężkości dla każdego z utworzonych skupień i ponownie wykonujemy kroki 5–6. Jeżeli obiekty zostały przypisane do tych samych skupień co poprzednio to kończymy ich grupowanie. W przeciwnym wypadku – jeżeli przynajmniej jeden obiekt został przeniesiony do innej grupy – wracamy do kroku 7.

Problemy, które możemy napotkać przy stosowaniu algorytmu centroidów to nieprawidłowy wybór liczby skupień  $k$  oraz błędne wylosowanie początkowych skupień, które decyduje o późniejszym grupowaniu obiektów (Géron, 2022). Rozwiązaniem pierwszego problemu jest tzw. **metoda łokcia** opisana w podrozdziale 2.2.2. Natomiast drugiego problemu możemy się pozbyć narzucając z góry początkowe skupienia albo uruchamiając algorytm kilkukrotnie, ale przy każdej iteracji losując inne początki skupień.

---

<sup>6</sup>Wymienione metryki są jedynie przykładowymi sposobami pomiaru odległości.

Wariancję skupienia możemy otrzymać aplikując poniższą formułę:

$$Var(C_k) = \sum_{x_i \in C_k}^k (x_i - \mu_k)^2 \quad \text{dla } j = 1, 2, \dots, k, \quad (2.19)$$

gdzie  $k$  – liczba skupień,

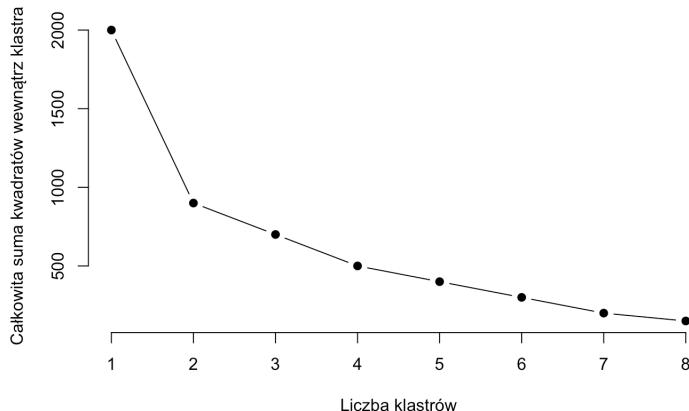
$x_i$  – obserwacja należąca do  $k$ -tego skupienia,

$\mu_k$  – środek ciężkości  $k$ -tego skupienia.

### 2.2.2 Metoda łokcia

Metoda łokcia to technika służąca do wyboru optymalnej liczby klas w analizie skupień, polegająca na obliczeniu sumy kwadratów odległości pomiędzy punktami wewnętrz skupienia a jego środkiem ciężkości dla różnej ilości skupień (tzw. **inercji**<sup>7</sup>). Wyniki prezentowane są na wykresie, gdzie na osi odciętych umieszczona jest liczba skupień, a na osi rzędnych wartości inercji (Géron, 2022).

$$WSS = \sum_{x_i \in C_k} \sum_{x_j \in C_k} ||x_i - x_j||^2 \quad (2.20)$$



Rysunek 2.1. Przykładowy wykres inercji zależnej od liczby skupień  $k$

Źródło: Opracowanie własne

Otrzymany wykres zazwyczaj przypomina kształtem łokieć, dlatego metodę nazywamy **metodą łokcia**. Na wykresie 2.1 widzimy, że wartość inercji maleje wraz ze wzro-

<sup>7</sup>Im mniejsza wartość inercji, tym otrzymane skupienia są bardziej jednorodne.

stem liczby skupień, a następnie powoli stabilizuje się po osiągnięciu tzw. **punktu łokcia na wykresie**<sup>8</sup>, który jest uważany za optymalną liczbę skupień.

Warto podkreślić, że najważniejszą informacją jest dla nas zmiana inercji a nie jej najmniejsza wartość, ponieważ przy każdym zwiększeniu liczby skupień wartość inercji odpowiadającej tej liczbie będzie malała. Dlatego punkt łokcia na wykresie charakteryzuje się najczęściej największym spadkiem całkowitej sumy kwadratów odległości.

### 2.2.3 Metoda Warda

Innym sposobem grupowania (inaczej metodą porządkowania nieliniowego) stosowanym w analizie skupień jest aplikowanie **algorytmu Warda**, którego celem jest budowa hierarchii zagnieżdżonych skupień według malejącego podobieństwa obiektów.

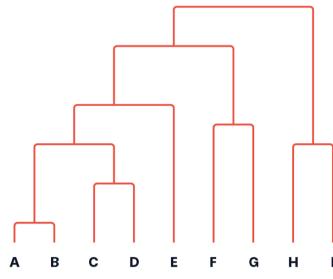
Algorytm rozpoczyna od przypisania obiektów do indywidualnych skupień, w których znajduje się tylko wybrany obiekt. Następnie w każdej iteracji algorytmu utworzone skupienia są ze sobą łączone, tworząc w ten sposób hierarchię skupień w górę. Taką strategię działania nazywamy **strategią aglomeracyjną**.

Decyzja o tym, które obiekty należy połączyć w jedno skupienie jest podejmowana na podstawie obliczonego **kryterium powiązań**. Metryka ta wykorzystuje analizę wariancji i dąży do minimalizacji sumy kwadratów różnic w nowo utworzonych skupieniach. Celem metody jest utworzenie grup, które charakteryzują się jak najmniejszym wewnętrznym zróżnicowaniem.

Po zakończeniu grupowania otrzymane skupienia możemy zaprezentować przy pomocy tzw. **dendrogramu** (rys.2.2) prezentującego siłę związków między obiektami na podstawie przyjętych kryteriów. Dodatkowo z utworzonego dendrogramu jesteśmy w stanie wskazać kolejność utworzenia skupień – im krótsza gałąź tym szybciej skupienie zostało utworzone (Sarkar, 2019).

---

<sup>8</sup>Optymalna liczba skupień może odpowiadać punktowi łokcia na wykresie, ale może również być od niego różna o 1.



Rysunek 2.2. Przykładowy dendrogram

Źródło: [datavizproject.com](http://datavizproject.com)

Po zapoznaniu się z głównymi ideami stojącymi za metodą Warda warto przejść przez bardziej szczegółowo zapisaną procedurą metody, którą możemy podzielić na następujące etapy (Podogrodzka i in., 2011):

1. Tworzymy macierz odległości  $D = [d_{ij}]$  między obiektami wykorzystując jedną z metryk wymienionych w procedurze metody k-srednich w punkcie 5.
2. Na podstawie macierzy odległości  $D$  szukamy pary obiektów ( $p$  i  $q$ ) charakteryzujących się najmniejszą odległością od siebie:

$$d_{pq} = \min_{i,j} \{d_{ij}\}, \quad (2.21)$$

gdzie  $n$  – liczba obiektów ( $i, j = 1, 2, \dots, n \wedge i \neq j$ ).

3. Łączymy obiekty  $p$  i  $q$  w jedno skupienie i określamy odległości pozostałych obiektów od nowo utworzonego skupienia  $p \cup q$  korzystając z następującego wzoru:

$$d_{(p \cup q)r} = \frac{n_p + n_r}{n_p + n_q + n_r} \cdot d_{pr} + \frac{n_q + n_r}{n_p + n_q + n_r} \cdot d_{qr} + \frac{n_r}{n_p + n_q + n_r} \cdot d_{pq}, \quad (2.22)$$

gdzie  $n_s$  – liczebność  $s$ -tego skupienia ( $r = 1, 2, \dots, n \wedge r \neq p \wedge r \neq q \Rightarrow s=\{p,q,r\}$ ).

4. Nowemu skupieniu  $p \cup q$  nadajemy nazwę  $p$ , po czym usuwamy obiekt  $q$ , a następnie zmniejszamy indeksy skupień większych od  $q$  o jeden i tworzymy nową macierz odległości  $D$  o wymiarze mniejszym o jeden. Kroki 2–4 powtarzamy do momentu, aż obiekty zostaną połączone w jedną klasę.

## Rozdział 3

# Atrakcyjności miejsca zamieszkania w przekroju powiatów województwa kujawsko-pomorskiego

### 3.1 Charakterystyka danych

Dane wykorzystane w analizie atrakcyjności zamieszkania powiatów zostały dostarczone przez [Bank Danych Lokalnych Głównego Urzędu Statystycznego](#) oraz serwis [wasaedukacja.pl](#). Zmienne, zostały dobrane w taki sposób, aby zbiór odzwierciedlał różne aspekty życia w powiatach (i miastach na prawach powiatów) województwa kujawsko-pomorskiego. Łącznie zbiór danych składa się z 19 zmiennych:

- **n01** – wskaźnik cen lokali (zł) – *destymulanta*,
- **n02** – wskaźnik powierzchni mieszkania (m<sup>2</sup>/mieszkańca) – *stymulanta*,
- **n03** – wskaźnik mieszkań (na 1 tys. mieszkańców) – *stymulanta*,
- **p01** – wskaźnik ofert pracy (na 1 tys. mieszkańców) – *stymulanta*,
- **p02** – wskaźnik bezrobocia (%) – *destymulanta*,
- **p03** – wskaźnik dochodu (zł) – *stymulanta*,
- **k01** – wskaźnik edukacji podstawowej (%) – *stymulanta*,

- **k02** – wskaźnik nauczania języków obcych w szkole podstawowej (%) – *stymulanta*,
- **k03** – wskaźnik nauczania języków obcych w szkole licealnej (%) – *stymulanta*,
- **k04** – wskaźnik opieki nad dziećmi (na 1 tys. mieszkańców) – *stymulanta*,
- **k05** – wskaźnik kultury (na 1 tys. mieszkańców) – *stymulanta*,
- **k06** – wskaźnik rozrywki (na 1 tys. mieszkańców) – *stymulanta*,
- **m01** – wskaźnik dostępności lekarstw (na 1 tys. mieszkańców) – *stymulanta*,
- **m02** – wskaźnik medycyny (na 1 tys. mieszkańców) – *stymulanta*,
- **s01** – wskaźnik zieleni (% udział w powierzchni powiatu) – *stymulanta*,
- **b01** – wskaźnik wykrywalności sprawców przestępstw (%) – *stymulanta*,
- **b02** – wskaźnik przestępstw (na 1 tys. mieszkańców) – *destymulanta*,
- **b03** – wskaźnik miejscowych zagrożeń (na 1 tys. mieszkańców) – *destymulanta*,
- **t01** – wskaźnik przystanków komunikacji miejskiej (na 1 tys. mieszkańców) – *stymulanta*.

Prefiksy zmiennych zostały zdefiniowane zgodnie z pogrupowaniem wskaźników zaprezentowanym w rozdziale 1.3:

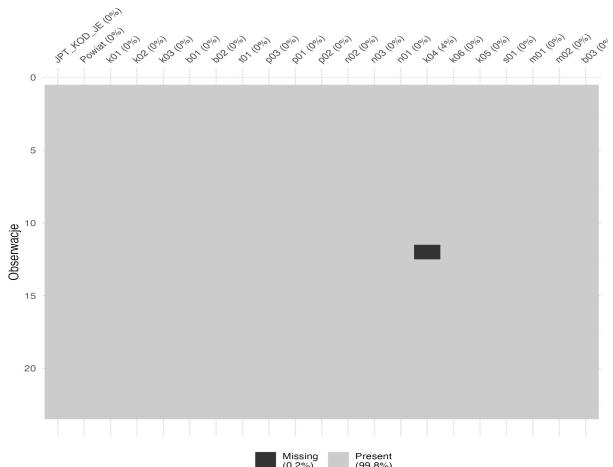
**Tabela 3.1. Prefiksy otrzymanych grup wskaźników**

Symbol	Grupa wskaźników
<b>n</b>	Sytuacja na rynku <b>nieruchomości</b>
<b>p</b>	Sytuacja na rynku <b>pracy</b>
<b>k</b>	Sytuacja społeczno–kulturowa
<b>m</b>	Opieka <b>medyczna</b>
<b>s</b>	Stan środowiska naturalnego
<b>b</b>	Poziom <b>bezpieczeństwa regionu</b>
<b>t</b>	Transport

**Źródło:** Opracowanie własne

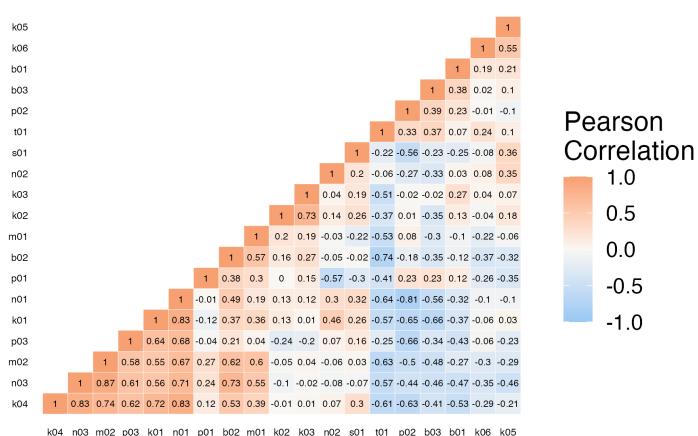
Ze względu na brak danych dla 2020 r. dotyczących wyników z egzaminu maturalnego dla powiatów województwa kujawsko-pomorskiego, do oceny jakości edukacji wykorzystujemy m.in. informacje o wynikach egzaminów ósmoklasisty.

W wykorzystywanym zbiorze danych pojawiła się jedna wartość pusta dla powiatu rypińskiego w kolumnie **k04** (rys. 3.1), dlatego przed przystąpieniem do obliczeń, pustą wartość zastępujemy średnią wartością wskaźnika opieki nad dziećmi, równą **102**. Po uzupełnieniu zbioru danych, staje się możliwa analiza siły zależności pomiędzy wybranymi zmiennymi uwzględnionymi w badaniu. Do zaprezentowania wyników wykorzystujemy macierz korelacji, której elementami są wartości współczynników korelacji Pearsona (rys. 3.2).



Rysunek 3.1. Braki danych w zbiorze danych

Źródło: Opracowanie własne



Rysunek 3.2. Korelacja Pearsona między zmiennymi

Źródło: Opracowanie własne

### 3.2 Rankingi porządkowania liniowego

W celu zbadania atrakcyjności zamieszkania powiatów (oraz miast na prawach powiatu) województwa kujawsko–pomorskiego zostały zaimplementowane trzy metody porządkowania liniowego: metoda sum standaryzowanych, metoda wzorca rozwoju Hellwiga oraz metoda TOPSIS. Każda z wymienionych metod została zaaplikowana dopiero po odpowiednim wyczyszczeniu i przygotowaniu zbioru danych.

Pierwszy ranking (tabela 3.2) został stworzony z wykorzystaniem bezwzorcowej metody sum standaryzowanych zgodnie z procedurą opisaną w rozdziale 2.1.1. Rozpoczynamy od transformacji zmiennych o charakterze destymulanty (zmiennej **n01**, zmiennej **p02**, zmiennej **b02** oraz zmiennej **b03**) do postaci zmiennych o charakterze stymulanty zgodnie ze wzorem 2.1. Następnie standaryzujemy wartości zmiennych (wzór 2.3) i obliczamy sumę ich wartości dla każdego rozpatrywanego obiektu. Procedurę kończymy wyznaczeniem miary  $m_i$ , według której najbardziej atrakcyjnym obszarem pod względem zamieszkania jest **Bydgoszcz** z wynikiem 0.564. Warto jednak zauważyć, że różnica między pierwszym a drugim miejscem (zajmowanym przez powiat bydgoski) jest znikoma, a powodem tego jest fakt, że Bydgoszcz leży wewnątrz powiatu bydgoskiego. Zwróćmy również uwagę na trzecie miejsce rankingu, które zajmuje druga stolica województwa – Toruń. Powiaty na niższych miejscach w rankingu znacznie odbiegają od trzech najlepszych obszarów, dlatego powinniśmy je traktować jako pewien "plan B".

**Tabela 3.2. Ranking powiatów otrzymany metodą sum standaryzowanych**

Nr	Powiat	Miara
1	Bydgoszcz	0.564
2	bydgoski	0.520
3	Toruń	0.517
4	radziejowski	0.453
5	żniński	0.438
6	brodnicki	0.422
7	Włocławek	0.414
8	świecki	0.405

**Tabela 3.2 – kontynuacja tabeli z poprzedniej strony**

Nr	Powiat	Miara
9	sępoleński	0.400
10	tucholski	0.400
11	aleksandrowski	0.384
12	toruński	0.381
13	Grudziądz	0.380
14	włocławski	0.379
15	golubsko-dobrzyński	0.367
16	wąbrzeski	0.361
17	nakielski	0.356
18	lipnowski	0.353
19	chełmiński	0.345
20	inowrocławski	0.344
21	rypiński	0.324
22	mogileński	0.316
23	grudziądzki	0.310

### Źródło: Opracowanie własne

Drugą metodą zastosowaną do oceny atrakcyjności zamieszkania jest metoda wzorca rozwoju Hellwiga, której efektem końcowym jest ranking przedstawiony w tabeli 3.3. Tabela 3.3 została posortowana według malejącej wartości miary rozwoju Hellwiga, porządkując zbiór powiatów od najlepszego do najgorszego obszaru. Proces obliczania miary rozwoju, również rozpoczyna się od standaryzacji danych zgodnie ze wzorem 2.13. Następnie zostaje utworzony tzw. **wzorzec**, który będzie symbolizował pewną teoretyczną miejscowości idealną pod każdym względem. Kolejnym krokiem jest obliczenie podobieństwa naszych obiektów do utworzonego wzorca używając wzoru 3.3. Na koniec pozostaje tylko obliczenie  $h_i$  (wzór 2.16).

Najwyższą miarę rozwoju otrzymał powiat **bydgoski**. Powodem tak wysokiego wyniku może być fakt, że powiat ten zawiera w swoim obszarze jedną ze stolic województwa kujawsko–pomorskiego (Bydgoszcz) lub fakt, że leży on niedaleko drugiej

stolicy województwa (Torunia). Jak widzimy w rankingu 3.3 stolice zajmują odpowiednio drugie i trzecie miejsce, co oznacza, że ich atrakcyjność przyciąga również obywateli do miejscowości leżących w ich okolicach, m.in. w powiecie bydgoskim. W przypadku metody Hellwiga, oceny atrakcyjności kolejnych obszarów rankingu nie różnią się znaczco od trzech najlepszych opcji – jak widzimy w rankingu różnica między maksymalną a minimalną wartością miary jest równa jedynie 0.3. Należy jednak zauważyć, że powiaty na miejscach niższych niż 11 otrzymały ujemną ocenę atrakcyjności, co może sugerować, że najlepiej pozostawić te rejony na koniec naszych poszukiwaniach nowego miejsca zamieszkania. Powodem nietypowych wartości ujemnych w ocenie jest brak znormalizowania wartości miary, a nie błąd w metodzie.

**Tabela 3.3. Ranking powiatów otrzymany metodą Hellwiga**

Nr	Powiat	Hellwiga
1	bydgoski	0.211
2	Bydgoszcz	0.183
3	Toruń	0.153
4	brodnicki	0.114
5	żniński	0.106
6	świecki	0.093
7	tucholski	0.046
8	radziejowski	0.034
9	Włocławek	0.032
10	sępoleński	0.017
11	toruński	0.010
12	Grudziądz	-0.00039
13	golubsko-dobrzyński	-0.005
14	inowrocławski	-0.009
15	aleksandrowski	-0.018
16	włocławski	-0.021
17	nakielski	-0.022
18	wąbrzeski	-0.033
19	chełmiński	-0.038

**Tabela 3.3 – kontynuacja tabeli z poprzedniej strony**

Nr	Powiat	Hellwiga
20	lipnowski	-0.044
21	mogileński	-0.047
22	rypiński	-0.068
23	grudziądzki	-0.104

**Źródło:** Opracowanie własne

Ostatnią tabelą prezentującą wyniki porządkowania liniowego jest tabela 3.4 zawierająca grupę miast i powiatów uporządkowanych według malejącej miary syntetycznej (wzór 2.12) otrzymanej metodą TOPSIS. Spadek miary świadczy o mniejszej atrakcyjności zamieszkania danego obszaru. Metoda TOPSIS<sup>1</sup> podobnie do wcześniej opisanych metod również zawiera pewną normalizację zbioru danych, utworzenie wzorca oraz miarę odległości. Nowym konceptem jest jednak tzw. **antywzorzec**, czyli najgorszy obszar pod względem atrakcyjności zamieszkania biorąc pod uwagę dowolne z rozważanych kryteriów. Nowy element metody nie wpłynął jednak znacząco na otrzymane wyniki. Podobnie jak w metodzie Hellwiga pierwsze miejsce zajmuje powiat bydgoski, a drugie jedna ze stolic województwa kujawsko–pomorskiego – tym razem jest to Toruń a nie Bydgoszcz, która znajduje się na miejscu trzecim. Kierując się wynikami metody TOPSIS powinniśmy podchodzić do obszarów gorszych od Bydgoszczy jak w przypadku kierowania się metodą sum standaryzowanych, czyli możemy je mieć na uwadze, ale nie powinniśmy rozpoczynać od nich naszych poszukiwań miejsca zamieszkania.

**Tabela 3.4. Ranking powiatów otrzymany metodą TOPSIS**

Nr	Powiat	TOPSIS
1	bydgoski	0.507
2	Toruń	0.494
3	Bydgoszcz	0.490
4	żniński	0.470

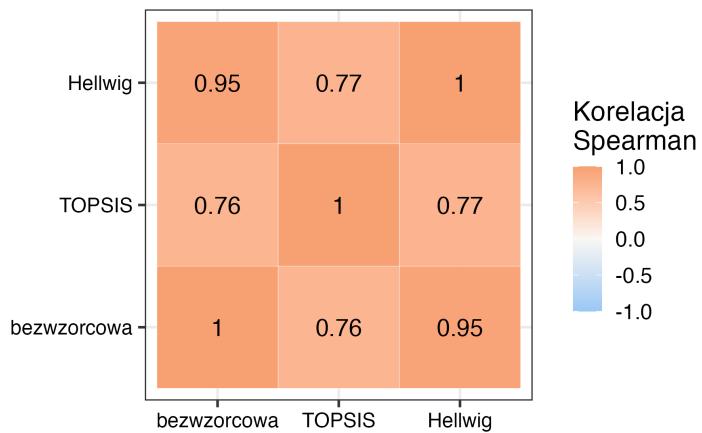
<sup>1</sup>Ze szczegółami metody można się zapoznać w rozdziale 2.1.2.

**Tabela 3.4 – kontynuacja tabeli z poprzedniej strony**

Nr	Powiat	TOPSIS
5	tucholski	0.445
6	sępoleński	0.437
7	Włocławek	0.381
8	Grudziądz	0.377
9	świecki	0.367
10	grudziądzki	0.363
11	radziejowski	0.359
12	brodnicki	0.350
13	wąbrzeski	0.342
14	toruński	0.338
15	chełmiński	0.334
16	nakielski	0.321
17	włocławski	0.321
18	lipnowski	0.316
19	inowrocławski	0.306
20	aleksandrowski	0.304
21	golubsko-dobrzyński	0.281
22	rypiński	0.281
23	mogileński	0.265

**Źródło:** Opracowanie własne

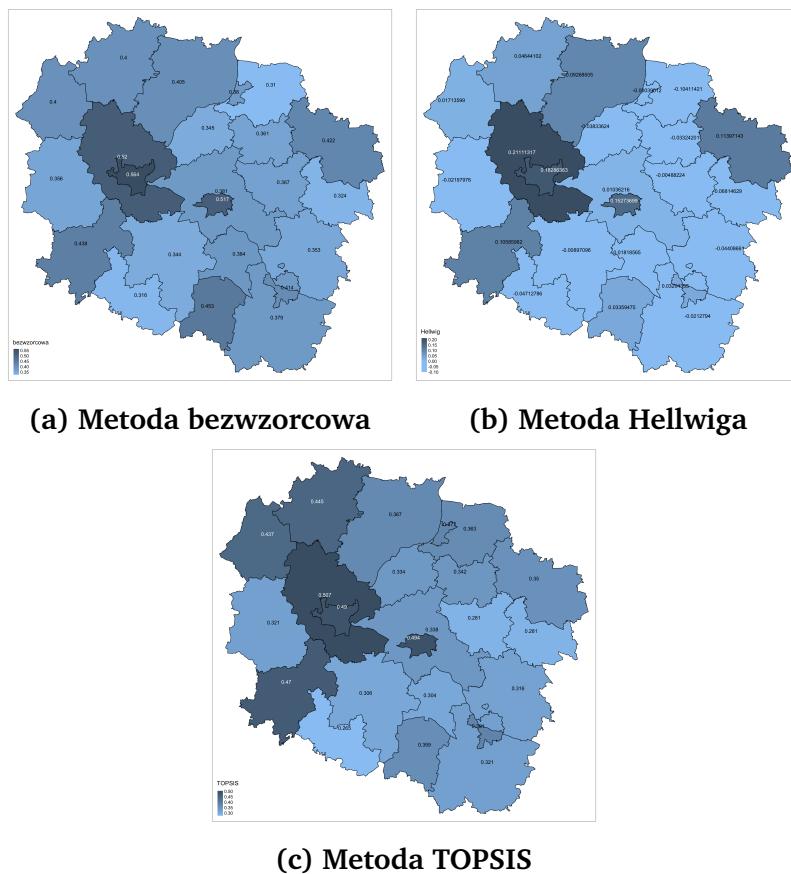
Uzyskane rankingi nie charakteryzują się istotnym zróżnicowaniem, co możemy udowodnić wykorzystując korelację Spearman'a. Na rys. 3.3) widzimy, że zbieżność wyników otrzymanych metodą Hellwiga oraz wyników otrzymanych metodą sum standaryzowanej jest silniejsza niż między wynikami uzyskanymi tymi metodami a metodą TOPSIS. Przyczyną może być m.in. uwzględnianie podobieństwa obszarów do najgorszej możliwości (antywzorca) w ostatniej metodzie.



Rysunek 3.3. Korelacja Spearman'a między pozycjami powiatów

Źródło: Opracowanie własne

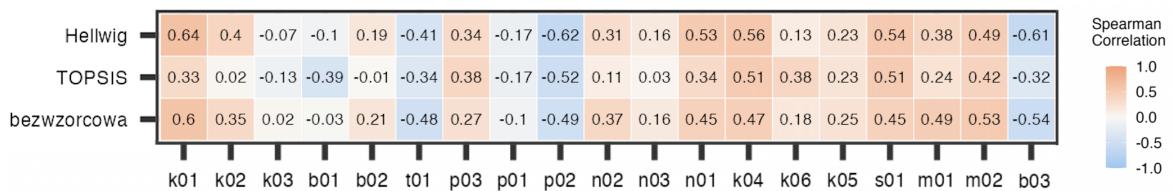
Na zestawie kartogramów (rys. 3.4) widzimy, że wcześniej wysunięte wnioski na temat zależności między atrakcyjnościami powiatów otrzymanymi różnymi metodami, mają odzwierciedlenie w rzeczywistości.



Rysunek 3.4. Kartogramy z wynikami metod porządkowania liniowego

Źródło: Opracowanie własne

Decyzja o miejscu zamieszkania na podstawie zaprezentowanych metod jest dość jednoznaczna i powinniśmy wybrać Toruń, Bydgoszcz lub okolice Bydgoszczy. Co jednak wyróżnia te trzy obszary na tle innych rozpatrując wymienione wskaźniki? We fragmencie macierzy korelacji Spearmana zilustrowanej na rys. 3.5 widzimy, że na miarę syntetyczną silnie wpływa m.in. zmienna **k01** (wskaźnik edukacji podstawowej), zmienna **p02** (wskaźnik bezrobocia), zmienna **n01** (wskaźnik cen lokalni), zmienna **k04** (wskaźnik opieki nad dziećmi) i zmienna **p03** (wskaźnik dochodu). Co ciekawe wzrost wartości zmiennej **n01** (o charakterze destymulanty) wpływa pozytywnie na pozycję obszaru w rankingu, co może być spowodowane charakterystyką popytu, które mówi im większa ilość osób wykazuje zainteresowanie produktem tym jego cena ulega wzrostowi. Również zmienna **b01** wykazuje ciekawe właściwości, ponieważ z interpretacji wynika, że jeżeli wykrywalność sprawców przestępstw wzrośnie to atrakcyjność zamieszkania spadnie, co przeczy logicznemu rozumowaniu i może świadczyć o błędny przypisaniu wskaźnikowi charakteru *stymulanty*.



Rysunek 3.5. Wpływ poszczególnych zmiennych na wyniki

Źródło: Opracowanie własne

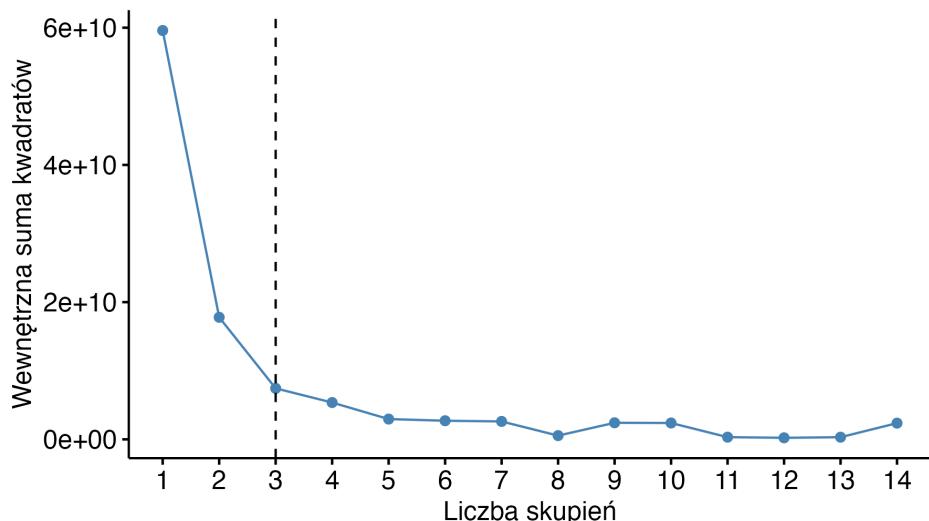
### 3.3 Grupowanie powiatów z wykorzystaniem analizy skupień

Do znalezienia grup podobnych powiatów przy pomocy metody k-średnich stosujemy istniejącą już w bibliotece *stats* metodę *kmeans(x, centers, nstart)*, gdzie *x* to zbiór danych składający się wyłącznie ze zmiennych o numerycznym typie danych, *centers*<sup>2</sup> to liczba skupień inicjujących grupowanie, a *nstart* określa liczbę sposobów wylosowania *k* klastrów, z których zostaje wybrany tylko jeden, który (według algorytmu) najlepiej pogrupuje obiekty.

<sup>2</sup>*centers* to odpowiednik symbolu *k*, którego używaliśmy do wyjaśnienia metody w rozdziale 2.2.2

Najlepszą strategią przy decydowaniu o wartości hiperparametru *centers* jest przetestowanie jakości grupowania powiatów dla różnych ilości klastrów np. poprzez obliczenie odległości obiektów należących do danego skupienia a środkiem jego ciężkości – metoda łokcia (rozdział 2.2.2). Po obliczeniu odległości, sumujemy je i traktujemy jako wskaźnik wydajności algorytmu, którego spadek świadczy o poprawie działania algorytmu.

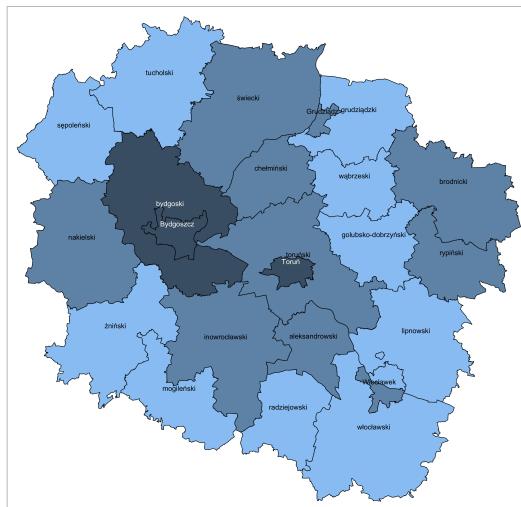
Patrząc na rys. 3.6 widzimy, że największą poprawę uzyskaliśmy w momencie zwiększenia liczby skupień do poziomu 2 klastrów. Jednak metoda łokcia pozwala na ustalenie optymalnej liczby skupień większej o 1 od punktu łokcia na wykresie, co wykorzystujemy w naszym badaniu i ustalamy *centers* na poziomie 3. W efekcie otrzymujemy podział województwa kujawsko–pomorskiego widoczny na rys. 3.7.



Rysunek 3.6. Optymalna liczba skupień (metoda łokcia)

Źródło: Opracowanie własne

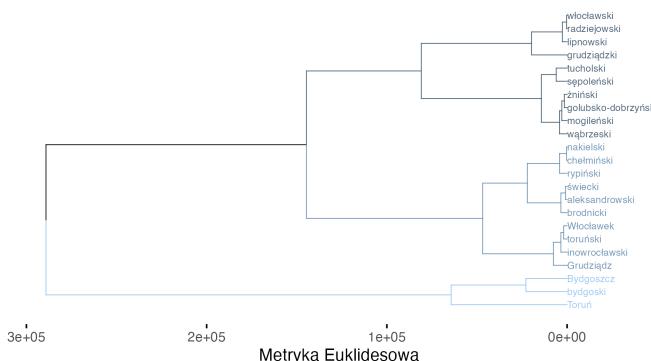
Można łatwo zauważyć, że do pierwszej grupy należą 3 obszary, które w poprzedniej części rozdziału zajęły pierwsze miejsca w rankingach porządkowania liniowego: Toruń, Bydgoszcz i powiat bydgoski. Do drugiej grupy należy 10 powiatów: włocławski, radziejowski, lipnowski, grudziądzki, tucholski, sępoleński, żniński, golubsko–dobrzyński, mogileński i wąbrzeski. Do trzeciego skupienia zostały przypisane pozostałe powiaty: nakielski, chełmiński, rypiński, świecki, aleksandrowski, brodnicki, Włocławek, toruński, inowrocławski i Grudziądz.



**Rysunek 3.7. Skupienia powiatów otrzymane metodą k-średnich**

**Źródło:** Opracowanie własne

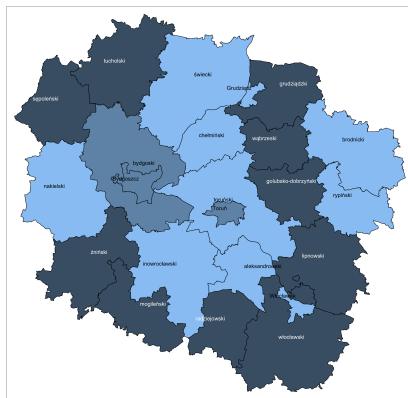
Drugą metodą użytą do grupowania powiatów według stopnia atrakcyjności zamieszkania jest hierarchiczna metoda Warda. Podstawową zasadą działania tej metody jest minimalizacja wariancji określającej odległość utworzonego skupienia od każdego innego, obliczonej przy pomocy normy Euklidesowej. Podobnie jak w przypadku metody k-średnich, metoda Warda posiada gotową funkcję w bibliotece *stats* noszącą nazwę *hclust(d,method)*, gdzie *d* oznacza macierz odległości utworzonej na podstawie zbioru danych. Macierz odległości została wygenerowana przy pomocy funkcji *dist()* również dostępnej w języku R. Natomiast w parametrze *method* definiujemy metodę aglomeracji, którą chcemy użyć do znalezienia klastrów (w naszym przypadku będzie to '*ward.D2*').



**Rysunek 3.8. Dendrogram połączeń metodą Warda**

**Źródło:** Opracowanie własne

Po zastosowaniu hierarchicznej metody analizy skupień otrzymaliśmy identyczne klastry, jak w przypadku metody  $k$ -średnich (rys. 3.9).



**Rysunek 3.9. Skupienia powiatów otrzymane metodą Warda**

**Źródło:** Opracowanie własne

Jednak dzięki metodzie Warda otrzymaliśmy bardziej szczegółowe informacje na temat przebiegu podziału obszarów na grupy. Na rys. 3.8 widać, że grupa składająca się z Torunia, Bydgoszczy i powiatu bydgoskiego została utworzona w końcowej fazie podziału oraz znaczaco odstaje od pozostałych jednostek województwa. Po wywołaniu funkcji *merge*, która generuje tabelę 3.5 widać, że grupa ta została utworzona dopiero w 19 kroku działania algorytmu, a Bydgoszcz z powiatem bydgoskim została połączona dopiero w kroku 17.

**Tabela 3.5. Wizualizacja procesu tworzenia dendrogramu w formie tabeli**

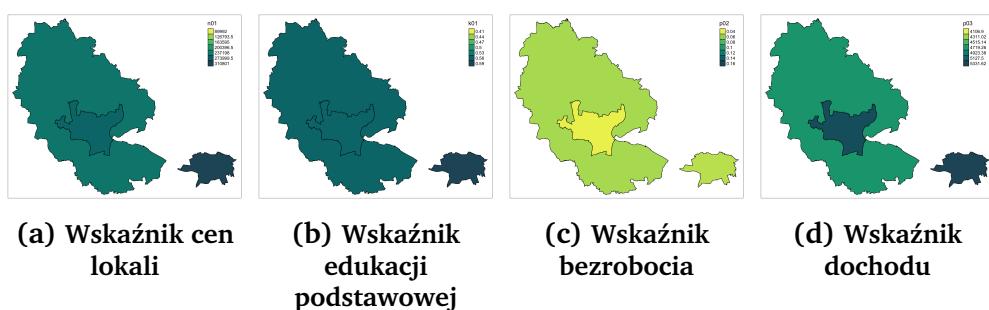
	Obiekt 1	Obiekt 2	Nr	Obiekt
1	-11	-18	1	aleksandrowski
2	-4	-10	2	brodnicki
3	-1	-14	3	bydgoski
4	-5	-19	4	chełmiński
5	-15	-23	5	golubsko-dobrzyński
6	-8	1	6	grudziądzki
7	-9	4	7	inowrocławski
8	-2	3	8	lipnowski
9	-7	5	9	mogileński
10	-12	2	10	nakielski
11	-17	7	11	radziejowski
12	-13	-16	12	rypiński
13	-21	9	13	sepoleński
14	11	12	14	świecki
15	-6	6	15	toruński
16	8	10	16	tucholski
17	-3	-20	17	wąbrzeski
18	13	16	18	włocławski
19	-22	17	19	żniński
20	14	15	20	Bydgoszcz
21	18	20	21	Grudziądz
22	19	21	22	Toruń
			23	Włocławek

**Źródło:** Opracowanie własne

### 3.4 Interpretacja wyników analizy taksonomicznej

Wyniki rankingów oraz otrzymanych skupień<sup>3</sup> nie powodują w nas zbyt dużego zdziwienia ze względu na charakterystykę obszarów w otrzymanych rankingach<sup>4</sup>.

Zacznijmy od głębszej analizy grupy trzech najatrakcyjniejszych obszarów w województwie kujawsko–pomorskim: Torunia, Bydgoszczy i powiatu bydgoskiego. Obie miejscowości i ich okolice posiadają dobrze rozwiniętą infrastrukturę, w tym sieć dróg, kolej, linii autobusowych i tramwajowych, co ułatwia przemieszczanie się po regionie. Niedaleko Torunia leży autostrada A1, która umożliwia szybki dojazd do każdego zakątku kraju, a nawet Niemiec. Grupę miejscowości charakteryzuje również wysoki poziom edukacji, dzięki takim placówkom jak Akademickie Liceum w Toruniu, Uniwersytet im. Mikołaja Kopernika czy Collegium Medicum w Bydgoszczy, dzięki czemu są to atrakcyjne miejscowości dla licealistów, studentów i naukowców. Ponadto, dzięki zainteresowaniu dużych korporacji tj. Zalando, które zdecydowało się na utworzenie oddziału w tym regionie<sup>5</sup>, zaobserwowane w ostatnim czasie wzrost zatrudnienia i poprawę sytuacji ekonomicznej na tym terenie. Niestety ceny lokali na terenie powiatów z 1. grupy są najwyższymi cenami panującymi w województwie. Nie zmienia to jednak faktu, że miasta Bydgoszcz i Toruń, oferując wiele korzyści dla mieszkańców, stanowią bardzo atrakcyjne miejsca do zamieszkania niezależnie od preferencji osoby.



Rysunek 3.10. Kartogramy powiatów w grupie 1

Źródło: Opracowanie własne

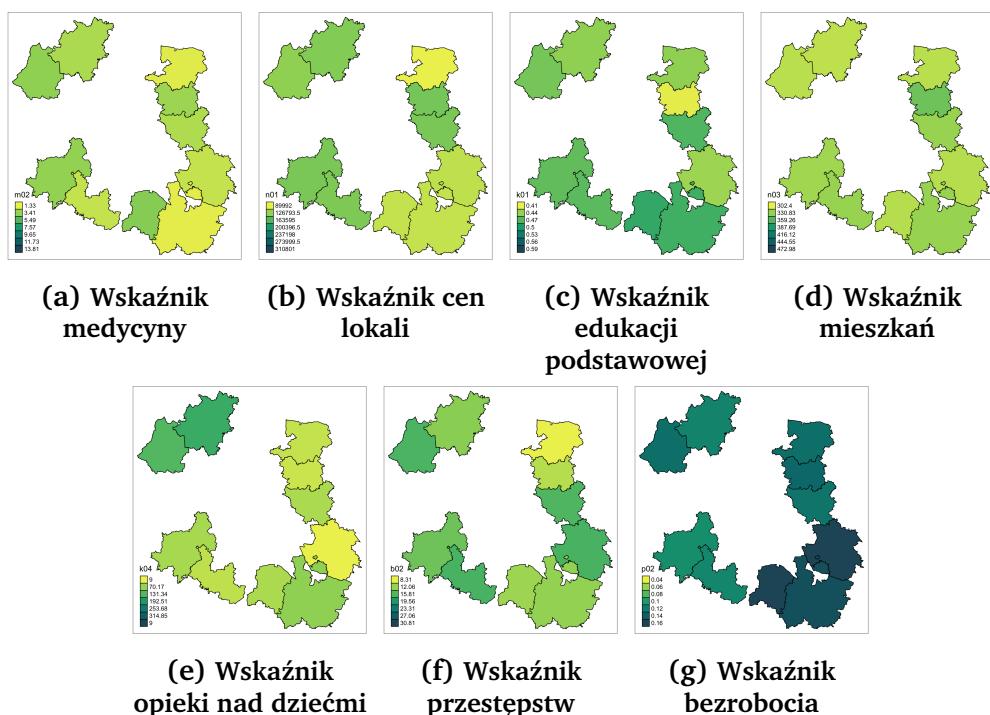
Do drugiej grupy 10 powiatów (włocławski, radziejowski, lipnowski, grudziądzki, tucholski, sępoleński, żniński, golubsko–dobrzyński, mogileński i wąbrzeski) przypisano

<sup>3</sup>Szczegółły skupień zostały przedstawione w poprzednim podrozdziale 3.3.

<sup>4</sup>Kartogramy prezentujące nasilenie poszczególnych wskaźników w otrzymanych skupieniach zostały umieszczone w repozytorium.

<sup>5</sup>[www.bydgoszcz.pl](http://www.bydgoszcz.pl)

obszary wyróżniające się bardzo niskimi wartościami wskaźnika medycyny, wskaźnika opieki nad dziećmi oraz wskaźnika mieszkani. Z drugiej strony ceny mieszkań na tym terenie są wyjątkowo niskie, a wskaźnik przestępstw i miejscowych zagrożeń nie odbiega zbytnio od średniej wartości dla całego województwa. Niestety, można tutaj zaobserwować bardzo wysoką stopę bezrobocia. Wszystkie te cechy utworzonej grupy mogą zniechęcić młodych (lub przyszłych) rodziców do wybrania obszaru jako swojego nowego miejsca zamieszkania. Jednak dla osób posiadających pracę zdalną może się to okazać najbardziej atrakcyjną lokalizacją po względem zamieszkania.

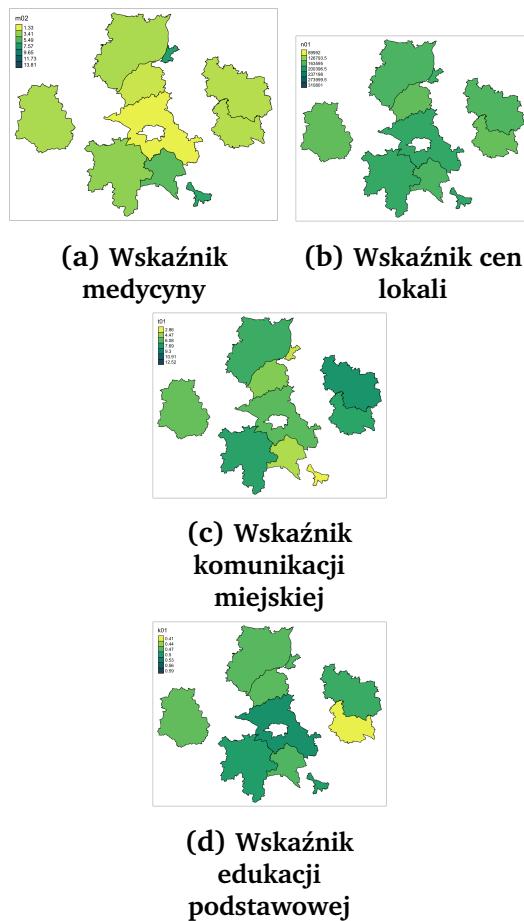


Rysunek 3.11. Kartogramy powiatów w grupie 2

**Źródło:** Opracowanie własne

Powiaty w trzecim skupieniu to głównie sąsiedzi powiatu bydgoskiego. Reguły tej nie spełniają jednak 2 z 8 powiatów: powiat brodnicki i rypiński oraz miejscowości Grudziądz i Włocławek. Nie są to zbyt atrakcyjne regiony, ponieważ mają one średnio rozwiniętą komunikacją miejską, a obywatele tych obszarów mogą narzekać m.in. na małą dostępność lekarzy i wydarzeń kulturalnych, słabą edukację w szkołach podstawowych (lub licealnych) oraz niski poziom zarobków w porównaniu do wartości panujących w całym województwie. Z drugiej strony, warto zauważyć, że ceny dostępnych mieszkań w powiatach należących do ostatniego skupienia nie są zbyt wysokie. Jest to pewne zadośćuczynienie za wymienione wcześniej problemy. Dlatego osoby, które nie

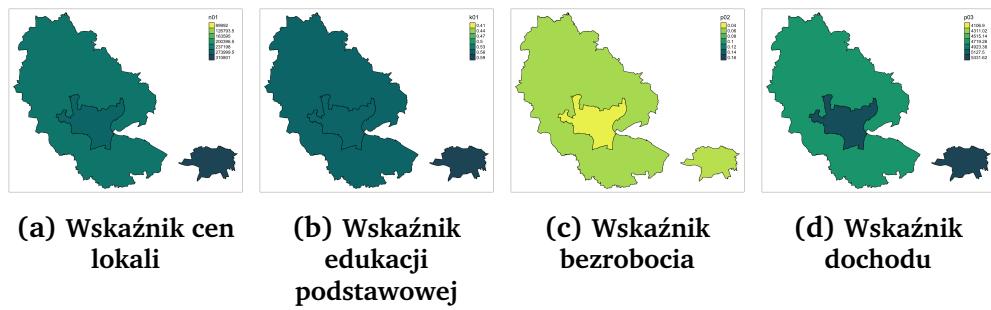
mają dużych wymagań co do infrastruktury, a cenią sobie spokojne i zaciszne miejsce zamieszkania, mogą rozpoczęć poszukiwanie ofert mieszkań od miejscowości należących do tego skupienia.



Rysunek 3.12. Kartogramy powiatów w grupie 3

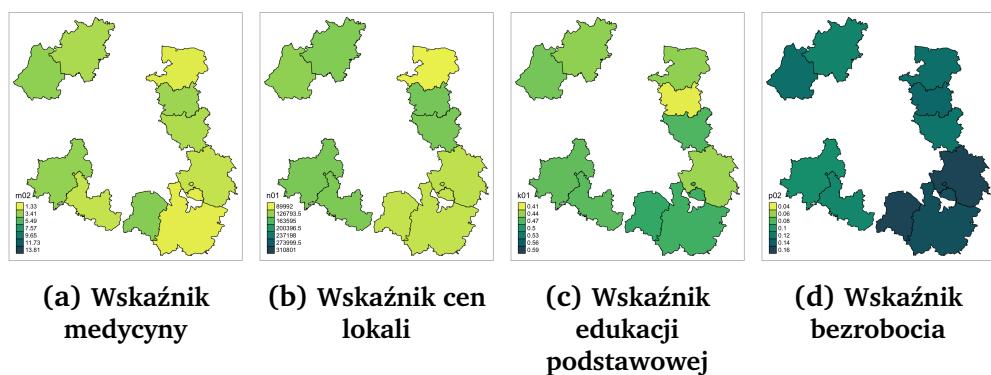
Źródło: Opracowanie własne

Na zakończenie chciałabym tylko podkreślić, że wybór miejsca zamieszkania jest bardzo indywidualny i zależy od wielu czynników, tj. preferencje osobiste, styl życia czy relacje rodzinne, które ciężko włączyć do analizy. Wyniki rankingów oraz analiza skupień mogą pomóc w dokonaniu świadomego wyboru, jednak należy pamiętać, że każdy człowiek ma inne oczekiwania i potrzeby. Dlatego warto dokładnie rozważyć swoje priorytety i kierować się nimi przy wyborze miejsca zamieszkania, aby móc się w pełni cieszyć z dokonanego wyboru.



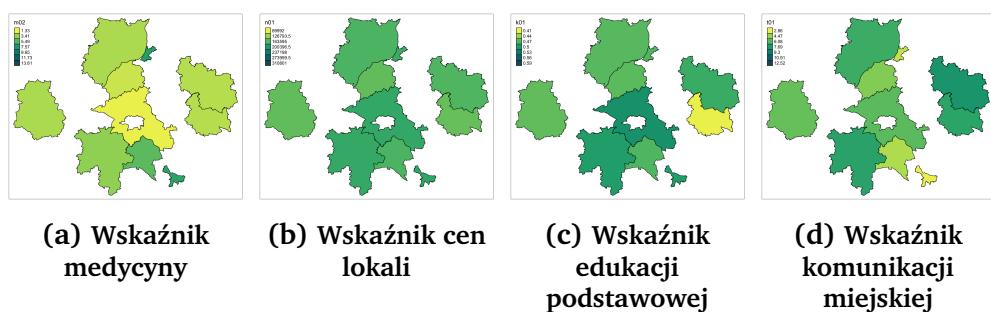
Rysunek 3.13. Kartogramy powiatów w grupie 1

Źródło: Opracowanie własne



Rysunek 3.14. Kartogramy powiatów w grupie 2

Źródło: Opracowanie własne



Rysunek 3.15. Kartogramy powiatów w grupie 3

Źródło: Opracowanie własne

# Podsumowanie

Celem niniejszej pracy było zbadanie i przeanalizowanie zróżnicowania atrakcyjności powiatów województwa kujawsko–pomorskiego w 2020 r. z punktu widzenia potencjału do zamieszkania. Użyto popularnych metod analizy wielowymiarowej, których procedury zostały szczegółowo omówione w rozdziale 2. Metody sum standaryzowanych, miernika rozwoju Hellwiga i TOPSIS pozwoliły na obliczenie miar syntetycznych odzwierciedlających potencjalną atrakcyjność zamieszkania powiatów.

Następnie zaaplikowano metodę k–średnich oraz metodę Warda w celu zidentyfikowania skupień podobnych jednostek administracyjnych pod względem atrakcyjności zamieszkania każdego z powiatów.

Analiza wykazała, że najatrakcyjniejszymi obszarami do życia są stolice województwa Toruń i Bydgoszcz oraz okolice Bydgoszczy (powiat bydgoski). Wspomniane obszary zajmują różne miejsca w rankingach w zależności od metody porządkowania liniowego, jednak zawsze znajdują się one w grupie trzech najlepszych powiatów. Gorzej radzą sobie regiony oddalone od najatrakcyjniejszych powiatów - największych miast na prawach powiatu. Na podstawie badania można wywnioskować, że zamieszkanie w centralnych częściach województwa będzie optymalną decyzją.

Przedstawione wyniki uwidaczniają znaczącą różnicę pomiędzy atrakcyjnością zamieszkania w obszarach stolic województwa, a pozostałymi powiatami. Istniejące dysproporcje związane z jakością życia na terenach różnych powiatów są możliwe do zwalczania pod warunkiem, że władze województwa podejmą odpowiednie działania oraz obywatele powiatów zaangażują się w podwyższenie badanej atrakcyjności. Jednak należy pamiętać, że nawet drastyczne rozwiązania będą miały wpływ na atrakcyjność dopiero na wiele lat po ich wdrożeniu.

# Bibliografia

- Województwo kujawsko-pomorskie [dostęp: 2023-02-17 12:44Z]. (2023). Dostępny w Internecie: [https://pl.wikipedia.org/wiki/Województwo\\_kujawsko-pomorskie](https://pl.wikipedia.org/wiki/Województwo_kujawsko-pomorskie).
- Bąk, A., i in. (2015). Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie pllord. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (384), 33–41.
- Bąk, A., i in. (2018). Analiza porównawcza wybranych metod porządkowania liniowego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (508), 19–28.
- Bąk, A. (2018). Zastosowanie metod wielowymiarowej analizy porównawczej do oceny stanu środowiska w województwie dolnośląskim. *Wiadomości Statystyczne. The Polish Statistician*, (1), 7–20.
- Bogacka, E., i Siniecka, A. (2016). Poczucie bezpieczeństwa mieszkańców miasta. Przykład Poznania. *Rozwój Regionalny i Polityka Regionalna*, (33), 57–71.
- Dziechciarz, J. (1986). *Ekonometria z elementami programowania matematycznego i analizy porównawczej*.
- Dziechciarz, J., i Anna, B. (2003). *Ekonometria. Metody Przykłady Zadania*.
- Gambin, A. (2012). Statystyczna analiza danych (molekularnych): mieszaniny rozkładów normalnych, algorytm k-średnich.
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc."
- Ginsbert-Gebert, A. (1984). Rozwój społeczno-gospodarczy a ochrona środowiska.
- Gniadek, A., i Marcisz, E. (2014). Zdrowie środowiskowe w miejscu zamieszkania: czynniki zagrożenia.
- Hwang, C., i Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag.

- Jędrzyński, J. (2020). Rynek pierwotny czy wtórny? <https://rynekpierwotny.pl>.
- Kozioł, L. (2013). The Concept of the Classification of Tourist Motivation Factors. *International Journal of Contemporary Management*, (12 (1)), 132.
- Krzton, W. (2017). Pojęcie i istota bezpieczeństwa jednostki. *Zeszyty Naukowe Wyższej Szkoły Informatyki, Zarządzania i Administracji w Warszawie*, 3(40), 42–54.
- Mortimer-Szymczak, H. (1995). Rynek pracy i bezrobocie (pojęcia ogólne).
- Nowosielska, E. (1995). Bezrobocie-przegląd aktualnych tendencji, problemów i wyników badań (w krajach rozwiniętych). *Zeszyty Instytutu Geografii i Przestrzennego Zagospodarowania PAN*.
- Olszewski, A., i Pilich, M. (2022). *Prawo oświatowe. Komentarz*. Wolters Kluwer.
- Podogrodzka, M., i in. (2011). Analiza zjawisk społeczno-ekonomicznych z zastosowaniem metod taksonomicznych. *Wiadomości Statystyczne. The Polish Statistician*, 56(11), 26–41.
- Sarkar, D. (2019). *Text analytics with Python: a practitioner's guide to natural language processing*. Springer.
- Tuleja, P., Florczak-Wątor, M., Naleziński, B., Radziewicz, P., i Czarny, P. (2019). *Konstytucja Rzeczypospolitej Polskiej. Komentarz*. Wolters Kluwer.
- Urząd Statystyczny. (2021). Rocznik Statystyczny Województwa Kujawsko-Pomorskiego 2021.

# **Spis tabel**

3.1	Prefiksy otrzymanych grup wskaźników . . . . .	28
3.2	Ranking powiatów otrzymany metodą sum standaryzowanych . . . . .	30
3.3	Ranking powiatów otrzymany metodą Hellwiga . . . . .	32
3.4	Ranking powiatów otrzymany metodą TOPSIS . . . . .	33
3.5	Wizualizacja procesu tworzenia dendrogramu w formie tabeli . . . . .	39

# Spis rysunków

1.1	Podział administracyjny województwa kujawsko-pomorskiego . . . . .	10
2.1	Przykładowy wykres inercji zależnej od liczby skupień $k$ . . . . .	24
2.2	Przykładowy dendrogram . . . . .	26
3.1	Braki danych w zbiorze danych . . . . .	29
3.2	Korelacja Pearsona między zmiennymi . . . . .	29
3.3	Korelacja Spearman'a między pozycjami powiatów . . . . .	35
3.4	Kartogramy z wynikami metod porządkowania liniowego . . . . .	35
3.5	Wpływ poszczególnych zmiennych na wyniki . . . . .	36
3.6	Optymalna liczba skupień (metoda łokcia) . . . . .	37
3.7	Skupienia powiatów otrzymane metodą k–średnich . . . . .	38
3.8	Dendrogram połączeń metodą Warda . . . . .	38
3.9	Skupienia powiatów otrzymane metodą Warda . . . . .	39
3.10	Kartogramy powiatów w grupie 1 . . . . .	40
3.11	Kartogramy powiatów w grupie 2 . . . . .	41
3.12	Kartogramy powiatów w grupie 3 . . . . .	42
3.13	Kartogramy powiatów w grupie 1 . . . . .	43
3.14	Kartogramy powiatów w grupie 2 . . . . .	43
3.15	Kartogramy powiatów w grupie 3 . . . . .	43

# Kody języka R

3.1	Przygotowanie danych . . . . .	50
3.2	Kartogram województwa kujawsko–pomorskiego . . . . .	51
3.3	Macierz korelacji zmiennych . . . . .	51
3.4	Metoda bezwzorcowa . . . . .	52
3.5	Metoda TOPSIS . . . . .	53
3.6	Metoda Hellwig . . . . .	53
3.7	Metoda kmeans z wykorzystaniem metody łokcia . . . . .	53
3.8	Metoda Warda . . . . .	54
3.9	Kartogramy z wynikami metod . . . . .	55
3.10	Macierz korelacji rankingów . . . . .	56
3.11	Kod generujący kartogramy w podrozdziale 3.4 . . . . .	57
3.12	Korelacja wartości miary syntetycznej i zmiennych . . . . .	57

```

library(dplyr)
2 library(ggplot2)
library(readxl)
4 library(stringr)
library(visdat)
6 library(sf)
library(cowplot)
8 dane<- read_xlsx("dane.xlsx", sheet = "dane") %>% rename(JPT_KOD_JE =
  Kod)

10 name<-names(which(colSums(is.na(dane))>0))
filter(dane, is.na(dane$`wskaźnik opieki nad dziećmi`))
12 row.names(dane)<-dane$Powiat

14 dane<-dane %>%
  rename('n01' = 'wskaźnik cen lokali') %>%
  rename('n02' = 'wskaźnik powierzchni mieszkania') %>%
  rename('n03' = 'wskaźnik mieszkań') %>%
  rename('p01' = 'wskaźnik ofert pracy') %>%
  rename('p02' = 'wskaźnik bezrobocia') %>%
  rename('p03' = 'wskaźnik dochodu') %>%
  rename('k01' = 'wskaźnik edukacji podstawowej') %>%
  rename('k02' = 'wskaźnik nauczania języków obcych w szkole
    podstawowej') %>%
  rename('k03' = 'wskaźnik nauczania języków obcych w szkole
    licealnej') %>%
  rename('k04' = 'wskaźnik opieki nad dziećmi') %>%
  rename('k05' = 'wskaźnik kultury') %>%
  rename('k06' = 'wskaźnik rozrywki') %>%
  rename('s01' = 'wskaźnik zieleni') %>%
  rename('m01' = 'wskaźnik dostępności lekarstw') %>%
  rename('m02' = 'wskaźnik medycyny') %>%
  rename('b01' = 'wskaźnik wykrywalności sprawców przestępstw') %>%
  rename('b02' = 'wskaźnik przestępstw') %>%
  rename('b03' = 'wskaźnik miejscowych zagrożeń') %>%
  rename('t01' = 'wskaźnik przystanków komunikacji miejskiej')

34 dane['k02']<-dane['k02']/100
36 dane['k03']<-dane['k03']/100
38 dane['p02']<-dane['p02']/100
38 theme_set(theme_bw(base_size=16))

40 missing_data<-vis_miss(dane)+ylab('Obserwacje')
ggsave(filename='obrazy/missing.jpg', plot = missing_data)
42
#zmieniamy nazwe Kod na JPT_KOD_JE
44 dane['k04'] <- replace(dane['k04'], is.na(dane['k04']), mean(dane$k04
  , na.rm=TRUE))

46 dane$Powiat<-str_replace(dane$Powiat, "Powiat ", "")

48 x_ij <- dane %>% select(3:21) %>%as.matrix()
row.names(x_ij) <-dane%>%select(2)%>%as.vector()%>%array()%>%unlist()
50
powiaty<-read_sf('powiaty/powiaty.shp')
52 kuj_pom <- powiaty[substr(powiaty$JPT_KOD_JE, 0, 2) == '04', ]
kuj_pom$JPT_NAZWA_ <-str_replace(kuj_pom$JPT_NAZWA_, "powiat ", "")
54 kuj_pom_miasta <- kuj_pom[substr(kuj_pom$JPT_KOD_JE, 0, 3) == '046',
  ]

```

```
kuj_pom_pow <- kuj_pom[(substr(kuj_pom$JPT_KOD_JE, 0, 3) != '046'), ]
```

**Listing 3.1. Przygotowanie danych**

```
mapa<-ggplot(kuj_pom)+  
  geom_sf(data = kuj_pom, color = 'white', fill='#98C8F4', size=15) +  
  geom_sf(data = kuj_pom_miasta, color = 'white', fill='#7094B5',  
    size=15) +  
  geom_sf_text(  
    data = kuj_pom,  
    aes(label = JPT_NAZWA_),  
    #family = "Calibri",  
    size = 3,  
    color = 'black')+  
  theme_classic()  
  theme(axis.line.x.bottom = element_blank(),  
        axis.line.y.left = element_blank(),  
        axis.ticks = element_blank(),  
        axis.title.x = element_blank(),  
        axis.title.y = element_blank(),  
        axis.text.x = element_blank(),  
        axis.text.y = element_blank())  
  
ggsave(  
  filename='obrazy/mapa.png',  
  plot = mapa,  
  scale = 1,  
  width = 8,  
  height = 8)
```

**Listing 3.2. Kartogram województwa kujawsko-pomorskiego**

```
library(reshape2)  
library(ggthemes)  
library(extrafont)  
library(ggplot2)  
library(rmetrics)  
library(cowplot)  
  
get_upper_tri <- function(data){  
  data[lower.tri(data)]<- NA  
  return(data)  
}  
  
reorder_data <- function(data){  
  dd <- as.dist((1-data)/2)  
  hc <- hclust(dd)  
  data <- data[hc$order, hc$order]  
}  
  
data<-round(cor(x_ij),2)  
data <- reorder_data(data)  
upper_tri <- get_upper_tri(data)  
melted_data <- melt(upper_tri, na.rm = TRUE)  
  
ggheatmap <- ggplot(melted_data, aes(Var2, Var1, fill = value))+  
  geom_tile(color = "white")  
  scale_fill_gradient2(low = "#b5e2fa", high = "#f7a072", mid = "#  
    f9f7f3",  
    limit = c(-1,1),
```

```

28   name="Pearson\nCorrelation") +
29   geom_text(aes(Var2, Var1, label = value), color = "black", size =
30             2) +
31   theme(
32     axis.title.x = element_blank(),
33     axis.title.y = element_blank(),
34     panel.grid.major = element_blank(),
35     panel.border = element_blank(),
36     panel.background = element_blank(),
37     axis.ticks = element_blank(),
38     axis.text.x = element_text(colour = "black", size = 6),
39     axis.text.y = element_text(colour = "black", size = 6)
40   )
41
42 save_plot(plot=ggheatmap, filename='obrazy/corr.png')

```

**Listing 3.3.** Macierz korelacji zmiennych

```

library(dplyr)
1 stymulanta <- c('+', '+', '+', '+', '-',
2   '+', '+', '+', '+', '+', '+', '+', '+', '+',
3   '+', '+', '+', '+', '+', '+', '+', '+', '+')
4 colnames(x_ij)

5 x<-x_ij

6 for(j in 1:19) {
7   if(stymulanta[j]=='-') {
8     x[,j]<-(1/x_ij[,j])
9     print(x[,j])
10   }
11 }
12
13 z_ij<-scale(x, center = TRUE, scale = TRUE)
14
15 p_i <- matrix(0,nrow(z_ij),1, dimnames=list(row.names(z_ij)))
16 for(i in 1:nrow(z_ij)) {
17   p_i[i] <- sum(z_ij[i,])
18 }
19
20 z_oj<-matrix(0,1,ncol(z_ij))
21 colnames(z_oj)<-colnames(z_ij)
22 for(j in 1:ncol(z_ij)) {
23   z_oj[j] <- max(z_ij[,j])
24 }
25
26 z_moj<-matrix(0,1,ncol(z_ij))
27 colnames(z_moj)<-colnames(z_ij)
28 for(j in 1:ncol(z_ij)) {
29   z_moj[j] <- min(z_ij[,j])
30 }
31
32 p_o<-sum(z_oj)
33 p_mo<-sum(z_moj)
34
35 m_i<- ((p_i-p_mo)/(p_o-p_mo))%>%round(3)%>%as.data.frame(row.names=
36   row.names(p_i))
37 colnames(m_i)<-'bezwzorcowa'
38

```

**Listing 3.4.** Metoda bezwzorcowa

```

1 if(!require('MCDA')) {
2   install.packages('MCDA')
3   library('MCDA')
4 }
5 weights <- rep(1,19)
6 criteria <- c('max','max','max','max','min','max','max','max','max','
7   'max','max','min','max','max','max','max','max','max','max','min')
8 M<-TOPSIS(x_ij, weights, criteria)%>%round(3) %>% as.data.frame(row.
9   names=names(x_ij))
10 colnames(M)<-'TOPSIS'
11 ranking<-merge(m_i, M, by='row.names')
12 row.names(ranking)<-ranking$Row.names
13 ranking<-ranking%>%select(-1)

```

**Listing 3.5. Metoda TOPSIS**

```

1 w<-rep(1,19) #wagi
2 scaled<-scale(x_ij)*w
3 criteria <- c('max','max','max','max','min','max','max','max','max','
4   'max','max','min','max','max','max','max','max','max','max','min')
5 wzorzec<-scaled[1,] #wzorzec
6 for(j in 1:ncol(scaled)) {
7   if(criteria[j]=='max') {
8     wzorzec[j]<- max(scaled[,j])
9   }else if(criteria[j]=='min') {
10     wzorzec[j]<- min(scaled[,j])
11   }
12 }
13
14 dpl <- scaled[,1]
15 for(i in 1:nrow(scaled)) {
16   dpl[i]<-sqrt(sum((scaled[i,]-wzorzec)^2))
17 }
18
19 dpl_matrix<-dpl%>%as.matrix()
20
21 d_bar<-sum(dpl)/nrow(dpl_matrix)
22
23 s_d<-z_ij[,1]
24 for(i in 1:nrow(z_ij)){
25   s_d[i] <- sqrt(sum((dpl[i]-d_bar)^2)/nrow(dpl_matrix))
26 }
27 s_d<-s_d%>%as.matrix()
28
29 d<-d_bar+2*s_d
30
31 h_i <- (1-(dpl/d)) %>%round(8) %>% as.data.frame(row.names=names(dpl))
32 colnames(h_i)<-'Hellwig'
33
34 ranking<-merge(ranking, h_i, by='row.names')
35 row.names(ranking)<-ranking$Row.names
36 ranking<-ranking%>%select(-1)

```

**Listing 3.6. Metoda Hellwig**

```

1 library(factoextra)
2 library(purrr)

```

```

4 #Przykładowy wykres metody łokcia (rozdział 2)
5 plot(c(1:8), c(2000,900,700,500,400,300,200,150),
6       type="b", pch = 19, frame = FALSE,
7       xlab="Liczba klastrów",
8       ylab="Całkowita suma kwadratów wewnątrz klastra"
9     )
10
11 elbow<-c()
12 difference<-c()
13
14 k_centers<-c(1:19)
15
16 for(i in k_centers){
17   elbow[i]<- kmeans(x_ij, centers = i, nstart = 25)$tot.withinss
18   if(i > 1){
19     difference[i]<-elbow[i]-elbow[i-1]
20   }
21 }
22 min(difference, na.rm=TRUE)
23
24 library(cluster)
25 library(factoextra)
26
27
28 silhouette<-fviz_nbclust(x_ij, kmeans, k.max = 14,method =
29   "silhouette")+
30   labs(title= "Optymalna liczba klastrów", subtitle = 'Indeks
31   silhouette')+
32   xlab("Liczba skupień") +
33   ylab("Silhouette")
34 save_plot(plot=silhouette, filename='obrazy/silhouette.png')
35
36 elbow<-fviz_nbclust(x_ij, kmeans, k.max = 14,method = "wss")+
37   geom_vline(xintercept = 3, linetype = 2)+
38   labs(title= "")+
39   xlab("Liczba skupień") +
40   ylab("Wewnętrzna suma kwadratów")
41 save_plot(plot=elbow, filename='obrazy/elbow.png')
42
43 kmeans_final<-kmeans(x_ij, centers = 3, nstart = 25)
44 clusters<-kmeans_final$cluster%>%as.data.frame(row.names=names(kmeans
45   _final$cluster))
46 colnames(clusters)<- 'kmeans'
47
48 ranking<-merge(ranking, clusters, by='row.names')
49 row.names(ranking)<-ranking$Row.names
50 ranking<-ranking%>%select(-1)

```

**Listing 3.7. Metoda kmeans z wykorzystaniem metody łokcia**

```

1 library(relations)
2 library(dendextend)
3 library(set)
4 library(factoextra)
5 library(dplyr)
6 #Ward
7
8 ward<-cutree(hclust(dist(as.matrix(x_ij)), method = "euclidean"),
9   method='ward.D2'), k=3)%>%as.data.frame()

```

```

1 colnames(ward)<- 'ward'
10 ranking<-merge(ranking, ward, by='row.names')
row.names(ranking)<-ranking$Row.names
12 ranking<-ranking%>%select(-1)

14 hc.ward<-hclust(dist(as.matrix(x_ij)), method = "euclidean"), method=
  'ward.D2')
wardMerge<-as.data.frame(hc.ward$merge)%>%
  rename('Obiekt 1'=V1)%>%rename('Obiekt 2'=V2)
wardMerge
18 plot(hc.ward, xlab="Powiaty", ylab = "Metryka Euklidesowa", main =
  ',lwd=1)
par(lwd=2, mar=c(0,0,0,0))
20 rect<-rect.hclust(hc.ward, k=3, border = 'steelblue')

22 dend<-fviz_dend(hc.ward, cex = 0.45, k=3,
  rect = TRUE,
24   k_colors = c("#98C8F4", "steelblue", "#496075"),
  rect_border = "black",
26   rect_fill = FALSE,
  horiz = TRUE,
28   lwd =0.25,
  rect_lty = 0, ylab="Metryka Euklidesowa",
  main= "")
30 save_plot(plot=dend, filename='obrazy/ward_drzewo.png')

```

**Listing 3.8. Metoda Warda**

```

library(tmap)
2 library(tmaptools)
#install.packages('openxlsx')
4 library(openxlsx)
height_value = 10
6 color_palette <-c('#98C8F4', '#7094B5', '#496075')

8 ranking$JPT_NAZWA_<-rownames(ranking)
kuj_pom_rank <- left_join(kuj_pom, ranking, by = "JPT_NAZWA_")
10 kuj_pom_rank$Hellwig<- array(kuj_pom_rank$Hellwig)
kuj_pom_rank$bezwzorcowa<- array(kuj_pom_rank$bezwzorcowa)

12 bezwzorcowa<-tm_shape(kuj_pom_rank) +
  tm_polygons(col = "bezwzorcowa",
14   style = "cont",
    palette =color_palette,
16   legend.reverse = TRUE,
18   border.col = 'black',
   border.lwd = 0.6
20   )+
  tm_text(text="bezwzorcowa",
22   size = 0.8,
     auto.placement=TRUE,
24   ymod=0.6)
tmap_save(bezwzorcowa, 'obrazy/bezwzorcowa.png', height= height_value
  )

26 Hellwig<-tm_shape(kuj_pom_rank) +
  tm_polygons(col = "Hellwig",
28   style = "cont",
    palette =color_palette,
30   legend.reverse = TRUE,

```

```

32         border.col = 'black',
33         border.lwd = 0.6
34     )+
35     tm_text(text="Hellwig",
36             size = 0.8,
37             auto.placement=TRUE,
38             ymod=0.8)
39 tmap_save(Hellwig, 'obrazy/Hellwig.png', height= height_value)
40
41 TOPSIS<-tm_shape(kuj_pom_rank) +
42     tm_polygons(col = "TOPSIS",
43                 style = "cont",
44                 palette =color_palette,
45                 legend.reverse = TRUE,
46                 border.col = 'black',
47                 border.lwd = 0.6
48             )+
49     tm_text(text="TOPSIS",
50             size = 0.8,
51             auto.placement=TRUE,
52             ymod=0.87)
53 tmap_save(TOPSIS, 'obrazy/TOPSIS.png', height= height_value)
54
55 Ward<-tm_shape(kuj_pom_rank) +
56     tm_polygons(col='ward',border.col='black',palette = c('#98C8F4',
57     '#7094B5', '#496075'))+
58     tm_layout(legend.outside=TRUE)+ 
59     tm_legend(show=FALSE)+ 
60     tm_text(text='JPT_NAZWA_', size = 0.8,
61             auto.placement=TRUE,
62             ymod=0.9)
63 tmap_save(Ward, 'obrazy/Ward.png', height= height_value)
64
65 kmean<-tm_shape(kuj_pom_rank) +
66     tm_polygons(col='kmeans',border.col='black',palette = c('#7094B5',
67     '#98C8F4', '#496075'))+
68     tm_legend(show=FALSE)+ 
69     tm_text(text='JPT_NAZWA_',
70             size = 0.8,
71             ymod=0.9)
72 tmap_save(kmean, 'obrazy/kmeans.png', height= height_value)

```

**Listing 3.9. Kartogramy z wynikami metod**

```

1 library(reshape2)
2 library(ggthemes)
3 library(extrafont)
4 library(ggplot2)
5 library(remotes)
6 library(cowplot)
7 library(dplyr)
8
9 dataRanking<-round(cor(ranking%>%select(1:3), method='spearman'),2)
10 melted_dataRanking <- melt(dataRanking, na.rm = TRUE)
11
12 rankingHeatmap <- ggplot(melted_dataRanking, aes(Var2, Var1, fill =
13     value))+ 
14     geom_tile(color = "white")+
15     scale_fill_gradient2(low = "#b5e2fa", high = "#f7a072", mid = "#f9f7f3",
16     )

```

```

    limit = c(-1,1),
16   name="Korelacja\nSpearman") +
geom_text(aes(Var2, Var1, label = value), color = "black", size =
5) +
18 theme(
  axis.title.x = element_blank(),
20  axis.text.x = element_text(size = 12, colour = "black"),
21  axis.text.y = element_text(size = 12, colour = "black"),
22  axis.title.y = element_blank())
save_plot(plot=rankingHeatmap, filename='obrazy/corrRanking.png')

```

**Listing 3.10.** Macierz korelacji rankingów

```

library(grDevices)
2 library(tmap)
3 library(tmaptools)
#install.packages('openxlsx')
4 library(openxlsx)
height_value = 5
5 dane1<-dane%>%rename(JPT_NAZWA_=Powiat)

8 dane_ranking<-left_join(ranking, dane1, by = "JPT_NAZWA_")
9 kuj_pom_dane <- left_join(kuj_pom, dane_ranking, by = "JPT_NAZWA_")

12 for(cl in c(1:3)){
  for(num_col in c(44:62)){
    col_name<-colnames(kuj_pom_dane)%>%filter(kuj_pom_dane$ward==cl)
14 %>%select(num_col))[1]
    minimum<-min(dane[,col_name])
16    maximum<-max(dane[,col_name])
    step<-round(((maximum-minimum)/6), digits=2)

18   picture<-tm_shape(kuj_pom_dane)%>%filter(kuj_pom_dane$ward==cl)) +
20     tm_polygons(col = col_name,
21                  style = "cont",
22                  palette = rev(hcl.colors(7, "ag_GrnY1"))

24       ,
26         border.col = 'black',
27         border.lwd = 0.05,
28         breaks = seq(minimum, maximum, step),
29         legend.show = FALSE
30       )+
31         tm_add_legend(type = "fill",
32                       size = 0.1,
33                       col = rev(hcl.colors(7, "ag_GrnY1")),
34                       labels = round(seq(minimum, maximum,
35                         step), 2),
36                         title=col_name)
      file_name<-paste("charakterystyka",cl,"/",col_name,".png",sep="")
      tmap_save(picture, filename = file_name, height= height_value)
    }
  }

```

**Listing 3.11.** Kod generujący kartogramy w podrozdziale 3.4

```

library(reshape2)
2 rank<-ranking%>%rename( 'Powiat'='JPT_NAZWA_')
linearOrdearAnalysis <- left_join(dane, rank%>%select(1,2,3,6), by =
  "Powiat")
4 melted_linearOrd<-melt(round(cor(linearOrdearAnalysis%>%select(3:24),

```

```

  method = 'spearman'),2))

corrLinear<-ggplot(
  data =melted_linearOrd , aes(x=Var1, y=Var2, fill=value)) +
geom_tile(color = "white")+
scale_fill_gradient2(low = "#98C8F4", high = "#f7a072", mid = "#f9f7f3",
                     midpoint = 0, limit = c(-1,1),
                     space = "Lab",
                     name="Spearman\nCorrelation")+
geom_text(aes(Var2, Var1, label = value), color = "black", size =
  1.5)+
theme(
  axis.title.x = element_blank(),
  axis.text.x = element_text(size = 5, colour = "black"),
  axis.text.y = element_text(size = 5, colour = "black"),
  axis.title.y = element_blank())
save_plot(plot=corrLinear, filename='obrazy/zmienne.png')

```

**Listing 3.12. Korelacja wartości miary syntetycznej i zmiennych**