

# Jakość życia w wybranych województwach w 2019 roku



PRZYGOTOWANE I ZAPREZENTOWANE PRZEZ

Hanna Sarnecka  
Aleksandra Zych  
Maja Świtaj  
Weronika Pijanowska

# SPIS TREŚCI



---

02

Wprowadzenie

---

09

Wstępna analiza

---

13

Porządkowanie liniowe

---

35

Analiza głównych składowych

---

41

Analiza skupień

---

53

Podsumowanie

---

54

Bibliografia

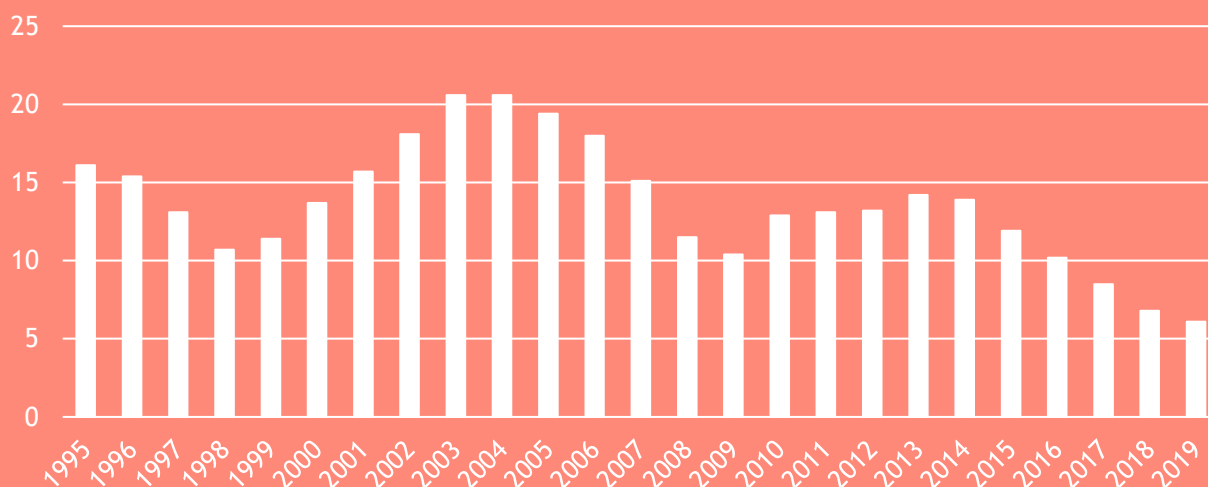
## Skrócony opis stanu gospodarki Polski w 2019 roku i latach poprzedzających

Wyjaśnienie stanu gospodarczego Polski w 2019 roku nie może pozostać pozbawione kontekstu. Polska gospodarka w roku 2018 zanotowała 5-procentowy wzrost PKB, a inflacja średnioroczna wyniosła 1,6%<sup>1</sup>. Deficyt sektora finansów publicznych był rekordowo niski i nie przekroczył 0,5 proc. PKB. Już w wtedy mówiono o rekordowo niskim bezrobociu, które wg GUS w kwartalnym badaniu BAEL wynosiło za III kw. br. 3,8 proc. (wyrównane sezonowo). Niskie bezrobocie sprzyjało wzrostowi wynagrodzeń w okolicach 7 proc. r/r. Tyle wystarczyło, by większość obywateli poczuła wzrost siły nabywczej swoich zarobków, a jednocześnie nie jest to na tyle dużo, aby zagrozić stabilności finansowej przedsiębiorstw. 2018 rok przyniósł również długo wyczekiwane dodatnie saldo migracji, co przynajmniej minimalnie zmniejszyło negatywne skutki starzenia się polskiego społeczeństwa. Dzięki napływowi do Polski Ukraińców czy Białorusinów mniejszy deficyt zanotował również Fundusz Ubezpieczeń Społecznych.

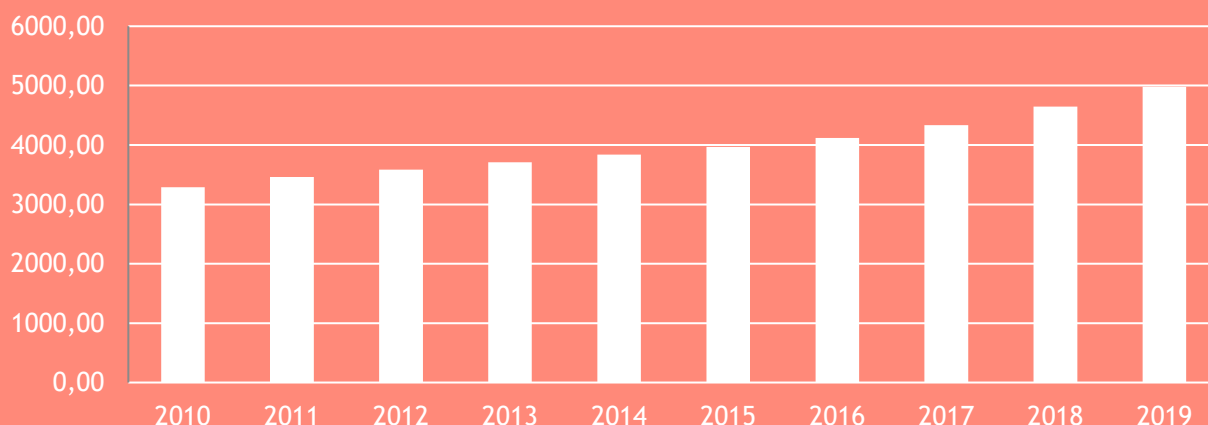
Na początku 2019 PKB rósł o prawie 5 proc, jednak ostatecznie na koniec roku zanotowano PKB poniżej 4 proc. - to najgorszy wynik od kilku lat<sup>2</sup>. PKB poszło w dół przeciwnie do bezrobocia, które w roku 2019 znalazło się na najniższym poziomie od blisko 30 lat. Stopa bezrobocia wynosiła 5,1 proc. Nie tylko spadało bezrobocie, ale i rosło zatrudnienie.

Pracownicy mogli również liczyć na podwyżki, średnio na poziomie 5,3 proc<sup>3</sup>.

### Stopa bezrobocia



### Przeciętne miesięczne wynagrodzenie



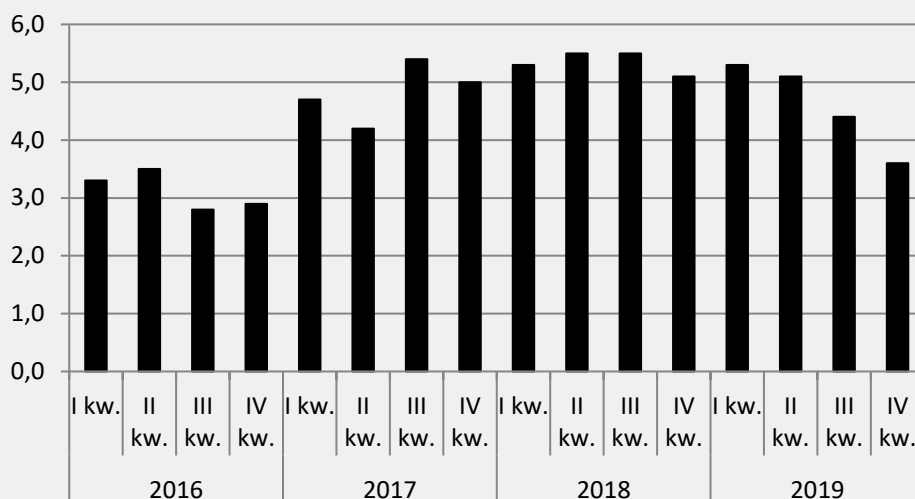
# 5,1%

## WZROST PKB 2018

# 4,0%

## WZROST PKB 2019

### Produkt krajowy brutto



Produkt Krajowy Brutto na przestrzeni lat wykazywał tendencję wzrostową. Wyjątek stanowił III i IV kwartał 2016 roku, po których jednak nastąpił znaczny wzrost na początku 2017 roku. Pierwsze niepokojące dane zanotowano w III kwartale 2019 roku, a już w IV kwartale wartość PKB była podobna jak na początku 2016 roku - niestety wtedy odnotowaliśmy już znaczny spadek.

### Jakość życia w Polsce w 2018 roku

Porównując do 2015 w roku 2018 nastąpiła poprawa sytuacji materialnej gospodarstw domowych. Wzrost odsetek gospodarstw domowych zadowolonych nie tylko ze swoich warunków życia, ale także z życia ogólnie rzecz biorąc.

Sytuacja dochodowa była wyraźnie zróżnicowana ze względu na miejsce zamieszkania. W największym stopniu ubóstwem dochodowym dotknięci byli mieszkańcy wsi (21%), dane wskazywały, że im większe miasto w które zamieszkujemy, tym mniejszy odsetek rodzin dotkniętych ubóstwem.



Badania dotyczyły również jakości środowiska naturalnego w miejscu zamieszkania. Wykazano, że zanieczyszczenie powietrza stanowiło poważny problem zimą dla 15% gospodarstw domowych w kraju, natomiast latem dla 6%. Problem był najmniej dotkliwie odczuwany na terenach wiejskich (zimą - 8%, latem - 3% gospodarstw domowych), ale nasilał się wraz ze wzrostem wielkości miast.

Zadowolenie z terenów zielonych w Polsce plasuje się na wyjątkowo wysokim poziomie. Mieszkańcy Polski na ogół (79% osób w wieku 16 lat i więcej) byli zadowoleni z terenów zielonych w swoim miejscu zamieszkania, przy czym mieszkańcy wsi byli z terenów zielonych bardziej zadowoleni (85%), niż mieszkańcy miast (75%).



Następnym, bardzo istotnym elementem jakości życia jest wskaźnik poczucia bezpieczeństwa w miejscu zamieszkania, który jest definiowany jako odsetek osób w wieku 16 lat i więcej deklarujących, że chodząc samotnie po zmroku w okolicy swojego miejsca zamieszkania, czują się bardzo bezpiecznie lub raczej bezpiecznie.

Mieszkańcy wsi wychodząc z domu po zmroku, czuli się w 2018 r. bardziej bezpiecznie (94%) niż mieszkańcy miast (86%). Poziom wskaźnika w miastach był jednak zróżnicowany, najniższy w miastach liczących od 100

tys. do 500 tys. mieszkańców (83%), najwyższy w miastach liczących poniżej 20 tys. mieszkańców (91%).

Mieszkańcy Polski mają zaufanie głównie do osób, które znają osobiście. Prawie wszyscy ufają swojej najbliższej rodzinie (98%). Również zaufanie do przyjaciół i znajomych, a także do współpracowników jest bardzo rozpowszechnione - odsetki osób ufających utrzymują się na poziomie odpowiednio 94 i 82%. Nieco mniejsze, ale również zdecydowanie dominujące nad nieufnością jest zaufanie do sąsiadów. Odsetek osób ufających sąsiadom wyniósł w pierwszej połowie 2018 r. 75%. W przypadku zaufania do sąsiadów zaobserwowano bardzo wyraźne zróżnicowanie ze względu na klasę miejscowości zamieszkania. Na wsi sąsiadom ufało 83%, a w miastach - 71% osób.

Policja cieszyła się zaufaniem 72% osób. Największe zaufanie do policji mieli mieszkańcy wsi (76%). W miastach było ono mniej powszechne (69%), a najmniej ufali policji mieszkańcy największych miast (67%).

Jednym z najbardziej podstawowych wskaźników stosowanych do oceny dobrobytu subiektywnego jest wskaźnik poziomu zadowolenia (satysfakcji) z życia ogólnie rzecz biorąc. Zakłada się bowiem, że oceniając poziom satysfakcji z własnego życia, dana osoba bierze pod uwagę wszystkie te jego aspekty, które uważa za istotne. W 2018 roku zadowolonych ze swojego życia ogólnie rzecz biorąc, było 83% mieszkańców Polski w wieku co najmniej 16 lat i jest to o 5 p. proc. więcej niż w 2015 r.<sup>4</sup>

Nie zaobserwowano wyraźnych różnic w poziomie zadowolenia z życia pomiędzy poszczególnymi klasami miejscowości ze względu na ich wielkość. Najwyższy poziom zadowolenia odnotowano na wsiach i w miastach poniżej 20 tys. Mieszkańców, było to aż po 84% osób zadowolonych. Najmniejszy - w miastach od 100 do 500 tys. oraz od 20 do 100 tys. mieszkańców po 81%<sup>5</sup>.

Podsumowując, jakość życia i poziom zadowolenia z życia wyraźnie wzrósł. Na jakość powietrza skarżyło się około 10,5% ludności, natomiast z terenów zielonych było zadowolonych aż 79%. Poczucie bezpieczeństwa deklarowało 90% mieszkańców Polski, co można uznać za wyjątkowo wysoki wynik. Ludzie czują się bezpiecznie również dlatego, że ufają osobom dookoła nich i mają zaufanie do służb mundurowych. Ogólnie zadowolonych z życia było aż 83 osoby na 100, czyli znacząca większość.





## PRZEDSTAWIENIE CELU I WYJAŚNIENIE PODSTAWOWYCH POJĘĆ

Jednym z nadrzędnych celów polityki społecznej i gospodarczej zarówno na szczeblu krajowym, regionalnym, jak i lokalnym jest wzrost jakości życia społeczeństwa. W ostatnich dwudziestu latach Unia Europejska coraz wyraźniej wskazuje istotność nie tylko wzrostu jakości życia, ale jednocześnie utrzymanie równowagi pomiędzy rozwojem gospodarczym i rozwojem społecznym, przy zachowaniu środowiska naturalnego oraz dziedzictwa kulturowego.

Prace nad stworzeniem systemu wskaźników monitorujących jakość życia prowadzone są od wielu lat przez instytucje międzynarodowe, jednakże dla potrzeb tej pracy wybraliśmy koncepcję pomiaru jakości życia w ramach Europejskiego Systemu Statystycznego. Została ona przedstawiona w raporcie końcowym Grupy Inicjatywnej ds. Pomiaru Postępu, Dobrostanu i Zrównoważonego Rozwoju (Eurostat, 2011a) oraz Zespołu Zadaniowego ds. Wielowymiarowego Pomiaru Jakości Życia (Eurostat, 2011b)<sup>6</sup>. Została przedstawiona lista dziedzin i subdziedzin kategorii jakości życia, a dla każdej z nich związane z nimi cele polityki społecznej. Ostatecznie zaproponowano, aby pomiar jakości życia obejmował dziewięć domen (dziedzin) jakości życia, w tym osiem wymiarów odnosi się do dziedzin związanych z warunkami życia, a dziewiąty dotyczy dobrostanu psychicznego.

CELEM ANALIZY JEST WYTYPOWANIE  
WOJEWÓDZTWA O NAJWYŻSZEJ  
JAKOŚCI ŻYCIA, NA PODSTAWIE  
WCZEŚNIEJ WYBRANYCH DZIEDZIN  
JAKOŚCI ŻYCIA.

Zmienne zostały wybrane na podstawie danych statystycznych za rok 2019 dostępnych na stronie głównego urzędu statystycznego

## Dziedzina 1 - Praca

x1 - S - Liczba pracujących ogółem na stu mieszkańców województwa

x2 - S - Liczba pracujących w sektorze publicznym na stu mieszkańców województwa

x3 - S - Liczba pracujących w sektorze prywatnym na stu mieszkańców województwa

x4 - D - Bezrobotni rejestrowani na stu mieszkańców województwa

x5 - S - Liczba wolnych miejsc pracy na tysiąc mieszkańców województwa

x6 - S - Wynagrodzenia w tysiącach

## Dziedzina 2 - Zdrowie

x7 - S - Oczekiwana długość życia w zdrowiu

x8 - S - Liczba lekarzy wykonujących zawód na tysiąc mieszkańców województwa



## Dziedzina 3 - Otaczające środowisko

x9 - S - Parki spacerowo-wypoczynkowe i zieleńce

x10 - D - Emisja zanieczyszczeń powietrza z zakładów szczególnie uciążliwych

## Dziedzina 4 - kultura i edukacja

x11 - S - Muzea łącznie z oddziałami, teatry, instytucje muzyczne oraz biblioteki

x12 - S - Widzowie i słuchacze (wg miejsca prezentacji) w teatrach i instytucjach muzycznych

x13 - S - Zdawalność egzaminów maturalnych

x14 - S - Liczba uczelni wyższych ogółem

## Dziedzina 5 - Bezpieczeństwo

x15 - D - Przestępstwa stwierdzone przez Policję w zakończonych postępowaniach przygotowawczych

x16 - D - Osoby w zamachach samobójczych

x17 - S - Wydatki na bezpieczeństwo publiczne i ochronę przeciwpożarową

# Macierz danych

Tabela 1. Początkowa macierz danych

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
dolnośląskie	9,82	2,12	7,70	20,53	11,71	5,32	59,30	14,63	26,22	10,24	6,79	37,51	78,70	33,00	7,76	10,96	13,34
kujawsko-pomorskie	5,74	1,50	4,23	26,52	4,87	4,49	59,80	7,77	9,23	9,73	4,51	28,68	78,30	17,00	3,83	4,01	7,91
lubelskie	5,46	1,70	3,76	29,32	3,11	4,56	59,40	10,68	9,70	4,99	6,32	11,70	80,10	17,00	3,36	6,67	9,88
lubuskie	2,74	0,72	2,03	5,51	2,62	4,56	60,60	3,26	9,19	3,32	2,76	40,11	79,50	6,00	2,58	3,69	6,15
łódzkie	8,18	1,89	6,29	23,69	8,84	4,79	58,70	14,21	7,44	38,21	5,57	23,84	79,90	23,00	4,58	12,52	11,69
małopolskie	11,78	2,91	8,87	23,63	15,31	5,10	60,10	16,72	17,95	9,00	8,71	37,90	84,80	29,00	8,00	10,75	10,90
mazowieckie	27,41	6,95	20,47	50,83	35,42	6,25	59,60	30,64	15,70	32,54	11,36	61,69	81,90	97,00	11,50	11,89	18,26
opolskie	2,67	0,72	1,95	7,21	2,60	4,71	59,90	3,14	4,98	13,80	3,31	20,82	80,30	6,00	1,79	2,15	7,20
podkarpackie	5,34	1,57	3,77	33,46	2,98	4,39	60,20	7,63	12,15	2,89	7,31	11,06	81,00	18,00	2,60	5,45	10,49
podlaskie	2,65	0,87	1,78	14,33	2,32	4,58	59,60	6,82	3,75	2,14	2,68	51,79	83,80	16,00	1,73	4,61	4,95
pomorskie	7,51	1,79	5,73	13,47	9,55	5,14	60,30	10,57	16,98	6,63	4,14	40,45	78,60	26,00	4,83	6,75	8,93
śląskie	15,55	4,21	11,34	21,19	18,69	5,18	59,30	21,11	29,82	32,43	8,73	41,33	79,80	34,00	10,89	16,20	9,70
świętokrzyskie	3,00	0,94	2,06	15,91	2,76	4,49	59,40	5,13	3,53	13,65	3,05	21,00	79,90	12,00	2,13	5,12	7,16
warmińsko-mazurskie	3,44	1,04	2,40	15,61	3,38	4,32	59,30	4,56	9,78	1,72	3,34	28,84	77,30	7,00	2,58	6,11	8,18
wielkopolskie	14,32	2,67	11,65	12,74	12,75	4,69	61,40	13,74	26,77	10,90	7,68	50,14	80,50	35,00	6,32	7,47	15,73
zachodniopomorskie	4,58	1,29	3,29	14,95	5,18	4,77	60,90	7,49	18,64	6,25	4,02	30,74	77,00	16,00	5,16	5,26	9,49

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Tabela pierwsza przedstawia dane, które pozyskaliśmy m.in. ze strony Głównego Urzędu Statystycznego przed wykonaniem jakichkolwiek transformacji.

Zmienne x1, x2 oraz x3 są podane w jednostce stu tysięcy, zmienne x4, x5, x6, x8 są podane w tysiącach, zmienna x10 podana jest w milionie ton/rok, x11 w setkach, 12 to liczba widzów na 1000 ludności, x13 to procentowa zdawalność, x15 podana jest w dziesięciu tysiącach, x16 w setkach, a x17 w jednostce dziesięciu milionów.



**Tabela 2. Przekształcona macierz danych**

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
dolnośląskie	33,85	7,31	26,54	7,08	40,36	53,24	59,30	50,43	90,38	35,31	23,40	37,51	78,70	113,75	26,75	37,78	45,97
kujawsko-pomorskie	27,62	7,24	20,38	12,76	23,44	44,94	59,80	37,38	44,42	46,84	21,71	28,68	78,30	81,82	18,42	19,30	38,09
lubelskie	25,78	8,01	17,77	13,84	14,69	45,65	59,40	50,42	45,81	23,55	29,84	11,70	80,10	80,28	15,86	31,50	46,67
lubuskie	27,02	7,05	19,96	5,43	25,82	45,60	60,60	32,16	90,58	32,72	27,20	40,11	79,50	59,14	25,48	36,37	60,66
łódzkie	33,17	7,65	25,52	9,61	35,84	47,90	58,70	57,60	30,17	154,94	22,58	23,84	79,90	93,26	18,56	50,76	47,41
małopolskie	34,64	8,57	26,07	6,95	45,02	50,98	60,10	49,15	52,79	26,46	25,61	37,90	84,80	85,28	23,53	31,61	32,06
mazowieckie	50,73	12,86	37,88	9,41	65,55	62,48	59,60	56,71	29,06	60,22	21,02	61,69	81,90	179,52	21,27	22,00	33,79
opolskie	27,03	7,30	19,74	7,31	26,36	47,12	59,90	31,81	50,48	139,90	33,55	20,82	80,30	60,82	18,18	21,79	73,00
podkarpackie	25,08	7,36	17,73	15,72	14,00	43,88	60,20	35,85	57,07	13,55	34,34	11,06	81,00	84,55	12,23	25,60	49,28
podlaskie	22,41	7,35	15,06	12,13	19,64	45,76	59,60	57,71	31,74	18,08	22,68	51,79	83,80	135,42	14,68	39,02	41,89
pomorskie	32,20	7,66	24,54	5,77	40,93	51,42	60,30	45,29	72,77	28,43	17,74	40,45	78,60	111,42	20,71	28,93	38,25
śląskie	34,30	9,29	25,01	4,67	41,23	51,78	59,30	46,57	65,78	71,54	19,26	41,33	79,80	75,00	24,01	35,73	21,39
świętokrzyskie	24,12	7,56	16,56	12,81	22,23	44,90	59,40	41,35	28,43	109,97	24,57	21,00	79,90	96,65	17,19	41,24	57,64
warmińsko-mazurskie	24,05	7,26	16,79	10,92	23,65	43,20	59,30	31,88	68,44	12,06	23,37	28,84	77,30	48,99	18,08	42,76	57,22
wielkopolskie	40,99	7,65	33,34	3,65	36,49	46,87	61,40	39,33	76,62	31,19	21,98	50,14	80,50	100,17	18,08	21,38	45,02
zachodniopomorskie	26,92	7,57	19,35	8,79	30,45	47,70	60,90	44,05	109,58	36,72	23,63	30,74	77,00	94,06	30,35	30,92	55,79

Źródło: opracowanie własne na podstawie obliczeń w programie *MS Excel*.

Tabela druga przedstawia dane podzielone przez liczbę mieszkańców danego województwa, a potem sprowadzone do odpowiedniej jednostki tak, aby można było łatwo przeprowadzać dalsze analizy.

Rzędy wielkości zostały przedstawione w rozdziale Wybrane Zmienne.

## Podstawowe miary dla zmiennych i histogramy

Tabela 3. Podstawowe miary dla zmiennych

Miary	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
Średnia	30,62	7,98	22,64	9,18	31,61	48,34	59,86	44,23	59,01	52,59	24,53	33,60	80,09	93,76	20,21	32,29	46,51
Mediana	27,33	7,56	20,17	9,10	28,40	47,00	59,70	44,67	54,93	34,02	23,39	34,12	79,90	89,27	18,49	31,55	46,32
Minimum	22,41	7,05	15,06	3,65	14,00	43,20	58,70	31,81	28,43	12,06	17,74	11,06	77,00	48,99	12,23	19,30	21,39
Maksimum	50,73	12,86	37,88	15,72	65,55	62,48	61,40	57,71	109,58	154,94	34,34	61,69	84,80	179,52	30,35	50,76	73,00
Odch.std	7,38	1,42	6,32	3,58	13,29	4,81	0,69	9,04	24,61	44,51	4,66	14,23	2,09	31,59	4,77	8,94	12,65
Wsp.zmn.	24,10	17,77	27,92	39,00	42,05	9,96	1,16	20,44	41,70	84,62	18,99	42,34	2,61	33,70	23,60	27,67	27,20
Skośność	1,49	3,09	1,12	0,21	0,97	1,82	0,67	0,08	0,49	1,45	0,97	0,18	0,83	1,30	0,51	0,28	0,10

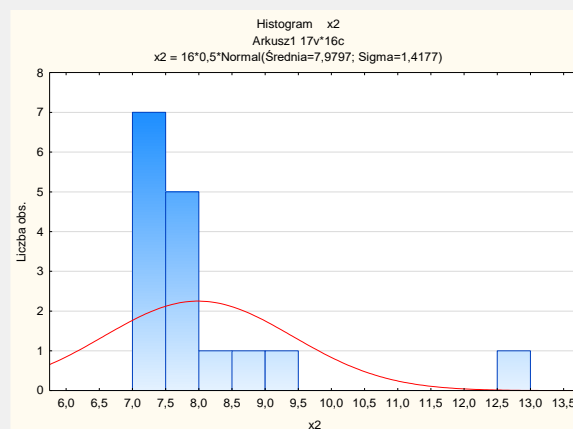
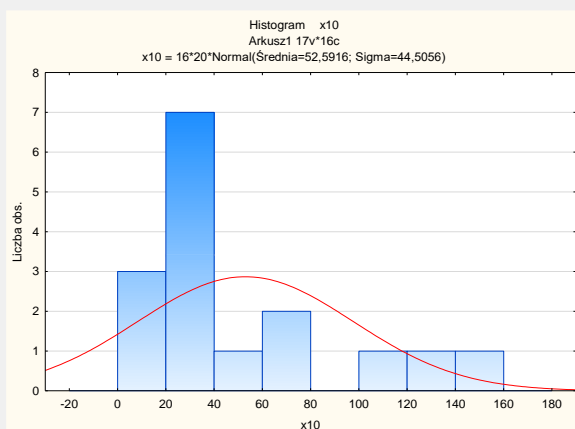
Źródło: opracowanie własne na podstawie obliczeń w programie Statistica.

Najmniejsze odchylenie standardowe ma zmienna x7 czyli oczekiwana długość życia w zdrowiu, oznacza to, iż w każdym województwie ludzie cieszą się zdrowiem podobnie długo. Natomiast największe odchylenie zanotowaliśmy w zmiennej x10, czyli emisji zanieczyszczeń powietrza z zakładów szczególnie uciążliwych. Dzięki temu wiemy, że rozbieżność emisji w poszczególnych województwach jest znacząca.

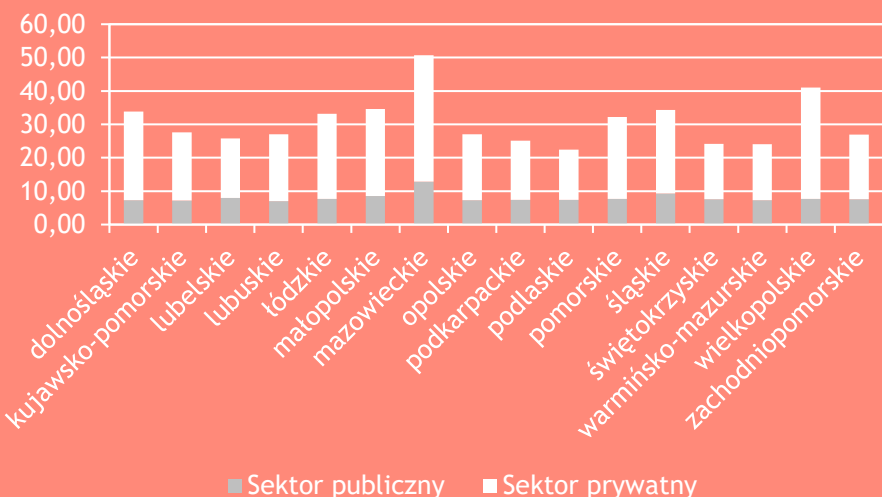
W tym przypadku badamy bardzo różnorodne cechy, z tego względu powinniśmy bardziej skupić się na współczynniku zmienności, aby dotrzeć do rzeczywistej siły rozproszenia cechy.

Po wyliczeniu wartości współczynnika zmienności dochodzimy do podobnych wniosków. Współczynnik zmiennej x10 wynosi aż 84%, co wskazuje silne zróżnicowanie. Współczynnik wspomnianej wcześniej zmiennej x7 wynosi tylko 1%, czyli bardzo niskie zróżnicowanie. Podobnie niskie zróżnicowania możemy zaobserwować również w zmiennych x6 na poziomie 10% oraz zmiennej x13 na poziomie 3%.

Zmienna x10 o największym rozproszeniu jest prawostronnie asymetryczna, na co wskazuje skośność na poziomie 1,45. Jednakże największą skośność posiada cecha x2 równa 3,09. W powyższym zestawieniu żadna zmienna nie wykazuje asymetrii lewostronnej.

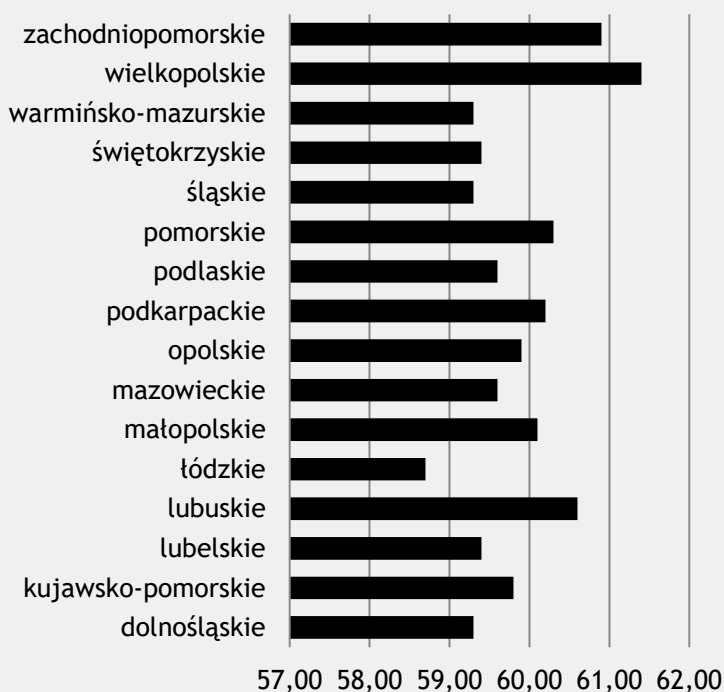


## Liczba osób pracujących w różnych sektorach na stu mieszkańców województwa

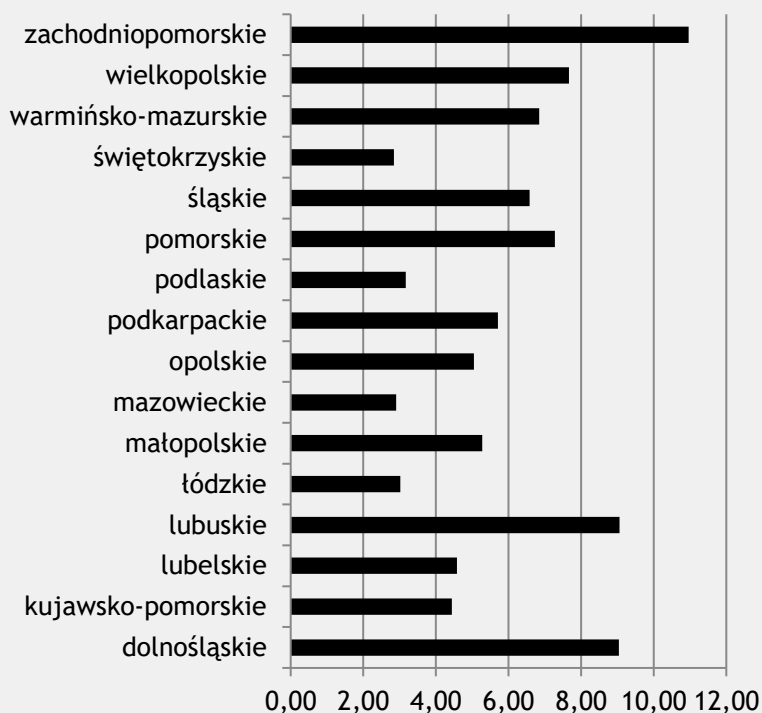


Najwięcej osób pracujących na stu mieszkańców jest w województwie mazowieckim. Liczba osób pracujących w sektorze publicznym jest znacznie wyższa niż w innych województwach, może być to spowodowane tym, iż na jego terenie znajduje się Warszawa, czyli stolica naszego kraju, w której stacjonuje aparat rządowy.

## Oczekiwana długość życia w zdrowiu

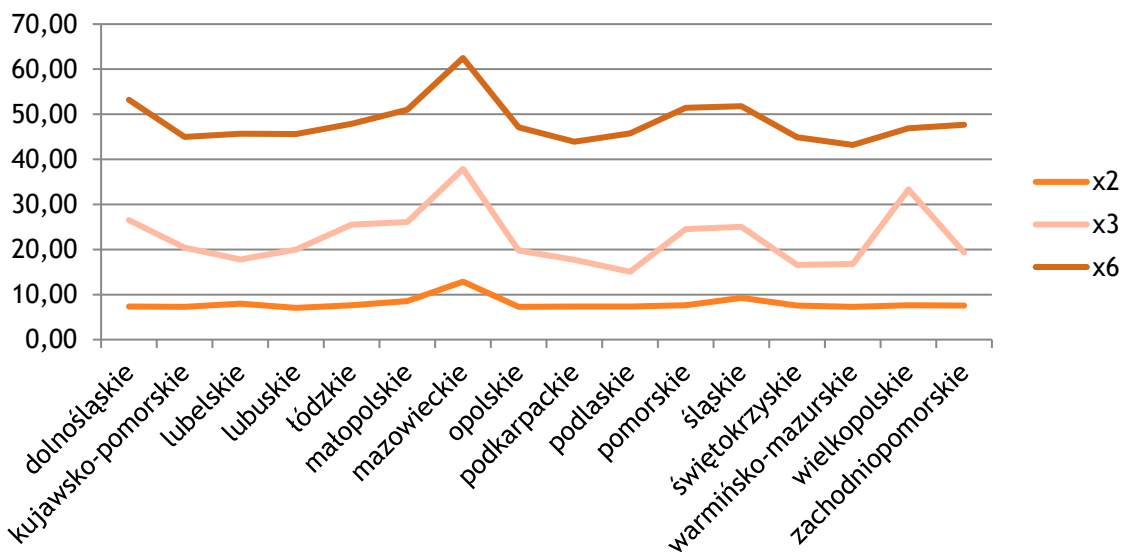


## Parki spacerowo-wypoczynkowe i zieleńce na dziesięć tysięcy osób



Najdłużej w zdrowiu żyją ludzie zamieszkujący województwa zachodniopomorskie, wielkopolskie oraz lubuskie, analogicznie w tych samych województwach jest największa ilość parków przypadających na dziesięć tysięcy osób. Wyjątkiem jest województwo dolnośląskie, które również ma na swoim terenie dużą ilość parków, ale ludzie żyją tam stosunkowo krótko w zdrowiu.

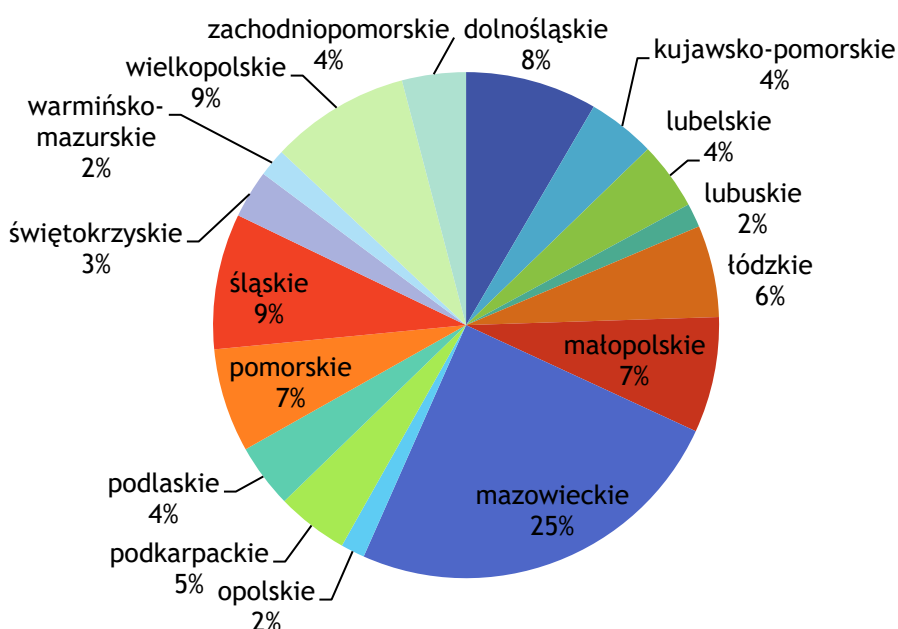
### Porównanie ilości osób pracujących w różnych sektorach do wysokości płac w poszczególnych województwach



Najbardziej podobnymi wartościami cechuje się województwo mazowieckie. Na jego wysokości możemy zauważyć wyraźny pik. Występuje tam największa ilość osób pracujących na jednego mieszkańca, co może być spowodowane wyższymi niż w innych rejonach zarobkami.

Dużą dysproporcję widzimy w województwie wielkopolskim. Inaczej niż w przypadku województwa mazowieckiego większa ilość osób pracujących nie idzie w parze z wyższymi niż przeciętnie zarobkami. Obserwujemy tu znacznie wyższą niż przeciętnie ilość osób pracujących w sektorze prywatnym.

### Wszystkie uniwersytety w podziale terytorialnym



Jak możemy zauważyć, największa ilość uniwersytetów znajduje się na terenie województwa mazowieckiego i jest to aż 25% wszystkich uczelni wyższych w naszym kraju. Na drugim miejscu z 9% wszystkich uniwersytetów są województwa wielkopolskie i śląskie. Natomiast najmniej jest w województwach warmińsko-mazurskim, lubuskim i opolskim.

# Korelacje

Tabela 4. Macierz korelacji zmiennych

Zmienna	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
x1	1,00	0,79	0,99	-0,50	0,91	0,85	0,11	0,40	-0,06	0,09	-0,42	0,65	0,23	0,60	0,27	-0,33	-0,51
x2		1,00	0,70	-0,11	0,77	0,85	-0,16	0,48	-0,33	0,06	-0,29	0,53	0,33	0,66	0,13	-0,26	-0,52
x3			1,00	-0,55	0,89	0,80	0,16	0,36	0,00	0,09	-0,43	0,64	0,20	0,56	0,29	-0,33	-0,48
x4				1,00	-0,58	-0,39	-0,38	0,05	-0,51	-0,08	0,42	-0,57	0,04	0,04	-0,62	0,07	0,16
x5					1,00	0,93	-0,01	0,45	0,01	0,13	-0,56	0,71	0,19	0,58	0,49	-0,15	-0,52
x6						1,00	-0,12	0,56	-0,07	0,08	-0,42	0,63	0,24	0,72	0,42	-0,21	-0,52
x7							1,00	-0,42	0,58	-0,42	0,08	0,23	-0,03	-0,05	0,23	-0,58	0,15
x8								1,00	-0,39	0,09	-0,42	0,37	0,44	0,72	0,06	0,30	-0,57
x9									1,00	-0,42	-0,06	0,07	-0,53	-0,32	0,69	-0,08	0,17
x10										1,00	0,07	-0,23	-0,04	-0,06	-0,05	0,23	0,27
x11											1,00	-0,66	0,17	-0,41	-0,36	-0,18	0,61
x12												1,00	0,30	0,63	0,34	-0,14	-0,47
x13													1,00	0,38	-0,31	-0,08	-0,34
x14														1,00	0,01	-0,17	-0,43
x15															1,00	0,08	-0,10
x16																1,00	0,09
x17																	1,00

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Współczynnik korelacji dla zmiennych x1 oraz x2 i x3 jest wyjątkowo wysoki, kolejno 0,79 i 0,99. Może być to spowodowane tym, że zmienne te opisują podobne zjawisko, czyli liczba pracujących ogółem, liczba pracujących w sektorze publicznym i liczba pracujących w sektorze prywatnym. Z tego względu w dalszych analizach nie będziemy uwzględniać cechy x1.

Podobnie wysoką korelację wykazują zmienne x5 i x6 które mówią nam kolejno o ilości wolnych miejsc pracy oraz wysokości wynagrodzenia.

Zmienna x6 koreluje również ze zmiennymi x1, x2, x3.

Ujemną korelację możemy zaobserwować między zmiennymi x11 oraz x12. X 11 to muzea łącznie z oddziałami, teatry, instytucje muzyczne oraz biblioteki, a x12 widzowie i słuchacze (wg miejsca prezentacji) w teatrach i instytucjach muzycznych na 1000 ludności. W tym przypadku to bardzo logiczne, że im mniej miejsc kultury, tym mniej słuchaczy i widzów.





## Wstęp

Porządkowanie liniowe ma pomóc określić, jak w 2019 r. kształtował się poziom jakości życia w województwach - w których z nich żyło się najlepiej, najgorzej lub przeciętnie. Kolejność obiektów jest wyznaczana na podstawie wybranych kryteriów. Tego typu porządkowanie spełniania dwa postulaty logiczne (prawo asymetrii i prawo przechodniości).

W porządkowaniu liniowym obiekty są uszeregowane ze względu na jedną cechę, ale jednocześnie obiekty te są wielowymiarowe. Aby pogodzić obie kwestie, dla zespołu cech należy wyznaczyć sumaryczną cechę. Najczęściej jest to syntetyczny wskaźnik, będący funkcją wartości zmiennych zespołu cech<sup>7</sup>.

Istnieją różne sposoby konstruowania takich funkcji i trudno jednoznacznie stwierdzić, która z nich jest najlepsza. W projekcie zostały wykorzystane następujące metody:

- metoda wzorca (z  $m_i$  i z  $q_i$ ),
- metoda wzorca z wagami (z  $m_i$  i z  $q_i$ ),
- metoda TOPSIS,
- metoda standaryzowanych sum,
- metoda sumy rang,
- metoda dystansów.



# Przygotowanie macierzy - redukcja zmiennych

Przed zastosowaniem metod porządkowania liniowego należy odpowiednio przygotować macierz, a dokładniej rozważane w niej zmienne. Może się bowiem okazać, że dla wiarygodności wyników niektóre z cech powinno się usunąć. Dotyczy to zmiennych, które pokrywają się, są bardzo podobne oraz takich, których współczynnik zmienności jest bardzo niski. W celu wyznaczenia omawianego współczynnika obliczono najpierw średnią z danej zmiennej i jej odchylenie standardowe. Następnie wyliczono już sam współczynnik zmienności (odchylenie podzielone przez średnią, pomnożone razy sto).

**Tabela 5. Macierz przed redukcją zmiennych**

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
dolnośląskie	33,85	7,31	26,54	7,08	40,36	53,24	59,30	50,43	90,38	35,31	23,40	37,51	78,70	113,75	26,75	37,78	45,97
kujawsko-pomorskie	27,62	7,24	20,38	12,76	23,44	44,94	59,80	37,38	44,42	46,84	21,71	28,68	78,30	81,82	18,42	19,30	38,09
lubelskie	25,78	8,01	17,77	13,84	14,69	45,65	59,40	50,42	45,81	23,55	29,84	11,70	80,10	80,28	15,86	31,50	46,67
lubuskie	27,02	7,05	19,96	5,43	25,82	45,60	60,60	32,16	90,58	32,72	27,20	40,11	79,50	59,14	25,48	36,37	60,66
łódzkie	33,17	7,65	25,52	9,61	35,84	47,90	58,70	57,60	30,17	154,94	22,58	23,84	79,90	93,26	18,56	50,76	47,41
małopolskie	34,64	8,57	26,07	6,95	45,02	50,98	60,10	49,15	52,79	26,46	25,61	37,90	84,80	85,28	23,53	31,61	32,06
mazowieckie	50,73	12,86	37,88	9,41	65,55	62,48	59,60	56,71	29,06	60,22	21,02	61,69	81,90	179,52	21,27	22,00	33,79
opolskie	27,03	7,30	19,74	7,31	26,36	47,12	59,90	31,81	50,48	139,90	33,55	20,82	80,30	60,82	18,18	21,79	73,00
podkarpackie	25,08	7,36	17,73	15,72	14,00	43,88	60,20	35,85	57,07	13,55	34,34	11,06	81,00	84,55	12,23	25,60	49,28
podlaskie	22,41	7,35	15,06	12,13	19,64	45,76	59,60	57,71	31,74	18,08	22,68	51,79	83,80	135,42	14,68	39,02	41,89
pomorskie	32,20	7,66	24,54	5,77	40,93	51,42	60,30	45,29	72,77	28,43	17,74	40,45	78,60	111,42	20,71	28,93	38,25
śląskie	34,30	9,29	25,01	4,67	41,23	51,78	59,30	46,57	65,78	71,54	19,26	41,33	79,80	75,00	24,01	35,73	21,39
świętokrzyskie	24,12	7,56	16,56	12,81	22,23	44,90	59,40	41,35	28,43	109,97	24,57	21,00	79,90	96,65	17,19	41,24	57,64
warmińsko-mazurskie	24,05	7,26	16,79	10,92	23,65	43,20	59,30	31,88	68,44	12,06	23,37	28,84	77,30	48,99	18,08	42,76	57,22
wielkopolskie	40,99	7,65	33,34	3,65	36,49	46,87	61,40	39,33	76,62	31,19	21,98	50,14	80,50	100,17	18,08	21,38	45,02
zachodniopomorskie	26,92	7,57	19,35	8,79	30,45	47,70	60,90	44,05	109,58	36,72	23,63	30,74	77,00	94,06	30,35	30,92	55,79
wsp. zmn.	23,33	17,20	27,03	37,76	40,72	9,64	1,12	19,79	40,38	81,94	18,38	41,00	2,53	32,63	22,86	26,79	26,33

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

W tabeli 5 zaznaczono, które ze zmiennych zostały wyłączone z analiz związanych z porządkowaniem liniowym.

Powody wykluczenia poszczególnych zmiennych:

- **x1 (pracujący ogółem)** - zawiera w sobie dane z dwóch innych zmiennych - x2 (pracujący w sektorze publicznym) i x3 (pracujący w sektorze prywatnym);
- **x7 (oczekiwana długość życia w zdrowiu)** - niski współczynnik zmienności (1,12);
- **x13 (zdawalność egzaminów maturalnych)** - niski współczynnik zmienności (2,53).

Po redukcji zostało zatem 14 zmiennych.

# Macierz danych wykorzystywana w analizach

Poniżej zamieszczono ostateczną macierz danych, która została wykorzystana w analizach związanych z porządkowaniem liniowym.

**Tabela 6. Macierz po redukcji zmiennych z podziałem na stymulanty i destymulanty**

Charakter zmiennej	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	7,31	26,54	7,08	40,36	53,24	50,43	90,38	35,31	23,40	37,51	113,75	26,75	37,78	45,97
kujawsko-pomorskie	7,24	20,38	12,76	23,44	44,94	37,38	44,42	46,84	21,71	28,68	81,82	18,42	19,30	38,09
lubelskie	8,01	17,77	13,84	14,69	45,65	50,42	45,81	23,55	29,84	11,70	80,28	15,86	31,50	46,67
lubuskie	7,05	19,96	5,43	25,82	45,60	32,16	90,58	32,72	27,20	40,11	59,14	25,48	36,37	60,66
łódzkie	7,65	25,52	9,61	35,84	47,90	57,60	30,17	154,94	22,58	23,84	93,26	18,56	50,76	47,41
małopolskie	8,57	26,07	6,95	45,02	50,98	49,15	52,79	26,46	25,61	37,90	85,28	23,53	31,61	32,06
mazowieckie	12,86	37,88	9,41	65,55	62,48	56,71	29,06	60,22	21,02	61,69	179,52	21,27	22,00	33,79
opolskie	7,30	19,74	7,31	26,36	47,12	31,81	50,48	139,90	33,55	20,82	60,82	18,18	21,79	73,00
podkarpackie	7,36	17,73	15,72	14,00	43,88	35,85	57,07	13,55	34,34	11,06	84,55	12,23	25,60	49,28
podlaskie	7,35	15,06	12,13	19,64	45,76	57,71	31,74	18,08	22,68	51,79	135,42	14,68	39,02	41,89
pomorskie	7,66	24,54	5,77	40,93	51,42	45,29	72,77	28,43	17,74	40,45	111,42	20,71	28,93	38,25
śląskie	9,29	25,01	4,67	41,23	51,78	46,57	65,78	71,54	19,26	41,33	75,00	24,01	35,73	21,39
świętokrzyskie	7,56	16,56	12,81	22,23	44,90	41,35	28,43	109,97	24,57	21,00	96,65	17,19	41,24	57,64
warmińsko-mazurskie	7,26	16,79	10,92	23,65	43,20	31,88	68,44	12,06	23,37	28,84	48,99	18,08	42,76	57,22
wielkopolskie	7,65	33,34	3,65	36,49	46,87	39,33	76,62	31,19	21,98	50,14	100,17	18,08	21,38	45,02
zachodniopomorskie	7,57	19,35	8,79	30,45	47,70	44,05	109,58	36,72	23,63	30,74	94,06	30,35	30,92	55,79

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Stosowanie niektórych metod porządkowania wymagało rozróżnienia **charakteru** badanych cech. Przyjęto podział na zmienne pozytywne (stymulanty) i negatywne (destymulanty).

W przypadku **stymulant** wystąpienie wysokich wartości jest korzystne. Niskie wartości nie są pożądane<sup>8</sup>. W projekcie do zmiennych pozytywnych należą: x2 (pracujący w sektorze publicznym), x3 (pracujący w sektorze prywatnym), x5 (wolne miejsca pracy), x6 (wynagrodzenia), x8 (lekarze wykonujący zawód), x9 (parki spacerowo-wypoczynkowe, zieleńce), x11 (muzea, teatry, instytucje muzyczne, biblioteki), x12 (widzowie, słuchacze w teatrach i instytucjach muzycznych), x14 (uczelnie wyższe) i x17 (bezpieczeństwo publiczne i ochrona przeciwpożarowa).

Odwrotnie wygląda sytuacja dla **destymulant**. Niskie wartości są pożądane, wysokie - niepożądane<sup>9</sup>. Do zmiennych negatywnych zaliczono: x4 (bezrobotni rejestrowani), x10 (emisja zanieczyszczeń powietrza z zakładów szczególnie uciążliwych), x15 (przestępstwa) oraz x16 (zamachy samobójcze).

Rozróżnienie charakteru cech zostało zaznaczone w tabeli 6.

# Metoda wzorca

Metoda wzorca (nazywana też metodą wzorca rozwoju) utrwaliła się w polskich badaniach, jednak w obcojęzycznej literaturze raczej się nie pojawia. Idea tej metody została zaprezentowana przez Z. Hellwiga w 1968 r. Polega ona na stworzeniu hipotetycznego wzorca, względem którego określa się odległości rzeczywistych punktów. Te odległości są podstawą porządkowania obiektów<sup>10</sup>.

Aby skorzystać z MW wartości zmiennych muszą być ujednolicone, zestandaryzowane - tak jak w tabeli 7. W celu otrzymania tych wartości obliczono najpierw średnią z danej zmiennej i jej odchylenie standardowe. Następnie od każdej wartości zmiennej odjęto średnią, po czym wynik tego działania podzielono przez odchylenie standardowe.

**Tabela 7. Macierz standaryzowana metody wzorca**

Charakter zmiennej	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	-0,49	<b>0,64</b>	-0,61	0,68	<b>1,05</b>	0,71	<b>1,32</b>	-0,40	-0,25	0,28	<b>0,65</b>	<b>1,42</b>	0,63	-0,04
kujawsko-pomorskie	<b>-0,54</b>	-0,37	1,03	-0,63	-0,73	-0,78	-0,61	-0,13	-0,63	-0,36	-0,39	-0,39	<b>-1,50</b>	-0,69
lubelskie	0,02	-0,79	<b>1,35</b>	<b>-1,31</b>	-0,58	0,71	-0,55	-0,67	<b>1,18</b>	<b>-1,59</b>	-0,44	<b>-0,94</b>	-0,09	0,01
lubuskie	<b>-0,68</b>	-0,44	<b>-1,08</b>	-0,45	-0,59	<b>-1,38</b>	<b>1,33</b>	-0,46	0,59	0,47	<b>-1,13</b>	<b>1,14</b>	0,47	<b>1,16</b>
łódzkie	-0,24	0,47	0,12	0,33	-0,09	<b>1,53</b>	<b>-1,21</b>	<b>2,38</b>	-0,43	-0,71	-0,02	-0,36	<b>2,13</b>	0,07
małopolskie	<b>0,43</b>	0,56	-0,64	<b>1,04</b>	0,57	0,56	-0,26	-0,61	0,24	0,31	-0,28	0,72	-0,08	<b>-1,18</b>
mazowieckie	<b>3,55</b>	<b>2,49</b>	0,07	<b>2,64</b>	<b>3,03</b>	<b>1,43</b>	<b>-1,26</b>	0,18	<b>-0,78</b>	<b>2,04</b>	<b>2,80</b>	0,23	-1,19	<b>-1,04</b>
opolskie	-0,50	-0,47	-0,54	-0,41	-0,26	<b>-1,42</b>	-0,36	<b>2,03</b>	<b>2,00</b>	<b>-0,93</b>	<b>-1,08</b>	-0,44	<b>-1,21</b>	<b>2,16</b>
podkarpackie	-0,45	-0,80	<b>1,89</b>	<b>-1,37</b>	<b>-0,96</b>	-0,96	-0,08	<b>-0,91</b>	<b>2,17</b>	<b>-1,64</b>	-0,30	<b>-1,73</b>	-0,77	0,23
podlaskie	-0,46	<b>-1,24</b>	0,85	<b>-0,93</b>	-0,55	<b>1,54</b>	-1,14	<b>-0,80</b>	-0,41	<b>1,32</b>	<b>1,36</b>	<b>-1,20</b>	0,78	-0,38
pomorskie	-0,23	0,31	-0,98	0,72	0,66	0,12	0,58	-0,56	<b>-1,51</b>	0,50	0,58	0,11	-0,39	-0,67
śląskie	<b>0,95</b>	0,39	<b>-1,30</b>	<b>0,75</b>	<b>0,74</b>	0,27	0,28	0,44	<b>-1,17</b>	0,56	-0,61	0,82	0,40	<b>-2,05</b>
świętokrzyskie	-0,30	<b>-0,99</b>	<b>1,05</b>	-0,73	<b>-0,74</b>	-0,33	<b>-1,28</b>	<b>1,33</b>	0,01	-0,91	0,09	-0,65	<b>1,03</b>	<b>0,91</b>
warmińsko-mazurskie	<b>-0,53</b>	<b>-0,96</b>	0,50	-0,62	<b>-1,10</b>	<b>-1,41</b>	0,40	<b>-0,94</b>	-0,26	-0,35	<b>-1,46</b>	-0,46	<b>1,21</b>	0,87
wielkopolskie	-0,24	<b>1,75</b>	<b>-1,60</b>	0,38	-0,31	-0,56	0,74	-0,50	-0,57	<b>1,20</b>	0,21	-0,46	<b>-1,26</b>	-0,12
zachodniopomorskie	-0,30	-0,54	-0,11	-0,09	-0,14	-0,02	<b>2,12</b>	-0,37	-0,20	-0,21	0,01	<b>2,19</b>	-0,16	0,76
<b>WZORZEC</b>	<b>3,55</b>	<b>2,49</b>	<b>-1,60</b>	<b>2,64</b>	<b>3,03</b>	<b>1,54</b>	<b>2,12</b>	<b>-0,94</b>	<b>2,17</b>	<b>2,04</b>	<b>2,80</b>	<b>-1,73</b>	<b>-1,50</b>	<b>2,16</b>
<b>ANTYWZORZEC</b>	<b>-0,68</b>	<b>-1,24</b>	<b>1,89</b>	<b>-1,37</b>	<b>-1,10</b>	<b>-1,42</b>	<b>-1,28</b>	<b>2,38</b>	<b>-1,51</b>	<b>-1,64</b>	<b>-1,46</b>	<b>2,19</b>	<b>2,13</b>	<b>-2,05</b>

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Mając tak znormalizowane zmienne, można było przystąpić do konstruowania dwóch modelowych obiektów. Pierwszy z nich czyli wzorec pozytywny jest nierzeczywistym obiektem o „najlepszych” wartościach zmiennych. W przypadku stymulant wybiera się wartości najwyższe, dla destymulant - najniższe.

Drugim nierzeczywistym obiektem był wzorec negatywny (nazywany też antywzorcem) z „najgorszymi” wartościami zmiennych. Kiedy zmienna jest pozytywna, poszukuje się wartości najniższej, kiedy negatywna - najwyższej<sup>11</sup>. Wektory wzorca (pozytywnego) i antywzorca znajdują się w dwóch ostatnich wierszach tabeli 7.

# Metoda wzorca

W tabeli 7 najbardziej korzystne wartości zaznaczono na niebiesko, a najmniej korzystne - na czerwono.

W przypadku aż **sześciu cech najlepsze było województwo mazowieckie**. Omawiane województwo przodowało w liczbie: pracujących w sektorze publicznym (x2), pracujących w sektorze prywatnym (x3), wolnych miejsc pracy (x5), widzów i słuchaczy w teatrach i instytucjach muzycznych (x12), uczelni wyższych (x14) oraz w wysokości wynagrodzenia (x6).

Pod względem dwóch innych cech - liczby muzeów, teatrów, instytucji muzycznych oraz bibliotek (x11) i liczby przestępstw (x15) - najlepiej wyglądała sytuacja w woj. podkarpackim. W kujawsko-pomorskim popełniano najmniej zamachów samobójczych (x16). Warmińsko-mazurskie wypadło najbardziej korzystnie ze względu na emisję zanieczyszczeń (x10), w wielkopolskim najlepiej radzono sobie z bezrobociem (x4). Pod względem liczby parków i zieleńców (x9) wygrało zachodniopomorskie.

Opolskie przodowało w wysokości wydatków na bezpieczeństwo publiczne i ochronę przeciwpożarową (x17), ale zajęło ostatnie miejsce ze względu na liczbę lekarzy (x8). Podlaskie było najlepsze, jeśli chodzi o liczbę lekarzy (x8), najgorsze pod względem liczby pracujących w sektorze prywatnym (x3).

W przypadku **trzech cech**, takich jak: liczba bezrobotnych (x4), liczba wolnych miejsc pracy (x5), liczba widzów i słuchaczy w teatrach i instytucjach muzycznych (x12), najgorzej wyglądała sytuacja w województwie **podkarpackim**. Łódzkie wypadło najmniej korzystnie ze względu na dwie zmienne - emisja zanieczyszczeń (x10) i zamachy samobójcze (x16).

Skonstruowano również macierz pokazującą, jak bardzo rzeczywiste wartości zmiennych są oddalone od wartości wzorca pozytywnego (tabela 8).

**Tabela 8. Macierz z odległościami metody wzorca**

Charakter zmiennej	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	16,32	3,43	0,98	3,83	3,94	0,69	0,65	0,29	5,88	3,08	4,62	9,87	4,56	4,87
kujawsko-pomorskie	16,75	8,18	6,92	10,71	14,16	5,40	7,48	0,65	7,84	5,74	10,20	1,79	0,00	8,13
lubelskie	12,45	10,79	8,66	15,62	13,04	0,69	7,17	0,07	0,99	13,17	10,52	0,62	1,99	4,62
lubuskie	17,91	8,57	0,26	9,53	13,12	8,52	0,64	0,23	2,50	2,45	15,49	8,22	3,89	1,02
łódzkie	14,40	4,08	2,96	5,33	9,79	0,00	11,11	10,99	6,79	7,55	7,95	1,88	13,23	4,37
małopolskie	9,76	3,72	0,91	2,54	6,09	0,96	5,68	0,11	3,74	2,98	9,49	5,98	2,03	11,18
mazowieckie	0,00	0,00	2,76	0,00	0,00	0,01	11,42	1,25	8,71	0,00	0,00	3,83	0,10	10,25
opolskie	16,40	8,79	1,12	9,28	10,86	8,76	6,15	8,80	0,03	8,80	15,06	1,66	0,08	0,00
podkarpackie	16,06	10,84	12,13	16,05	15,93	6,24	4,86	0,00	0,00	13,51	9,64	0,00	0,53	3,75
podlaskie	16,07	13,91	5,99	12,73	12,87	0,00	10,67	0,02	6,68	0,52	2,08	0,28	5,19	6,45
pomorskie	14,34	4,75	0,38	3,66	5,63	2,02	2,39	0,14	13,54	2,38	4,96	3,36	1,24	8,05
śląskie	6,76	4,42	0,09	3,57	5,27	1,62	3,38	1,91	11,18	2,18	11,67	6,50	3,61	17,76
świętokrzyskie	14,87	12,13	6,99	11,33	14,22	3,49	11,60	5,16	4,69	8,72	7,34	1,15	6,43	1,57
warmińsko-mazurskie	16,64	11,88	4,41	10,60	17,12	8,71	2,98	0,00	5,91	5,69	18,21	1,60	7,35	1,66
wielkopolskie	14,39	0,55	0,00	5,10	11,22	4,41	1,91	0,20	7,50	0,70	6,73	1,60	0,06	5,22
zachodniopomorskie	14,86	9,16	2,20	7,44	10,06	2,44	0,00	0,33	5,63	5,05	7,80	15,37	1,80	1,97

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.



# Metoda wzorca

Od rzeczywistych wartości zmiennych odjęto wartość wzorcową, po czym wynik podniesiono do potęgi drugiej. Tak wyliczone wartości zsumowano dla każdego województwa, a następnie spierwiastkowano. W ten sposób obliczono  $d_{i0}$ , które można wyrazić także wzorem:

$$d_{i0} = \sqrt{\sum_{j=1}^p (z_{ij} - z_{0j})^2} \quad i = 1, 2, \dots, n$$

Im mniejsza jest taka odległość, tym bardziej dany obiekt jest podobny do wzorca.

Kolejne wyliczenia zależały od tego, jaką metodę przyjęto - standardową (z  $d_0$  i  $m_i$ ) lub zaprezentowaną w artykule A. Bąka<sup>12</sup> (z  $d_{0'}$  i  $q_i$ ). Według pierwszej z nich od wartości wzorca pozytywnego odjęto wartości negatywnego, podniesiono do potęgi drugiej, po czym zsumowano je i spierwiastkowano. W ten sposób otrzymano  $d_0$ :

$$d_0 = \sqrt{\sum_{j=1}^p (z_{0j} - z_{-0j})^2}$$

Przygotowanie  $d_{i0}$  oraz  $d_0$  pozwoliło obliczyć dla każdego województwa taksonomiczną miarę  $m_i$ .  $D_{i0}$  danego województwa podzielono przez  $d_0$  i iloraz ten odjęto od 1:

$$m_i = 1 - \frac{d_{i0}}{d_0}$$

Uporządkowania województw można również dokonać za sprawą wartości zmiennej agregatywnej  $q_i$ :

$$q_i = 1 - \frac{d_{i0}}{d_{0'}}$$

Jak widać, wzór nieznacznie różni się od tego na  $m_i$ . Tym razem trzeba było wyliczyć  $d_{0'}$ . Aby je otrzymać, należało ustalić średnią i odchylenie standardowe z wartości  $d_{i0}$  wszystkich województw, a następnie pomnożyć odchylenie razy dwa i dodać średnią<sup>13</sup>.

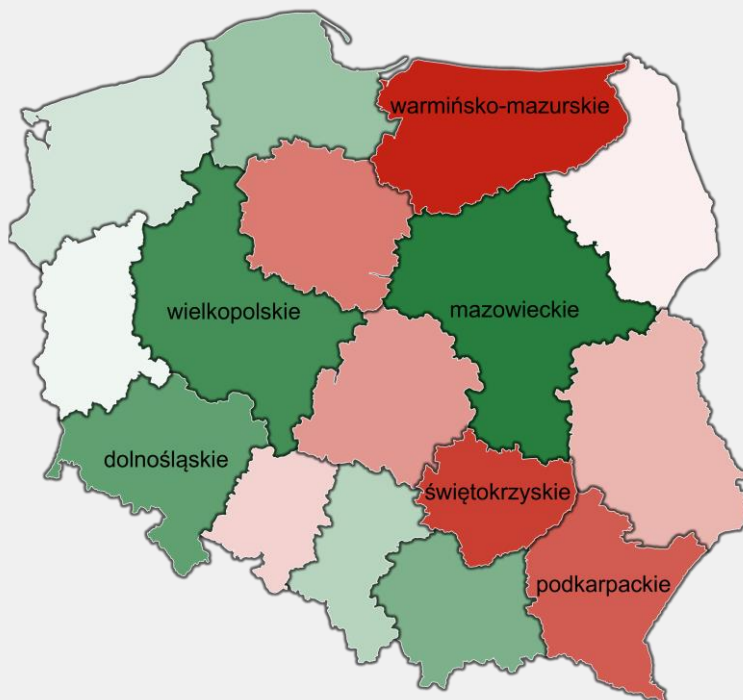
Wartości uzyskane za pomocą  $q_i$  oraz  $m_i$  nieco się różnią, ale ustalają taką samą kolejność badanych obiektów. Należą do przedziału  $[0; 1]$ . Jeśli któreś z województw byłoby dokładnie jak wzorec pozytywny  $m_i$  i  $q_i$  wyniosłyby 1, jeśli jak negatywny - 0. W związku z tym wyższy poziom miary oznacza wyższy poziom omawianego zjawiska czyli lepszą jakość życia<sup>14</sup>.

# Metoda wzorca

Ranking województw uzyskany za pomocą metody wzorca przedstawia się następująco:

Tabela 9. Ranking metody wzorca

	$m_i$	RANKING MW $m_i$	$q_i$	RANKING MW $q_i$
mazowieckie	0,562	1	0,467	1
wielkopolskie	0,454	2	0,336	2
dolnośląskie	0,439	3	0,317	3
małopolskie	0,429	4	0,305	4
pomorskie	0,422	5	0,296	5
śląskie	0,368	6	0,230	6
zachodniopomorskie	0,352	7	0,211	7
lubuskie	0,321	8	0,173	8
podlaskie	0,317	9	0,168	9
opolskie	0,308	10	0,158	10
lubelskie	0,292	11	0,137	11
łódzkie	0,292	12	0,137	12
kujawsko-pomorskie	0,279	13	0,122	13
podkarpackie	0,260	14	0,099	14
świętokrzyskie	0,259	15	0,098	15
warmińsko-mazurskie	0,249	16	0,086	16



Obsługiwane przez usługę Bing  
© GeoNames, Microsoft

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Korzystając z metody wzorca (wersja z  $m_i$ , jak i  $q_i$ ), można orzec, że w 2019 r. zdecydowanie najlepiej żyło się w województwie mazowieckim. Drugie miejsce należało do wielkopolskiego, a trzecie do dolnośląskiego. Stosunkowo wysoką jakość życia odnotowano także w przypadku województwa małopolskiego, pomorskiego oraz śląskiego. Najgorzej sytuacja wyglądała w podkarpackim, świętokrzyskim i warmińsko-mazurskim, które zajęło ostatnią lokatę.

# Metoda wzorca z wagami

Przedstawioną wcześniej metodę wzorca można zmodyfikować przez uwzględnienie wag poszczególnych zmiennych. Należy wtedy pamiętać o dwóch warunkach - 1) waga jednej zmiennej musi być większa lub równa 0; 2) suma wag wszystkich zmiennych jest równa 1.

Istnieją różne sposoby na różnicowanie wag. Przykładem jest metoda ekspercka, jednak osoby z niej korzystające powinny mieć odpowiedni zasób wiedzy, doświadczenie w danej dziedzinie. Nieumiejętnie dobrane wagi rzutowałyby na późniejsze wyniki. Zamiast tego można skorzystać z innej metody, np. dobierając wagi proporcjonalnie do stopnia informacyjności analizowanych cech (z wykorzystaniem współczynnika zmienności)<sup>15</sup>. Właśnie ona została zastosowana w dalszej części pracy.

Na początku skorzystano z macierzy po redukcji trzech zmiennych (tabela 6). Dla każdej zmiennej obliczono średnią i odchylenie standardowe, by dzięki temu otrzymać współczynnik zmienności. Pozwoliło to wyznaczyć kluczowe przy tej metodzie wagi poszczególnych zmiennych - podzielono współczynnik zmienności danej cechy przez sumę współczynników wszystkich czternastu cech. Skorzystano więc ze wzoru:

$$\alpha_j = \frac{V_j}{\sum_{j=1}^p V_j} \quad j = 1, 2, \dots, p$$

Przydatna była również macierz standaryzowana MW (tabela 7). Tym razem jej wartości pomnożono przez wagi, w konsekwencji odpowiednio przeliczone zostały też wartości wzorca i antywzorca - tabela 10.

**Tabela 10. Macierz standaryzowana metody wzorca z wagami**

Charakter zmiennej	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	-0,02	0,04	-0,05	0,06	0,02	0,03	0,12	-0,07	-0,01	0,03	0,05	0,07	0,04	0,00
kujawsko-pomorskie	-0,02	-0,02	0,09	-0,06	-0,02	-0,03	-0,06	-0,02	-0,03	-0,03	-0,03	-0,02	-0,09	-0,04
lubelskie	0,00	-0,05	0,11	-0,12	-0,01	0,03	-0,05	-0,12	0,05	-0,15	-0,03	-0,05	-0,01	0,00
lubuskie	-0,03	-0,03	-0,09	-0,04	-0,01	-0,06	0,12	-0,09	0,02	0,04	-0,08	0,06	0,03	0,07
łódzkie	-0,01	0,03	0,01	0,03	0,00	0,07	-0,11	0,44	-0,02	-0,07	0,00	-0,02	0,13	0,00
małopolskie	0,02	0,03	-0,05	0,10	0,01	0,03	-0,02	-0,11	0,01	0,03	-0,02	0,04	0,00	-0,07
mazowieckie	0,14	0,15	0,01	0,24	0,07	0,06	-0,11	0,03	-0,03	0,19	0,21	0,01	-0,07	-0,06
opolskie	-0,02	-0,03	-0,05	-0,04	-0,01	-0,06	-0,03	0,38	0,08	-0,09	-0,08	-0,02	-0,07	0,13
podkarpackie	-0,02	-0,05	0,16	-0,13	-0,02	-0,04	-0,01	-0,17	0,09	-0,15	-0,02	-0,09	-0,05	0,01
podlaskie	-0,02	-0,08	0,07	-0,09	-0,01	0,07	-0,10	-0,15	-0,02	0,12	0,10	-0,06	0,05	-0,02
pomorskie	-0,01	0,02	-0,08	0,07	0,01	0,01	0,05	-0,10	-0,06	0,05	0,04	0,01	-0,02	-0,04
śląskie	0,04	0,02	-0,11	0,07	0,02	0,01	0,03	0,08	-0,05	0,05	-0,05	0,04	0,02	-0,12
świętokrzyskie	-0,01	-0,06	0,09	-0,07	-0,02	-0,01	-0,12	0,25	0,00	-0,08	0,01	-0,03	0,06	0,05
warmińsko-mazurskie	-0,02	-0,06	0,04	-0,06	-0,02	-0,06	0,04	-0,17	-0,01	-0,03	-0,11	-0,02	0,07	0,05
wielkopolskie	-0,01	0,11	-0,14	0,03	-0,01	-0,03	0,07	-0,09	-0,02	0,11	0,02	-0,02	-0,08	-0,01
zachodniopomorskie	-0,01	-0,03	-0,01	-0,01	0,00	0,00	0,19	-0,07	-0,01	-0,02	0,00	0,11	-0,01	0,05
<b>WZORZEC</b>	0,14	0,15	-0,14	0,24	0,07	0,07	0,19	-0,17	0,09	0,19	0,21	-0,09	-0,09	0,13
<b>ANTYWZORZEC</b>	-0,03	-0,08	0,16	-0,13	-0,02	-0,06	-0,12	0,44	-0,06	-0,15	-0,11	0,11	0,13	-0,12

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

# Metoda wzorca z wagami

W tabeli 10 najbardziej korzystne wartości wyróżniono kolorem niebieskim, najmniej korzystne - czerwonym. Co prawda różnią się one od tych z macierzy bez wag, ale póki co województwa w ramach poszczególnych zmiennych zajmują takie same lokaty jak przy MW.

W ostatniej macierzy potrzebnej do metody z wagami (tabela 11) wykonano analogiczne działania, jak w macierzy z odległościami MW (tabela 8). Wyznaczono zatem odległości poszczególnych wartości zmiennych (tym razem z wagami) od tych najlepszych, wzorcowych.

**Tabela 11. Macierz z odległościami metody wzorca z wagami**

Charakter zmiennej	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	0,02	0,01	0,01	0,03	0,00	0,00	0,01	0,01	0,01	0,03	0,03	0,03	0,02	0,02
kujawsko-pomorskie	0,03	0,03	0,05	0,09	0,01	0,01	0,06	0,02	0,01	0,05	0,06	0,00	0,00	0,03
lubelskie	0,02	0,04	0,06	0,13	0,01	0,00	0,06	0,00	0,00	0,11	0,06	0,00	0,01	0,02
lubuskie	0,03	0,03	0,00	0,08	0,01	0,02	0,01	0,01	0,00	0,02	0,08	0,02	0,01	0,00
łódzkie	0,02	0,02	0,02	0,05	0,00	0,00	0,09	0,38	0,01	0,06	0,04	0,01	0,05	0,02
małopolskie	0,01	0,01	0,01	0,02	0,00	0,00	0,05	0,00	0,01	0,03	0,05	0,02	0,01	0,04
mazowieckie	0,00	0,00	0,02	0,00	0,00	0,00	0,10	0,04	0,02	0,00	0,00	0,01	0,00	0,04
opolskie	0,02	0,03	0,01	0,08	0,01	0,02	0,05	0,30	0,00	0,08	0,08	0,00	0,00	0,00
podkarpackie	0,02	0,04	0,09	0,14	0,01	0,01	0,04	0,00	0,00	0,12	0,05	0,00	0,00	0,01
podlaskie	0,02	0,05	0,04	0,11	0,01	0,00	0,09	0,00	0,01	0,00	0,01	0,00	0,02	0,02
pomorskie	0,02	0,02	0,00	0,03	0,00	0,00	0,02	0,00	0,02	0,02	0,03	0,01	0,00	0,03
śląskie	0,01	0,02	0,00	0,03	0,00	0,00	0,03	0,07	0,02	0,02	0,06	0,02	0,01	0,06
świętokrzyskie	0,02	0,05	0,05	0,10	0,01	0,01	0,10	0,18	0,01	0,07	0,04	0,00	0,02	0,01
warmińsko-mazurskie	0,03	0,04	0,03	0,09	0,01	0,02	0,02	0,00	0,01	0,05	0,10	0,00	0,03	0,01
wielkopolskie	0,02	0,00	0,00	0,04	0,01	0,01	0,02	0,01	0,01	0,01	0,04	0,00	0,00	0,02
zachodniopomorskie	0,02	0,03	0,02	0,06	0,00	0,00	0,00	0,01	0,01	0,04	0,04	0,04	0,01	0,01

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

# Metoda wzorca z wagami

Aby wyliczyć miary  $m_i$  oraz  $q_i$  skorzystano z tych samych wzorów, co w przypadku MW bez wag.

Obserwując poniżej zamieszczony ranking (tabela 12), można jednak zauważyć, że po uzupełnieniu metody wzorca wagami kolejność województw uległa pewnej zmianie.

Tabela 12. Ranking metody wzorca z wagami

	$m_i$	RANKING MW WAGI $m_i$	$q_i$	RANKING MW WAGI $q_i$
wielkopolskie	0,609	1	0,528	1
pomorskie	0,573	2	0,485	2
dolnośląskie	0,573	3	0,485	3
mazowieckie	0,571	4	0,482	4
małopolskie	0,534	5	0,438	5
zachodniopomorskie	0,493	6	0,388	6
lubuskie	0,476	7	0,368	7
śląskie	0,457	8	0,345	8
podlaskie	0,426	9	0,308	9
warmińsko-mazurskie	0,395	10	0,270	10
kujawsko-pomorskie	0,385	11	0,258	11
lubelskie	0,339	12	0,203	12
podkarpackie	0,332	13	0,194	13
świętokrzyskie	0,258	14	0,105	14
opolskie	0,244	15	0,088	15
łódzkie	0,199	16	0,033	16



Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Według tego rankingu województwem, w którym w 2019 r. jakość życia była najwyższa, było woj. wielkopolskie (drugie miejsce w rankingu MW). Kolejna lokata należała do pomorskiego, które bez uwzględnienia wag zajmowało dopiero piąte miejsce. Na podium znalazło się jeszcze dolnośląskie (również trzecie miejsce w rankingu MW).

W tym zestawieniu woj. mazowieckie zajęło dopiero czwartą lokatę, mimo że przy metodzie wzorca bez wag zdecydowanie przodowało -  $m_i$  i  $q_i$  były tam wyraźnie większe od miar w reszcie województw.

Także patrząc na ostatnie wiersze w rankingu, zauważa się zmiany. Najgorszym województwem pod względem jakości życia nie było już warmińsko-mazurskie, zajęło ono bowiem aż dziesiąte miejsce. Najsłabsze okazało się woj. łódzkie (dwunaste miejsce w rankingu MW), a zaraz potem opolskie (dziesiąte miejsce w rankingu MW).

Można zatem stwierdzić, że w przypadku części województw dodanie wag wpłynęło dość znacznie na zajmowaną pozycję.



# Metoda TOPSIS

Analizę metodą TOPSIS rozpoczynamy od wyliczenia pierwiastków wartości z macierzy właściwej (tabela 6), następnie obliczamy sumy dla każdej zmiennej. Kolejnym krokiem jest obliczenie ilorazu wartości z macierzy właściwej (tabela 6) i pierwiastka wcześniej obliczonej sumy.

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$$

Dodatkowo należy obliczyć wektor wzorca oraz wektor antywzorca, postępując tak jak przy metodzie wzorca. Współrzędne wzorca:

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

Współrzędne antywzorca:

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

W ten sposób powstaje nowa macierz, zaprezentowana poniżej.

**Tabela 13. Macierz dla metody TOPSIS**

	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	0,23	0,28	0,18	0,30	0,27	0,28	0,36	0,13	0,23	0,26	0,29	0,32	0,28	0,24
kujawsko-pomorskie	0,22	0,22	0,33	0,17	0,23	0,21	0,17	0,17	0,22	0,20	0,21	0,22	0,14	0,20
lubelskie	0,25	0,19	0,35	0,11	0,23	0,28	0,18	0,09	0,30	0,08	0,20	0,19	0,24	0,24
lubuskie	0,22	0,21	0,14	0,19	0,23	0,18	0,36	0,12	0,27	0,28	0,15	0,31	0,27	0,32
łódzkie	0,24	0,27	0,24	0,26	0,25	0,32	0,12	0,57	0,23	0,16	0,24	0,22	0,38	0,25
małopolskie	0,26	0,28	0,18	0,33	0,26	0,27	0,21	0,10	0,26	0,26	0,22	0,28	0,24	0,17
mazowieckie	0,40	0,40	0,24	0,48	0,32	0,31	0,11	0,22	0,21	0,42	0,46	0,26	0,16	0,18
opolskie	0,23	0,21	0,19	0,19	0,24	0,18	0,20	0,51	0,34	0,14	0,15	0,22	0,16	0,38
podkarpackie	0,23	0,19	0,40	0,10	0,23	0,20	0,22	0,05	0,34	0,08	0,21	0,15	0,19	0,26
podlaskie	0,23	0,16	0,31	0,14	0,24	0,32	0,12	0,07	0,23	0,36	0,34	0,18	0,29	0,22
pomorskie	0,24	0,26	0,15	0,30	0,26	0,25	0,29	0,10	0,18	0,28	0,28	0,25	0,22	0,20
śląskie	0,29	0,27	0,12	0,30	0,27	0,26	0,26	0,26	0,19	0,28	0,19	0,29	0,27	0,11
świętokrzyskie	0,23	0,18	0,33	0,16	0,23	0,23	0,11	0,40	0,25	0,14	0,25	0,21	0,31	0,30
warmińsko-mazurskie	0,22	0,18	0,28	0,17	0,22	0,18	0,27	0,04	0,23	0,20	0,12	0,22	0,32	0,30
wielkopolskie	0,24	0,36	0,09	0,27	0,24	0,22	0,30	0,11	0,22	0,35	0,25	0,22	0,16	0,23
zachodniopomorskie	0,23	0,21	0,22	0,22	0,25	0,24	0,43	0,14	0,24	0,21	0,24	0,37	0,23	0,29
WZORZEC	0,40	0,40	0,09	0,48	0,32	0,32	0,43	0,04	0,34	0,42	0,46	0,15	0,14	0,38
ANTYWZORZEC	0,22	0,16	0,40	0,10	0,22	0,18	0,11	0,57	0,18	0,08	0,12	0,37	0,38	0,11

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

# Metoda TOPSIS

Następnie należy wykonać zestawienie z wzorcem oraz z antywzorcem. W rezultacie powstają dwie nowe macierze - pierwsza powstała poprzez podniesienie do kwadratu różnicy wartości i wzorca z tabeli 13. Druga przez podniesienie do kwadratu różnicy wartości i antywzorca z tabeli 13.

**Tabela 14. Zestawienie z wzorcem**

	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	0,03	0,01	0,01	0,03	0,00	0,00	0,01	0,01	0,01	0,03	0,03	0,03	0,02	0,02
kujawsko-pomorskie	0,03	0,03	0,05	0,10	0,01	0,01	0,07	0,02	0,02	0,05	0,06	0,01	0,00	0,03
lubelskie	0,02	0,05	0,07	0,14	0,01	0,00	0,06	0,00	0,00	0,12	0,06	0,00	0,01	0,02
lubuskie	0,03	0,04	0,00	0,08	0,01	0,02	0,01	0,01	0,01	0,02	0,09	0,03	0,02	0,00
łódzkie	0,03	0,02	0,02	0,05	0,01	0,00	0,10	0,28	0,01	0,07	0,05	0,01	0,06	0,02
małopolskie	0,02	0,02	0,01	0,02	0,00	0,00	0,05	0,00	0,01	0,03	0,06	0,02	0,01	0,05
mazowieckie	0,00	0,00	0,02	0,00	0,00	0,00	0,10	0,03	0,02	0,00	0,00	0,01	0,00	0,04
opolskie	0,03	0,04	0,01	0,08	0,01	0,02	0,05	0,22	0,00	0,08	0,09	0,01	0,00	0,00
podkarpackie	0,03	0,05	0,09	0,14	0,01	0,01	0,04	0,00	0,00	0,12	0,06	0,00	0,00	0,02
podlaskie	0,03	0,06	0,05	0,11	0,01	0,00	0,09	0,00	0,01	0,00	0,01	0,00	0,02	0,03
pomorskie	0,03	0,02	0,00	0,03	0,00	0,00	0,02	0,00	0,03	0,02	0,03	0,01	0,01	0,03
śląskie	0,01	0,02	0,00	0,03	0,00	0,00	0,03	0,05	0,02	0,02	0,07	0,02	0,02	0,07
świętokrzyskie	0,03	0,05	0,05	0,10	0,01	0,01	0,10	0,13	0,01	0,08	0,04	0,00	0,03	0,01
warmińsko-mazurskie	0,03	0,05	0,03	0,09	0,01	0,02	0,03	0,00	0,01	0,05	0,11	0,00	0,03	0,01
wielkopolskie	0,03	0,00	0,00	0,05	0,01	0,01	0,02	0,00	0,02	0,01	0,04	0,00	0,00	0,02
zachodniopomorskie	0,03	0,04	0,02	0,07	0,01	0,01	0,00	0,01	0,01	0,05	0,05	0,05	0,01	0,01

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

**Tabela 15. Zestawienie z antywzorcem**

	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	0,00	0,01	0,05	0,04	0,00	0,01	0,06	0,19	0,00	0,03	0,03	0,00	0,01	0,02
kujawsko-pomorskie	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,16	0,00	0,01	0,01	0,02	0,06	0,01
lubelskie	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,23	0,01	0,00	0,01	0,03	0,02	0,02
lubuskie	0,00	0,00	0,07	0,01	0,00	0,00	0,06	0,20	0,01	0,04	0,00	0,00	0,01	0,04
łódzkie	0,00	0,01	0,02	0,03	0,00	0,02	0,00	0,00	0,00	0,01	0,01	0,02	0,00	0,02
małopolskie	0,00	0,01	0,05	0,05	0,00	0,01	0,01	0,22	0,01	0,03	0,01	0,01	0,02	0,00
mazowieckie	0,03	0,06	0,03	0,14	0,01	0,02	0,00	0,12	0,00	0,12	0,11	0,01	0,05	0,00
opolskie	0,00	0,00	0,05	0,01	0,00	0,00	0,01	0,00	0,03	0,00	0,00	0,02	0,05	0,07
podkarpackie	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,27	0,03	0,00	0,01	0,05	0,04	0,02
podlaskie	0,00	0,00	0,01	0,00	0,00	0,02	0,00	0,25	0,00	0,08	0,05	0,04	0,01	0,01
pomorskie	0,00	0,01	0,06	0,04	0,00	0,01	0,03	0,22	0,00	0,04	0,03	0,01	0,03	0,01
śląskie	0,00	0,01	0,08	0,04	0,00	0,01	0,02	0,09	0,00	0,04	0,00	0,01	0,01	0,00
świętokrzyskie	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,01	0,03	0,01	0,04
warmińsko-mazurskie	0,00	0,00	0,01	0,01	0,00	0,00	0,02	0,28	0,00	0,01	0,00	0,02	0,00	0,03
wielkopolskie	0,00	0,04	0,09	0,03	0,00	0,00	0,04	0,21	0,00	0,07	0,02	0,02	0,05	0,02
zachodniopomorskie	0,00	0,00	0,03	0,01	0,00	0,00	0,10	0,19	0,00	0,02	0,01	0,00	0,02	0,03

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

# Metoda TOPSIS

Dla powstałych zestawień obliczono sumy dla każdego województwa. Następnie obliczono odległości obiektów od wzorca czyli zastosowano poniższy wzór:

$$d_{i0}^+ = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j}^+)^2} \quad i = 1, 2, \dots, n$$

Czynność powtórzono dla antywzorca:

$$d_{i0}^- = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j}^-)^2} \quad i = 1, 2, \dots, n$$

**Tabela 16. Wyniki dla wzorca**

	SUMA	$d_{i0}^+$
dolnośląskie	0,24	0,49
kujawsko-pomorskie	0,48	0,70
lubelskie	0,56	0,75
lubuskie	0,36	0,60
łódzkie	0,70	0,84
małopolskie	0,29	0,53
mazowieckie	0,22	0,47
opolskie	0,63	0,80
podkarpackie	0,58	0,76
podlaskie	0,43	0,65
pomorskie	0,24	0,49
śląskie	0,37	0,61
świętokrzyskie	0,65	0,81
warmińsko-mazurskie	0,48	0,69
wielkopolskie	0,20	0,45
zachodniopomorskie	0,34	0,58

**Tabela 17. Wyniki dla antywzorca**

	SUMA	$d_{i0}^-$
dolnośląskie	0,46	0,68
kujawsko-pomorskie	0,28	0,53
lubelskie	0,34	0,59
lubuskie	0,45	0,67
łódzkie	0,14	0,38
małopolskie	0,44	0,66
mazowieckie	0,70	0,84
opolskie	0,24	0,49
podkarpackie	0,42	0,65
podlaskie	0,47	0,68
pomorskie	0,48	0,69
śląskie	0,33	0,57
świętokrzyskie	0,13	0,36
warmińsko-mazurskie	0,40	0,63
wielkopolskie	0,58	0,76
zachodniopomorskie	0,43	0,66

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

# Metoda TOPSIS

Ostatnim krokiem jest wyliczenie wartości zmiennej agregatywnej  $q_i$ , którą otrzymujemy przez obliczenie ilorazu wartości  $d_{i0}^-$  przez sumę  $d_{i0}^-$  i  $d_{i0}^+$ .

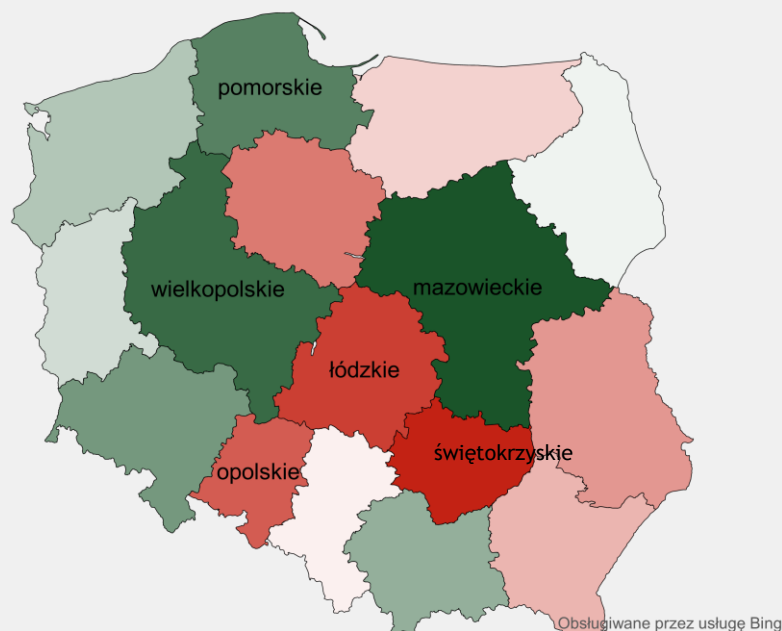
$$q_i = \frac{d_{i0}^-}{d_{i0}^- + d_{i0}^+}, \quad q_i = \llbracket 0; 1 \rrbracket$$

Na podstawie otrzymanej zmiennej tworzony jest ranking dla metody TOPSIS.

**Tabela 16. Ranking dla metody TOPSIS**

	$q_i$	Ranking TOPSIS
mazowieckie	0,639	1
wielkopolskie	0,630	2
pomorskie	0,586	3
dolnośląskie	0,580	4
małopolskie	0,554	5
zachodniopomorskie	0,532	6
lubuskie	0,527	7
podlaskie	0,511	8
śląskie	0,485	9
warmińsko-mazurskie	0,477	10
podkarpackie	0,462	11
lubelskie	0,439	12
kujawsko-pomorskie	0,433	13
opolskie	0,380	14
łódzkie	0,313	15
świętokrzyskie	0,309	16

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.



Według rankingu metodą TOPSIS, województwem z najlepszą jakością życia w 2019 r. było województwo mazowieckie. Na drugim i trzecim miejscu w rankingu były województwa wielkopolskie i pomorskie. Dobrą jakością życia mogą się również pochwalić województwa - dolnośląskie, małopolskie oraz zachodniopomorskie. Najgorzej w rankingu wypadły natomiast województwa - opolskie, łódzkie, świętokrzyskie.

# Metoda standaryzowanych sum

Metoda sum standaryzowanych polega na wyznaczeniu liniowej funkcji porządkującej  $W_i$ , która jest średnią arytmetyczną znormalizowanych wartości cech. Analizę rozpoczynamy od standaryzacji zmiennych, na których będziemy pracować - tabela 7. Następnie dla każdego województwa obliczamy sumę wartości zmiennych. Korzystając z wzoru:

$$w_i = \frac{1}{p} \sum_{j=1}^p z_{ij}$$

wyliczamy współczynnik  $w_i$ . Następnie współczynnik  $\widehat{w}_i$  według wzoru:

$$\widehat{w}_i = w_i - \min(w_i)$$

Ostatnim krokiem jest obliczenie naszej funkcji  $W_i$ . Wzór na nią jest następujący:

$$W_i = \frac{\widehat{w}_i}{\max(\widehat{w}_i)} = \frac{w_i - \min(w_i)}{\max(w_i - \min(w_i))}$$

**Tabela 17. Obliczenia do metody standaryzowanych sum**

	Suma	$w_i$	$\widehat{w}_i$	$W_i$
dolnośląskie	3,51	0,25	0,75	0,47
kujawsko-pomorskie	-4,74	-0,34	0,16	0,10
lubelskie	-2,99	-0,21	0,29	0,18
lubuskie	-1,19	-0,08	0,42	0,26
łódzkie	-4,58	-0,33	0,18	0,11
małopolskie	2,61	0,19	0,69	0,43
mazowieckie	15,63	1,12	1,62	1,00
opolskie	-1,09	-0,08	0,42	0,26
podkarpackie	-2,64	-0,19	0,31	0,19
podlaskie	-0,52	-0,04	0,47	0,29
pomorskie	2,88	0,21	0,71	0,44
śląskie	-0,26	-0,02	0,48	0,30
świętokrzyskie	-7,04	-0,50	0,00	0,00
warmińsko-mazurskie	-5,72	-0,41	0,09	0,06
wielkopolskie	6,29	0,45	0,95	0,59
zachodniopomorskie	-0,16	-0,01	0,49	0,30
	min	-0,50		
		max	1,62	

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.



# Metoda standaryzowanych sum

Na podstawie obliczonego współczynnika powstaje ranking dla metody standaryzowanych sum.

Tabela 18. Ranking dla metody standaryzowanych sum

	Ranking MSS
mazowieckie	1
wielkopolskie	2
dolnośląskie	3
pomorskie	4
małopolskie	5
zachodniopomorskie	6
śląskie	7
podlaskie	8
opolskie	9
lubuskie	10
podkarpackie	11
lubelskie	12
łódzkie	13
kujawsko-pomorskie	14
warmińsko-mazurskie	15
świętokrzyskie	16



Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Analizując dane metodą standaryzowanych sum, wśród województw z najlepszą jakością życia znalazły się: mazowieckie, wielkopolskie i dolnośląskie. Najgorszy wynik uzyskały województwa: kujawsko-pomorskie, warmińsko-mazurskie oraz świętokrzyskie.

# Metoda sumy rang

Metoda sumy rang polega na ustaleniu rang dla poszczególnych zmiennych. Dla stymulant najwyższe rangi są ustalane dla województw o największych wartościach, zaś dla destymulant najwyższą rangę otrzymują wartości najniższe. Na tej podstawie zostaje stworzona macierz z rangami, zaprezentowana poniżej. Ustala się zatem lokatę jednostki w zbiorze jednostek i przypisuje jej odpowiednią liczbę naturalną ( $l_{ij}$ ).

Tabela 19. Macierz z rangami

	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	12	3	6	5	2	4	3	9	8	8	3	15	12	9
kujawsko-pomorskie	15	8	13	12	13	12	12	11	13	11	11	8	1	13
lubelskie	4	12	15	15	11	5	11	4	3	15	12	3	8	8
lubuskie	16	9	3	10	12	14	2	8	4	6	15	14	11	2
łódzkie	6,5	5	10	7	6	2	14	16	11	12	8	9	16	7
małopolskie	3	4	5	2	5	6	9	5	5	7	9	12	9	15
mazowieckie	1	1	9	1	1	3	15	12	14	1	1	11	4	14
opolskie	13	10	7	9	8	16	10	15	2	14	14	7	3	1
podkarpackie	10	13	16	16	15	13	8	2	1	16	10	1	5	6
podlaskie	11	16	12	14	10	1	13	3	10	2	2	2	13	11
pomorskie	5	7	4	4	4	8	5	6	16	5	4	10	6	12
śląskie	2	6	2	3	3	7	7	13	15	4	13	13	10	16
świętokrzyskie	9	15	14	13	14	10	16	14	6	13	6	4	14	3
warmińsko-mazurskie	14	14	11	11	16	15	6	1	9	10	16	5,5	15	4
wielkopolskie	6,5	2	1	6	9	11	4	7	12	3	5	5,5	2	10
zachodniopomorskie	8	11	8	8	7	9	1	10	7	9	7	16	7	5

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Aby nadać analizie statystycznej walor analizy wymiarowej należy dokonać sumowania rang dla obiektów.

$$s_i = \sum_{j=1}^p l_{ij}, i = 1, 2, \dots, n$$

# Metoda sumy rang

Na podstawie obliczonej sumy powstaje ranking dla metody sumy rang.

Tabela 20. Ranking dla metody sumy rang

	SUMA RANG	Ranking MSR
wielkopolskie	84	1
mazowieckie	88	2
małopolskie	96	3
pomorskie	96	4
dolnośląskie	99	5
zachodniopomorskie	113	6
śląskie	114	7
podlaskie	120	8
lubelskie	126	9
lubuskie	126	10
opolskie	129	11
łódzkie	129,5	12
podkarpackie	132	13
warmińsko-mazurskie	147,5	14
świętokrzyskie	151	15
kujawsko-pomorskie	153	16

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.



Według metody sumy rang województwami z najlepszą jakością życia w 2019 r. są województwa wielkopolskie, mazowieckie i małopolskie. Trochę niżej, ale nadal na wysokim miejscu, znalazły się województwa pomorskie, dolnośląskie oraz zachodniopomorskie. Miano województw z najgorszą jakością życia otrzymały województwa warmińsko-mazurskie, świętokrzyskie oraz kujawsko-pomorskie.

# Metoda dystansów

Do wyliczenia rankingu metodą dystansów konieczne jest wyznaczenie wartości maksymalnych - dla stymulant, oraz wartości najmniejszych - dla destymulant. Następnie tworzona jest nowa macierz, w której dla stymulant miarą dystansu jest iloraz danej wartości zmiennej i obliczonego maksimum. W przypadku destymulant obliczone wcześniej minimum jest dzielone przez wartość zmiennej. Według poniższych wzorów:

$$w_i = \frac{x_{ij}}{\max(x_{ij})}, \text{ dla stymulant}$$

$$w_i = \frac{\min(x_{ij})}{x_{ij}}, \text{ dla destymulant}$$

Wyniki zostały zaprezentowane w tabeli 21.

**Tabela 21. Macierz z obliczeniami dla metody dystansów**

	S	S	D	S	S	S	S	D	S	S	S	D	D	S
	x2	x3	x4	x5	x6	x8	x9	x10	x11	x12	x14	x15	x16	x17
dolnośląskie	0,57	0,70	0,52	0,62	0,85	0,87	0,82	0,34	0,68	0,61	0,63	0,46	0,51	0,63
kujawsko-pomorskie	0,56	0,54	0,29	0,36	0,72	0,65	0,41	0,26	0,63	0,46	0,46	0,66	1,00	0,52
lubelskie	0,62	0,47	0,26	0,22	0,73	0,87	0,42	0,51	0,87	0,19	0,45	0,77	0,61	0,64
lubuskie	0,55	0,53	0,67	0,39	0,73	0,56	0,83	0,37	0,79	0,65	0,33	0,48	0,53	0,83
łódzkie	0,59	0,67	0,38	0,55	0,77	1,00	0,28	0,08	0,66	0,39	0,52	0,66	0,38	0,65
małopolskie	0,67	0,69	0,52	0,69	0,82	0,85	0,48	0,46	0,75	0,61	0,48	0,52	0,61	0,44
mazowieckie	1,00	1,00	0,39	1,00	1,00	0,98	0,27	0,20	0,61	1,00	1,00	0,58	0,88	0,46
opolskie	0,57	0,52	0,50	0,40	0,75	0,55	0,46	0,09	0,98	0,34	0,34	0,67	0,89	1,00
podkarpackie	0,57	0,47	0,23	0,21	0,70	0,62	0,52	0,89	1,00	0,18	0,47	1,00	0,75	0,68
podlaskie	0,57	0,40	0,30	0,30	0,73	1,00	0,29	0,67	0,66	0,84	0,75	0,83	0,49	0,57
pomorskie	0,60	0,65	0,63	0,62	0,82	0,78	0,66	0,42	0,52	0,66	0,62	0,59	0,67	0,52
śląskie	0,72	0,66	0,78	0,63	0,83	0,81	0,60	0,17	0,56	0,67	0,42	0,51	0,54	0,29
świętokrzyskie	0,59	0,44	0,28	0,34	0,72	0,72	0,26	0,11	0,72	0,34	0,54	0,71	0,47	0,79
warmińsko-mazurskie	0,56	0,44	0,33	0,36	0,69	0,55	0,62	1,00	0,68	0,47	0,27	0,68	0,45	0,78
wielkopolskie	0,60	0,88	1,00	0,56	0,75	0,68	0,70	0,39	0,64	0,81	0,56	0,68	0,90	0,62
zachodniopomorskie	0,59	0,51	0,41	0,46	0,76	0,76	1,00	0,33	0,69	0,50	0,52	0,40	0,62	0,76

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Następnie dla każdego województwa obliczana jest suma wartości zmiennych. Ostatnim krokiem jest podzielenie otrzymanych sum przez liczbę zmiennych. W ten sposób zostaje obliczony współczynnik  $W_i$ :

$$W_i = \frac{1}{p} \sum_{i=1}^p w_{ij}, i = 1, 2, \dots, n$$

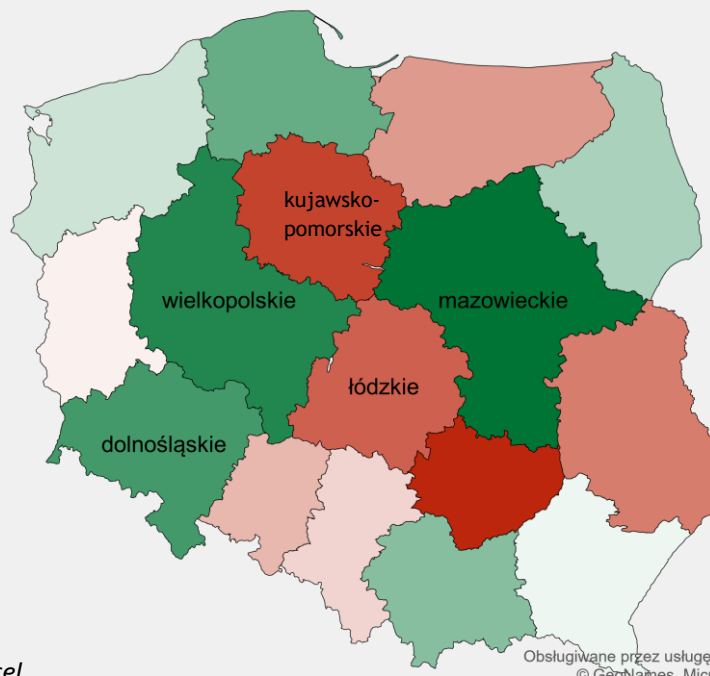
# Metoda dystansów

Na podstawie obliczonego współczynnika powstaje tabela z rankingiem, zaprezentowana poniżej.

Tabela 22. Ranking dla metody dystansów

	SUMA	W <sub>i</sub>	Ranking MD
mazowieckie	10,36	0,740	1
wielkopolskie	9,756	0,697	2
dolnośląskie	8,813	0,630	3
pomorskie	8,77	0,626	4
małopolskie	8,577	0,613	5
podlaskie	8,416	0,601	6
zachodniopomorskie	8,336	0,595	7
podkarpackie	8,299	0,593	8
lubuskie	8,237	0,588	9
śląskie	8,187	0,585	10
opolskie	8,054	0,575	11
warmińsko-mazurskie	7,903	0,565	12
lubelskie	7,644	0,546	13
łódzkie	7,565	0,540	14
kujawsko-pomorskie	7,513	0,537	15
świętokrzyskie	7,017	0,501	16

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.



Korzystając z metody dystansów można stwierdzić, że w 2019 r. najlepiej żyło się w województwie mazowieckim. Drugie miejsce należało do wielkopolskiego, a trzecie do dolnośląskiego. Wysoką jakość życia odnotowano także w województwach pomorskim, małopolskim oraz śląskim. Najgorzej wypadły województwa łódzkie, kujawsko-pomorskie i świętokrzyskie.

# Podsumowanie metod porządkowania liniowego

Podsumowując przedstawione metody, została zbudowana wspólna tabela uwzględniająca wszystkie 8 metod porządkowania liniowego.

Tabela 23. Zestawienie rankingów

	MW ( $m_i$ )	MW ( $q_i$ )	TOPSIS	MW WAGI ( $m_i$ )	MW WAGI ( $q_i$ )	MSS	MSR	MD
mazowieckie	1	1	1	4	4	1	2	1
wielkopolskie	2	2	2	1	1	2	1	2
dolnośląskie	3	3	4	3	3	3	5	3
małopolskie	4	4	5	5	5	5	3	5
pomorskie	5	5	3	2	2	4	4	4
śląskie	6	6	9	8	8	7	7	10
zachodniopomorskie	7	7	6	6	6	6	6	7
lubuskie	8	8	7	7	7	10	10	9
podlaskie	9	9	8	9	9	8	8	6
opolskie	10	10	14	15	15	9	11	11
lubelskie	11	11	12	12	12	12	9	13
łódzkie	12	12	15	16	16	13	12	14
kujawsko-pomorskie	13	13	13	11	11	14	16	15
podkarpackie	14	14	11	13	13	11	13	8
świętokrzyskie	15	15	16	14	14	16	15	16
warmińsko-mazurskie	16	16	10	10	10	15	14	12

Źródło: opracowanie własne na podstawie obliczeń w programie MS Excel.

Można zauważyć, że pomiędzy niektórymi metodami występują podobieństwa, jak i różnice. W tabeli 23 kolorem zielonym zaznaczono województwa, które znalazły się na podium, zaś kolorem czerwonym - trzy ostatnie miejsca. W ten sposób dobrze widać podobieństwo w metodach wzorca  $m_i$  i  $q_i$  oraz metodach wzorca z wagami. Zauważa się również, że rankingi dla metod wzorca z wagami najbardziej odbiegają od reszty rankingów.

Na podstawie zestawienia powyższych metod porządkowania liniowego można orzec, że w 2019 r. do województw z najwyższą jakością życia należały:





# Podsumowanie metod porządkowania liniowego

Tabela 24. Macierz korelacji metod porządkowania liniowego

	MW ( $m_i$ )	MW ( $q_i$ )	TOPSIS	MW WAGI ( $m_i$ )	MW WAGI ( $q_i$ )	MSS	MSR	MD
MW ( $m_i$ )	1,00	<b>1,00</b>	0,87	0,84	0,84	<b>0,96</b>	<b>0,95</b>	0,86
MW ( $q_i$ )	<b>1,00</b>	1,00	0,87	0,84	0,84	<b>0,96</b>	<b>0,95</b>	0,86
TOPSIS	0,87	0,87	1,00	<b>0,96</b>	<b>0,96</b>	0,90	0,89	0,94
MW WAGI ( $m_i$ )	0,84	0,84	<b>0,96</b>	1,00	<b>1,00</b>	0,84	0,84	0,85
MW WAGI ( $q_i$ )	0,84	0,84	<b>0,96</b>	<b>1,00</b>	1,00	0,84	0,84	0,85
MSS	<b>0,96</b>	<b>0,96</b>	0,90	0,84	0,84	1,00	<b>0,95</b>	0,94
MSR	<b>0,95</b>	<b>0,95</b>	0,89	0,84	0,84	<b>0,95</b>	1,00	0,89
MD	0,86	0,86	0,94	0,85	0,85	0,94	0,89	1,00

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

W celu najlepszej oceny zależności między metodami porządkowania liniowego sporządzono macierz korelacji - tabela 24. Współczynnik korelacji rang Spearmana ma wysokie wartości dla wszystkich metod. Jak wcześniej zaznaczono, identyczne wyniki uzyskano standardową metodą wzorca (z miarą  $m_i$ ) i metodą wzorca z artykułu A. Bąka (z  $q_i$ ). Analogicznie wyglądała zależność pomiędzy MW z wagami z  $m_i$  a MW z wagami z  $q_i$ .

Współczynnik korelacji Spearmana był wysoki również w przypadku porównania MW z MSS oraz TOPSIS z MW WAGI (0,96). Mocno skorelowane okazały się także MW z MSR i MSS z MSR (0,95).

Najbardziej różniły się te metody, dla których współczynnik korelacji był najniższy. W tym projekcie najniższa wartość omawianego współczynnika wyniosła 0,84 - czyli pomimo pewnych różnic nadal jest to stosunkowo wysoki poziom zależności. Spośród rozważanych metod taki wynik odnotowano w przypadku zestawienia metody wzorca z wagami z:

- metodą wzorca (bez wag),
- metodą standaryzowanych sum,
- metodą sumy rang.

# ANALIZA GŁÓWNYCH SKŁADOWYCH

## Wstęp

Analiza PCA polega na ortogonalnej transformacji układu badanych zmiennych  $X_j$  w zbiór nowych zmiennych, które są ich liniowymi kombinacjami. Podział przebiega w taki sposób, by pierwsza składowa wyjaśniała możliwie jak największą część zmienności danych.

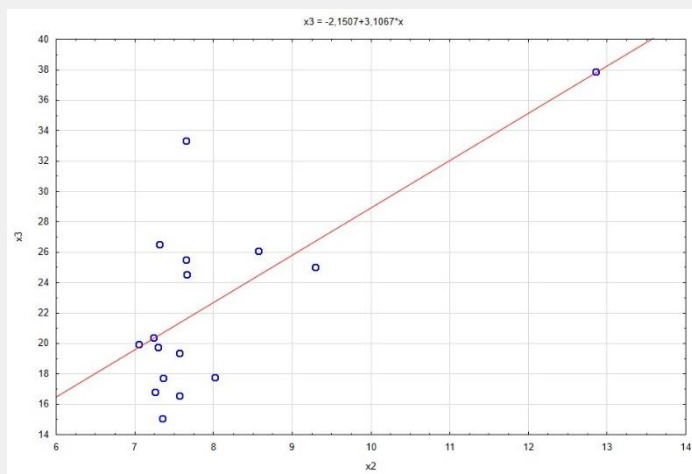
## Analiza korelacji z wykresami rozrzutu

Tabela 25. Macierz korelacji

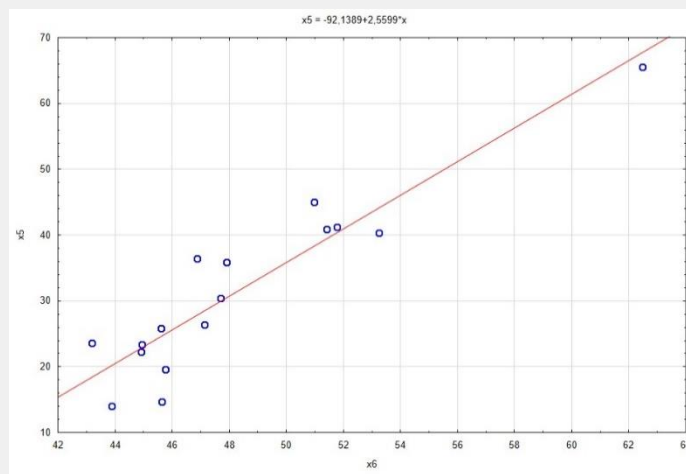
Zmienna	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17
x1	1	0,79	0,99	-0,5	0,91	0,85	0,11	0,4	-0,06	0,09	-0,42	0,65	0,23	0,6	0,27	-0,33	-0,51
x2	0,79	1	0,7	-0,11	0,77	0,85	-0,16	0,48	-0,33	0,06	-0,29	0,53	0,33	0,66	0,13	-0,26	-0,52
x3	0,99	0,7	1	-0,55	0,89	0,8	0,16	0,36	0	0,09	-0,43	0,64	0,2	0,56	0,29	-0,33	-0,48
x4	-0,5	-0,11	-0,55	1	-0,58	-0,39	-0,38	0,05	-0,51	-0,08	0,42	-0,57	0,04	0,04	-0,62	0,07	0,16
x5	0,91	0,77	0,89	-0,58	1	0,93	-0,01	0,45	0,01	0,13	-0,56	0,71	0,19	0,58	0,49	-0,15	-0,52
x6	0,85	0,85	0,8	-0,39	0,93	1	-0,12	0,56	-0,07	0,08	-0,42	0,63	0,24	0,72	0,42	-0,21	-0,52
x7	0,11	-0,16	0,16	-0,38	-0,01	-0,12	1	-0,42	0,58	-0,42	0,08	0,23	-0,03	-0,05	0,23	-0,58	0,15
x8	0,4	0,48	0,36	0,05	0,45	0,56	-0,42	1	-0,39	0,09	-0,42	0,37	0,44	0,72	0,06	0,3	-0,57
x9	-0,06	-0,33	0	-0,51	0,01	-0,07	0,58	-0,39	1	-0,42	-0,06	0,07	-0,53	-0,32	0,69	-0,08	0,17
x10	0,09	0,06	0,09	-0,08	0,13	0,08	-0,42	0,09	-0,42	1	0,07	-0,23	-0,04	-0,06	-0,05	0,23	0,27
x11	-0,42	-0,29	-0,43	0,42	-0,56	-0,42	0,08	-0,42	-0,06	0,07	1	-0,66	0,17	-0,41	-0,36	-0,18	0,61
x12	0,65	0,53	0,64	-0,57	0,71	0,63	0,23	0,37	0,07	-0,23	-0,66	1	0,3	0,63	0,34	-0,14	-0,47
x13	0,23	0,33	0,2	0,04	0,19	0,24	-0,03	0,44	-0,53	-0,04	0,17	0,3	1	0,38	-0,31	-0,08	-0,34
x14	0,6	0,66	0,56	0,04	0,58	0,72	-0,05	0,72	-0,32	-0,06	-0,41	0,63	0,38	1	0,01	-0,17	-0,43
x15	0,27	0,13	0,29	-0,62	0,49	0,42	0,23	0,06	0,69	-0,05	-0,36	0,34	-0,31	0,01	1	0,08	-0,1
x16	-0,33	-0,26	-0,33	0,07	-0,15	-0,21	-0,58	0,3	-0,08	0,23	-0,18	-0,14	-0,08	-0,17	0,08	1	0,09
x17	-0,51	-0,52	-0,48	0,16	-0,52	-0,52	0,15	-0,57	0,17	0,27	0,61	-0,47	-0,34	-0,43	-0,1	0,09	1

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

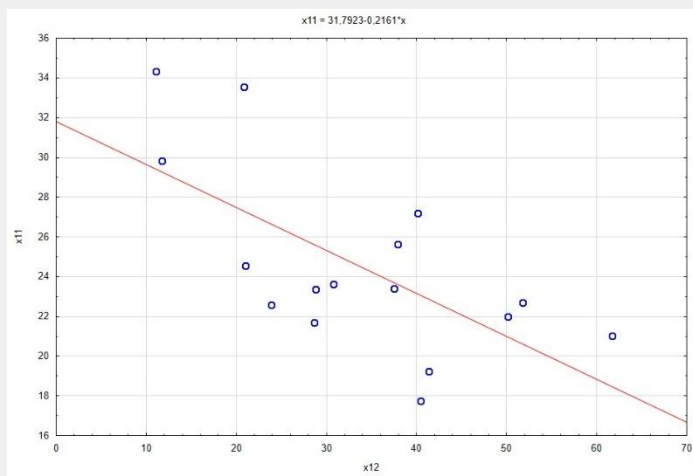
Powyżej ponownie prezentujemy macierz korelacji. Dla zmiennych o dodatnim ( $x_3$  względem  $x_2$ ,  $x_5$  względem  $x_6$ ), ujemnym ( $x_{11}$  względem  $x_{12}$ ) oraz równym zero ( $x_3$  względem  $x_9$ ) współczynnika korelacji zostaną przedstawione wykresy rozrzutu wygenerowane w programie *Statistica*.



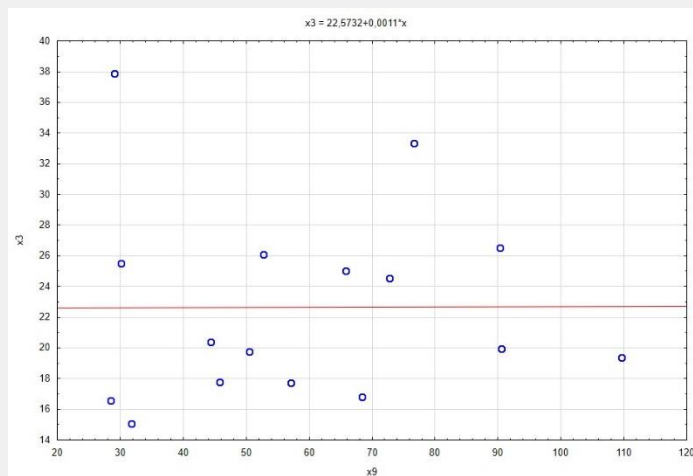
Wykres rozrzutu  $x_3$  względem  $x_2$



Wykres rozrzutu  $x_5$  względem  $x_6$



Wykres rozrzutu x11 względem x12



Wykres rozrzutu x3 względem x9

Z dalszej analizy wykluczaliśmy zmienną x1, by wyeliminować redundancję informacji (bardzo wysoki współczynnik korelacji ze zmienną x3), a także zmienne x7, x10, x13, x16 oraz x17 ze względu na słabą korelację z innymi zmiennymi, co wskazuje na ich nieistotność statystyczną.

## Wybór głównych składowych oraz ich równania



W celu ograniczenia liczby głównych składowych w dalszej analizie, korzysta się z różnych kryteriów pomagających wybrać odpowiednią ich liczbę. Jedną z metod jest wybranie najmniejszej liczby głównych składowych, których skumulowany procent tłumaczy, co najmniej, uprzednio wybrany procent skumulowany, np. 75%. Metoda, w której pozostawia się główne składowe, których wartości własne są większe od jedności, nazywa się kryterium Kaisera. Kryterium Cattella bazuje na analizie wykresu osypiska. Zgodnie z nim „odcina” się składowe, u których występuje łagodny spadek wartości własnych, czyli wszystkie składowe, które znajdują się na prawo od tzw. kolanka.

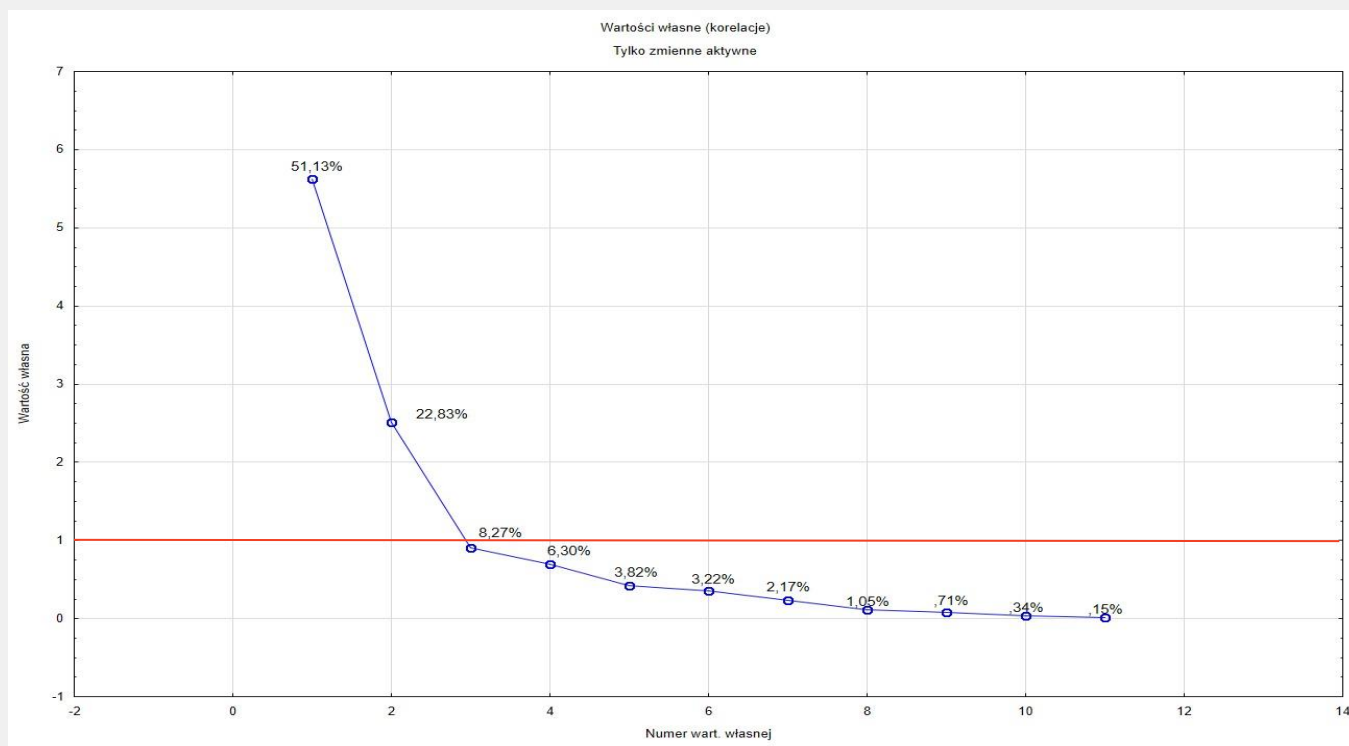
**Tabela 26. Wartości własne składowych**

Składowa	Wartość własna	% ogółu	Skumul. wartość własna	Skumul. %
PCA1	5,62	51,13	5,62	51,13
PCA2	2,51	22,83	8,14	73,96
PCA3	0,91	8,27	9,05	82,23
PCA4	0,69	6,30	9,74	88,53
PCA5	0,42	3,82	10,16	92,35
PCA6	0,35	3,22	10,51	95,57
PCA7	0,24	2,17	10,75	97,74
PCA8	0,12	1,05	10,87	98,79
PCA9	0,08	0,71	10,95	99,50
PCA10	0,04	0,34	10,98	99,85
PCA11	0,02	0,15	11,00	100,00

Tabela 26. przedstawia wartości własne wszystkich składowych. W celu ograniczenia liczby głównych składowych w dalszej części analizy, zastosowaliśmy kryterium Kaisera. Zgodnie z nim do analizy przyjęliśmy 2 główne składowe, ponieważ ich wartości własne są większe od jedności. Łącznie tłumaczą one 73,96% zmienności

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

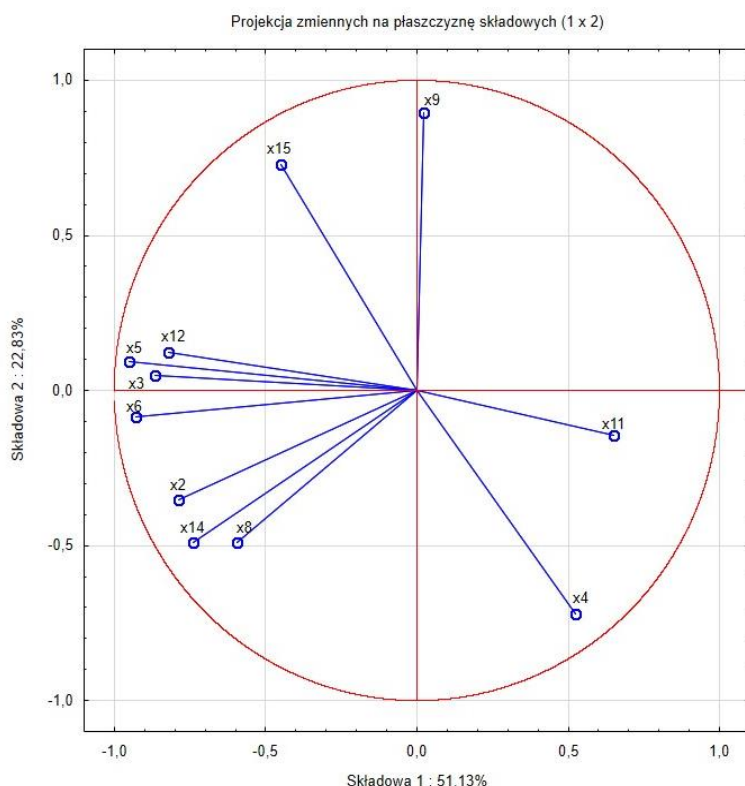
**Wykres osypiska**



Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Na powyższym wykresie kolanko tworzy się pomiędzy drugą a trzecią składową. Od trzeciej składowej wartości własne składowych nie różnią się znacząco. Sugeruje to dokonanie „cięcia” za drugą główną składową. Wniosek z kryterium Cattella pokrywa się z wnioskiem z kryterium Kaisera, co potwierdza słuszność wybrania do dalszej analizy dwóch pierwszych głównych składowych.

## Projekcja zmiennych na płaszczyznę składowych



Z wykresu obok możemy wywnioskować, że najmniejszy wpływ na pierwszą i drugą główną składową ma zmienna x11, której wektor jest najkrótszy. Zmienne x3, x5, x6 i x12 są silnie skorelowane ze sobą i mają duży wpływ na pierwszą składową. Znaczącą korelację zaobserwujemy także między zmiennymi x2, x8 i x14.

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

### Równania głównych składowych:

$$Y_1 = -0,333X_2 - 0,364X_3 + 0,221X_4 - 0,402X_5 - 0,392X_6 - 0,251X_8 + 0,009X_9 + 0,274X_{11} - 0,346X_{12} - 0,312X_{14} - 0,190X_{15}$$

$$Y_2 = -0,223X_2 + 0,031X_3 - 0,454X_4 + 0,060X_5 - 0,055X_6 - 0,310X_8 + 0,566X_9 - 0,092X_{11} + 0,077X_{12} - 0,308X_{14} + 0,459X_{15}$$

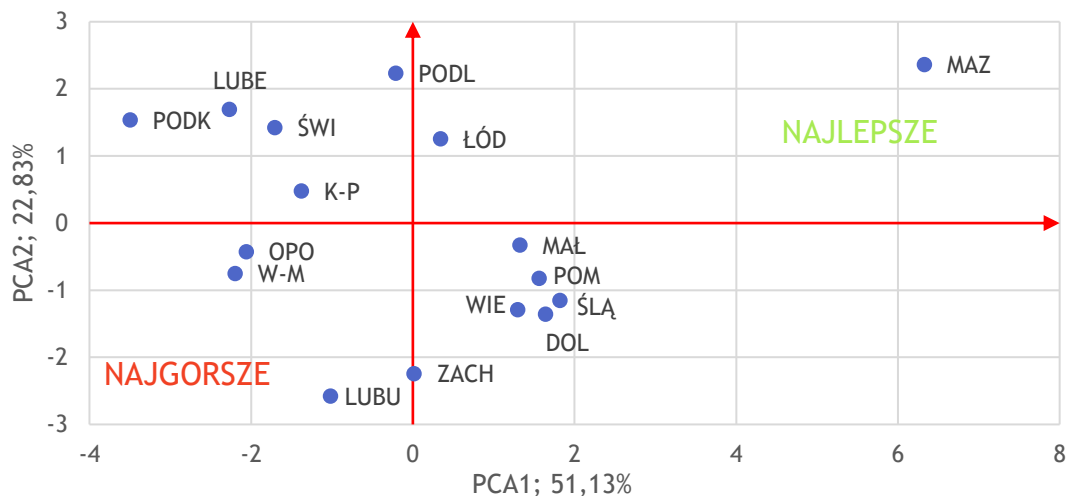
Tabela 27. Macierz korelacji składowych

Zmienna	Składowa 1	Składowa 2
x2	0,79	0,35
x3	0,86	-0,05
x4	-0,53	0,72
x5	0,95	-0,09
x6	0,93	0,09
x8	0,60	0,49
x9	-0,02	-0,90
x11	-0,65	0,15
x12	0,82	-0,12
x14	0,74	0,49
x15	0,45	-0,73

Macierz korelacji składowych pozwala zauważyć, że największy wkład w budowanie Składowej 1 mają zmienne x2, x3, x5, x6, x8, x11, x12 i x14. Spośród nich tylko zmienna x11 wykazuje korelację ujemną. Największy wkład w budowanie składowej 2 mają zmienne x4, x9 i x15, z których tylko zmienna x4 wykazuje korelację dodatnią.

W celu uzyskania najlepszych województw w pierwszej ćwiartce na wykresie poniżej, wartości składowych zostały pomnożone przez -1.

### Rzut obserwacji na płaszczyznę składowych (1 x 2)



#### Legenda

DOL - dolnośląskie  
 K-P - kujawsko-pomorskie  
 LUBE - lubelskie  
 LUBU - lubuskie  
 ŁÓD - łódzkie  
 MAŁ - małopolskie  
 MAZ - mazowieckie  
 OPO - opolskie  
 PODK - podkarpackie  
 PODL - podlaskie  
 POM - pomorskie  
 ŚŁĄ - śląskie  
 ŚWI - świętokrzyskie  
 W-M - warmińsko-mazurskie  
 WIE - wielkopolskie  
 ZACH - zachodniopomorskie

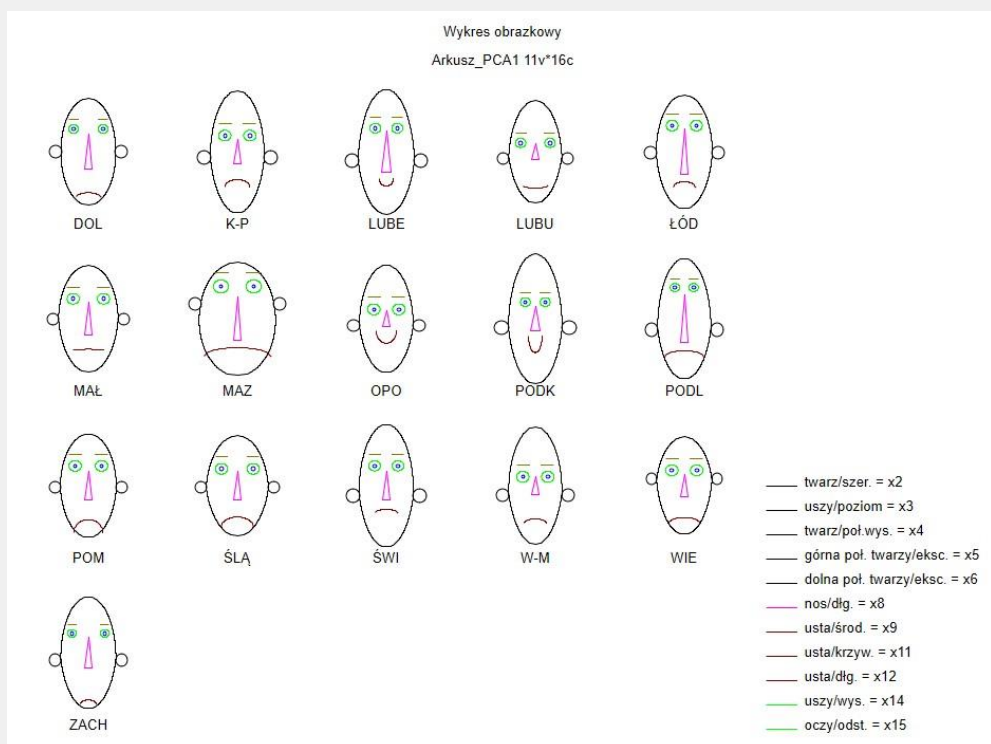
Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Najbardziej znacząca jest wysoka liczba wolnych miejsc pracy oraz wysokość wynagrodzenia przy pierwszej głównej składowej, z kolei przy drugiej składowej największe znaczenie mają parki i zieleńce. Najlepszymi województwami pod względem składowej pierwszej są mazowieckie i śląskie, natomiast pod względem składowej drugiej najlepsze są mazowieckie i podlaskie. Analizując rzut obserwacji na płaszczyznę składowych, możemy stwierdzić, że województwo mazowieckie zdecydowanie odstaje względem reszty kraju, wykazując wysokie wartości zarówno składowej pierwszej jak i składowej drugiej. Województwem najgorzej wypadającym pod względem obu składowych jest województwo lubuskie.

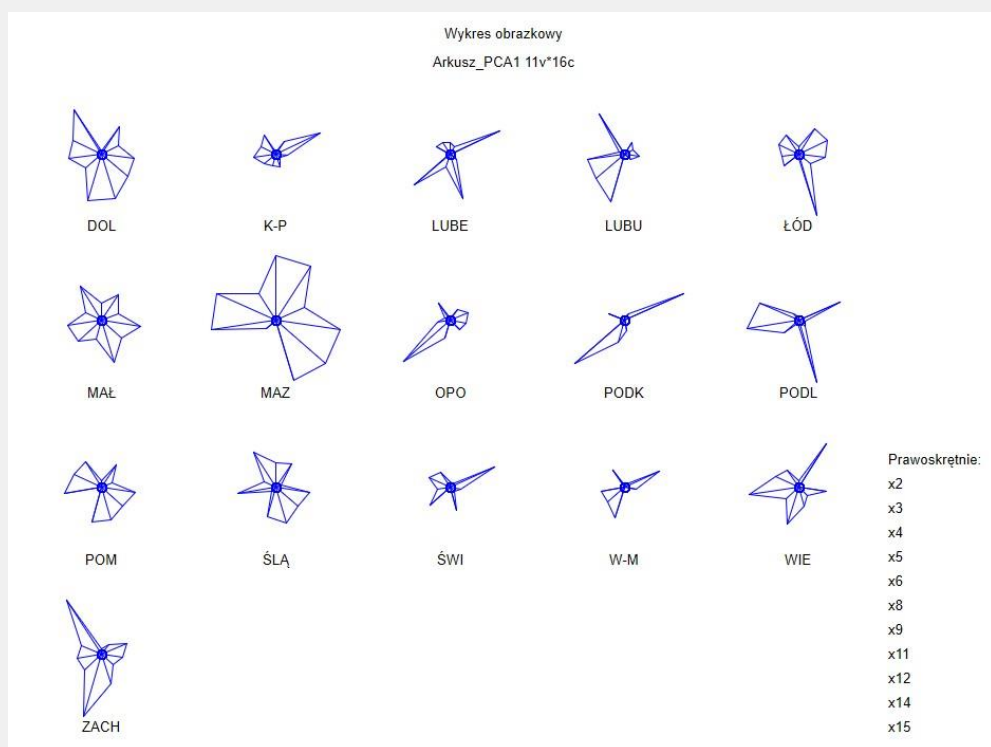




## Dodatkowe wizualizacje:



Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.



Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Na wykresie Chernoffa poszczególne zmienne odzwierciedlane są przez różne charakterystyki twarzy ludzkiej. Najmniejszej wartości każdej zmiennej przypisujemy minimalną wartość charakterystyki twarzy, tzw. najmniejsza twarz, natomiast maksymalna wartość każdej cechy otrzymuje twarz największą. W wyniku tego przekształcenia obrazy wszystkich rozpatrywanych danych znajdują się pomiędzy dwoma skrajnymi twarzami.

Na wykresie obrazkowym gwiazdy względne wartości wybranych zmiennych dla każdego przypadku reprezentowane są przez długości ramion gwiazdy (zgodnie z ruchem wskazówek zegara, począwszy od godziny 12:00). Końce ramion są połączone linią.

# ANALIZA SKUPIEŃ

## Wstęp

Analiza skupień opiera się na pomiarze stopnia podobieństwa lub zróżnicowania obiektów w wielowymiarowej przestrzeni. Pomiaru tego można dokonać za pomocą różnych metod - hierarchicznych lub niehierarchicznych, a jej celem jest wyodrębnienie grup jednostek do siebie podobnych. Do analizy wykorzystano macierz z 14 zmiennymi.

We wszystkich technikach analizy skupień ważną rolę pełnią pojęcia odległości. W projekcie wykorzystano następujące miary odległości:

- **Odległość miejska** - obliczana według poniższego wzoru:

$$\sum_i |x_i - y_i|$$

- **Odległość euklidesowa** - obliczana według poniższego wzoru:

$$\left\{ \sum_i (x_i - y_i)^2 \right\}^{\frac{1}{2}}$$

- **Odległość euklidesowa kwadratowa** - obliczana według poniższego wzoru:

$$\sum_i (x_i - y_i)^2$$

Tabela 28. Macierz odległości miejskiej

	DOL	K-P	LUBE	LUBU	ŁÓD	MAŁ	MAZ	OPO	PODK	PODL	POM	ŚLĄ	ŚWI	W-M	WIE	ZACH
DOL	0	185	189	130	258	111	272	322	219	182	69	149	248	202	103	99
K-P	185	0	102	163	225	121	275	194	120	177	148	155	157	149	137	155
LUBE	189	102	0	176	245	120	343	233	62	153	172	205	177	148	179	171
LUBU	130	163	176	0	320	160	391	214	180	242	158	174	240	99	145	109
ŁÓD	258	225	245	320	0	240	330	181	297	262	268	232	121	303	270	273
MAŁ	111	121	120	160	240	0	252	269	158	179	82	105	232	180	122	149
MAZ	272	275	343	391	330	252	0	421	385	227	249	270	331	415	261	350
OPO	322	194	233	214	181	269	421	0	231	352	303	248	158	226	279	273
PODK	219	120	62	180	297	158	385	231	0	192	187	237	205	128	176	183
PODL	182	177	153	242	262	179	227	352	192	0	167	253	209	211	183	232
POM	69	148	172	158	268	82	249	303	187	167	0	121	246	188	75	131
ŚLĄ	149	155	205	174	232	105	270	248	237	253	121	0	221	211	160	187
ŚWI	248	157	177	240	121	232	331	158	205	209	246	221	0	212	240	214
W-M	202	149	148	99	303	180	415	226	128	211	188	211	212	0	183	167
WIE	103	137	179	145	270	122	261	279	176	183	75	160	240	183	0	129
ZACH	99	155	171	109	273	149	350	273	183	232	131	187	214	167	129	0

Źródło: opracowanie własne na podstawie obliczeń w programie Statistica.

W tabeli 28 zaprezentowano macierz odległości miejskiej. Jak można na jej podstawie zauważyć, największa odległość (421) dzieli województwo mazowieckie i opolskie. Oznacza to, iż wspomniane województwa najbardziej się różnią. Z kolei najkrótszy dystans (62) i tym samym największe podobieństwo dostrzega się pomiędzy lubelskim a podkarpackim.

**Tabela 29. Macierz odległości euklidesowej**

	DOL	K-P	LUBE	LUBU	ŁÓD	MAŁ	MAZ	OPO	PODK	PODL	POM	ŚLĄ	ŚWI	W-M	WIE	ZACH
DOL	0,0	66,4	70,4	62,3	137,4	50,8	103,2	131,9	69,1	72,6	24,8	64,0	104,1	79,7	33,7	34,4
K-P	66,4	0,0	37,2	62,8	117,2	39,0	117,8	103,1	45,4	72,6	52,4	49,0	73,7	61,9	50,6	72,2
LUBE	70,4	37,2	0,0	64,8	137,3	46,1	133,0	123,9	23,2	71,4	60,4	73,1	92,3	51,4	63,2	73,8
LUBU	62,3	62,8	64,8	0,0	145,8	61,1	152,7	118,0	61,0	104,3	64,8	67,1	109,0	36,3	53,7	44,2
ŁÓD	137,4	117,2	137,3	145,8	0,0	134,5	141,9	64,1	151,1	147,8	138,6	100,0	52,9	157,9	139,7	145,6
MAŁ	50,8	39,0	46,1	61,1	134,5	0,0	109,8	127,8	52,7	66,0	35,3	50,2	97,5	59,7	39,1	66,1
MAZ	103,2	117,8	133,0	152,7	141,9	109,8	0,0	164,4	137,7	84,7	95,7	119,1	122,0	161,9	105,6	133,9
OPO	131,9	103,1	123,9	118,0	64,1	127,8	164,4	0,0	132,4	154,5	133,5	95,5	59,3	133,1	127,2	127,0
PODK	69,1	45,4	23,2	61,0	151,1	52,7	137,7	132,4	0,0	76,3	59,6	82,4	104,4	48,6	60,0	68,5
PODL	72,6	72,6	71,4	104,3	147,8	66,0	84,7	154,5	76,3	0,0	58,7	95,3	107,0	101,7	69,2	97,0
POM	24,8	52,4	60,4	64,8	138,6	35,3	95,7	133,5	59,6	58,7	0,0	59,9	101,6	74,4	23,0	49,2
ŚLĄ	64,0	49,0	73,1	67,1	100,0	50,2	119,1	95,5	82,4	95,3	59,9	0,0	75,8	80,4	58,6	71,2
ŚWI	104,1	73,7	92,3	109,0	52,9	97,5	122,0	59,3	104,4	107,0	101,6	75,8	0,0	116,7	102,6	111,5
W-M	79,7	61,9	51,4	36,3	157,9	59,7	161,9	133,1	48,6	101,7	74,4	80,4	116,7	0,0	68,4	69,7
WIE	33,7	50,6	63,2	53,7	139,7	39,1	105,6	127,2	60,0	69,2	23,0	58,6	102,6	68,4	0,0	46,6
ZACH	34,4	72,2	73,8	44,2	145,6	66,1	133,9	127,0	68,5	97,0	49,2	71,2	111,5	69,7	46,6	0,0

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Kolejna macierz (tabela 29) zawiera odległości euklidesowe. Ponownie największy dystans (164) jest widoczny pomiędzy województwem mazowieckim i opolskim. Tym razem jednak najbardziej podobnymi województwami okazały się być pomorskie oraz wielkopolskie (odległość równa 23).

Tabela 30. Macierz kwadratowej odległości euklidesowej

	DOL	K-P	LUBE	LUBU	ŁÓD	MAŁ	MAZ	OPO	PODK	PODL	POM	ŚLĄ	ŚWI	W-M	WIE	ZACH
DOL	0	4414	4951	3879	18890	2579	10648	17400	4772	5269	614	4097	10830	6351	1134	1185
K-P	4414	0	1386	3943	13727	1521	13888	10639	2057	5271	2748	2397	5424	3833	2561	5215
LUBE	4951	1386	0	4196	18838	2126	17693	15360	540	5096	3654	5338	8512	2643	3999	5447
LUBU	3879	3943	4196	0	21268	3731	23319	13934	3719	10888	4195	4496	11874	1320	2882	1953
ŁÓD	18890	13727	18838	21268	0	18090	20143	4110	22832	21858	19220	9999	2795	24936	19512	21186
MAŁ	2579	1521	2126	3731	18090	0	12046	16335	2780	4353	1245	2519	9513	3563	1528	4372
MAZ	10648	13888	17693	23319	20143	12046	0	27027	18972	7175	9166	14178	14880	26206	11143	17932
OPO	17400	10639	15360	13934	4110	16335	27027	0	17532	23873	17821	9124	3515	17706	16184	16137
PODK	4772	2057	540	3719	22832	2780	18972	17532	0	5819	3550	6789	10904	2357	3604	4688
PODL	5269	5271	5096	10888	21858	4353	7175	23873	5819	0	3441	9092	11446	10336	4793	9403
POM	614	2748	3654	4195	19220	1245	9166	17821	3550	3441	0	3585	10318	5542	528	2420
ŚLĄ	4097	2397	5338	4496	9999	2519	14178	9124	6789	9092	3585	0	5748	6471	3436	5064
ŚWI	10830	5424	8512	11874	2795	9513	14880	3515	10904	11446	10318	5748	0	13623	10527	12442
W-M	6351	3833	2643	1320	24936	3563	26206	17706	2357	10336	5542	6471	13623	0	4675	4854
WIE	1134	2561	3999	2882	19512	1528	11143	16184	3604	4793	528	3436	10527	4675	0	2173
ZACH	1185	5215	5447	1953	21186	4372	17932	16137	4688	9403	2420	5064	12442	4854	2173	0

Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

W tabeli 30 zaprezentowano macierz kwadratowej odległości euklidesowej. Znajdujące się w niej wartości prowadzą do takich samych wniosków, jak w przypadku odległości euklidesowych, ponieważ są to właśnie odległości euklidesowe podniesione do potęgi drugiej.



# Metody hierarchiczne

Najczęściej stosowanymi sposobami grupowania wielowymiarowych obiektów są metody hierarchiczne i dlatego to one zostaną omówione w pierwszej kolejności. W ich przypadku na każdym kroku elementy łączą się albo rozpadają na skupienia na danym poziomie - według odległości lub podobieństwa. W konsekwencji liczba skupień zmniejsza się lub zwiększa.

Prezentowane w tej części metody określa się również jako aglomeracyjne. Oznacza to, że na początku każda jednostka jest traktowana jako jednoelementowe skupienie, potem następuje łączenie jednostek i skupień. Jeśli proces nie zostanie na pewnym etapie przerwany, będzie trwać aż do momentu, kiedy wszystkie elementy utworzą jedno skupienie<sup>16</sup>.

Przetestowano następujące metody hierarchiczne: najbliższego sąsiada, najdalszego sąsiada, średniej grupowej (z wykorzystaniem odległości miejskiej i euklidesowej) oraz Warda (z kwadratową odległością euklidesową).

W dalszej części pracy zostaną omówione metody, dla których współczynnik korelacji kofenetycznej był najwyższy - metoda średniej grupowej (z odległością miejską i euklidesową), najbliższego sąsiada (z odległością euklidesową) i najdalszego sąsiada (z odległością euklidesową). Wysoka wartość oznacza bowiem, że dendrogram danej metody dobrze podsumowuje obserwowalne różnice między elementami lub ich podobieństwo. Warto dodać, że najmniej odpowiednia okazała się metoda najdalszego sąsiada z wykorzystaniem odległości miejskiej, a zaraz potem metoda Warda. Ich współczynniki były mniejsze niż zalecane 0,8.

**Tabela 31. Współczynniki korelacji kofenetycznej dla wybranych metod**

Najbliższy sąsiad (miejska)	Najbliższy sąsiad (euklidesowa)	Najdalszy sąsiad (miejska)	Najdalszy sąsiad (euklidesowa)	Średnia grupowa (miejska)	Średnia grupowa (euklidesowa)	Warda (kwadratowa euklidesowa)
0,825	0,870	0,753	0,828	0,863	0,895	0,784

Źródło: opracowanie własne na podstawie obliczeń w programie R.





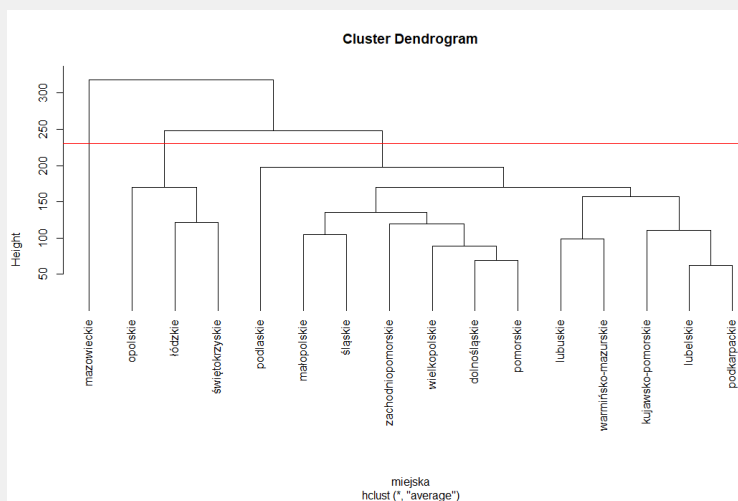
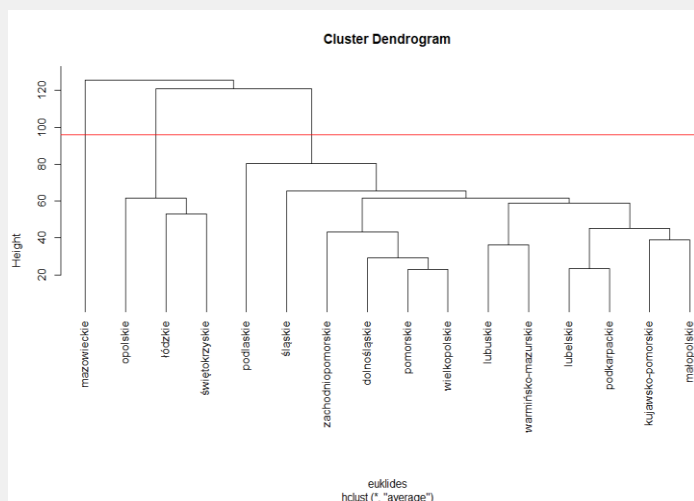
## Metoda średniej grupowej

Najwyższy współczynnik korelacji kofenetycznej, równy 0,895, odnotowano w przypadku metody średniej grupowej (średnich połączeń) z wykorzystaniem odległości euklidesowej. Wysoki współczynnik odnotowano też w przypadku odległości miejskiej - 0,863.

Metoda średnich połączeń pozwala zmniejszyć niepewność towarzyszącą metodom najbliższego i najdalszego sąsiada. Polega na przeliczaniu odległości, nie trzeba już wybierać najbliższego lub najdalszego sąsiada z dwóch różnych skupień obiektów. Metoda średniej grupowej jest także mniej wrażliwa na wartości ekstremalne. Jej podstawą jest „konceptcja średniej odległości, którą jest średnia arytmetyczna odległości między wszystkimi parami obiektów należących do dwóch porównywanych grup”<sup>17</sup>.

Poniżej zaprezentowano dendrogramy omawianej metody - z uwzględnieniem dwóch rodzajów odległości.

**Rysunek 1. Dendrogramy metody średniej grupowej z odległością euklidesową i miejską**



Źródło: opracowanie własne na podstawie obliczeń w programie R.

Czerwona linia pomaga w określeniu właściwej liczby grup województw. Wyliczenie jej położenia jest wynikiem pomysłu R. Mojeny z 1977 r., jednak za stałą przyjęto 1,25 - zgodnie z późniejszymi zaleceniami G.W. Miligana i M.C. Cooper.

Dla analizowanej metody wartość wyniosła 95,97 (z odległością euklidesową) i 230,54 (z miejską). W konsekwencji województwa podzielono na trzy skupienia. Pierwsze z nich stanowi tylko jedno województwo - mazowieckie. Kolejne skupienie jest trzelementowe i należą do niego województwa: opolskie, łódzkie oraz świętokrzyskie. Pozostałe 12 województw tworzy trzecie skupienie. Jak można zauważyć na podstawie dendrogramów (i opierając się na kryterium Mojeny), wyznaczono takie same skupienia dla odległości miejskiej i euklidesowej. Gdyby jednak zwrócić uwagę na zgrupowania występujące pod czerwoną linią, zauważyłoby się już pewne różnice.



Na koniec warto przyjrzeć się przebiegowi aglomeracji analizowanej metody (tabela 32). W ten sposób można bowiem sprawdzić dokładne odległości, na których dochodzi do połączeń określonych województw, łatwiej jest je odczytać niż z wcześniej zaprezentowanego dendrogramu. Widać, że najmniej oddalone od siebie (czyli najbardziej podobne) są województwa pomorskie oraz wielkopolskie. Zauważa się także, w jakim stopniu województwo mazowieckie odbiega od reszty.

**Tabela 32. Przebieg aglomeracji metody średniej grupowej z odległością euklidesową**

Odległość	Przebieg aglomeracji
22,98	PM, WP
23,24	LB, PK
29,22	DŚ, PM, WP
36,34	LS, WM
39,00	KP, MP
43,41	DŚ, PM, WP, ZP
45,35	KP, MP, LB, PK
52,87	ŁD, ŚK
58,90	KP, MP, LB, PK, LS, WM
61,46	DŚ, PM, WP, ZP, KP, MP, LB, PK, LS, WM
61,70	ŁD, ŚK, OP
65,58	DŚ, PM, WP, ZP, KP, MP, LB, PK, LS, WM, ŚL
80,46	DŚ, PM, WP, ZP, KP, MP, LB, PK, LS, WM, ŚL, PL
121,03	DŚ, PM, WP, ZP, KP, MP, LB, PK, LS, WM, ŚL, PL, ŁD, ŚK, OP
125,56	DŚ, PM, WP, ZP, KP, MP, LB, PK, LS, WM, ŚL, PL, ŁD, ŚK, OP, MZ

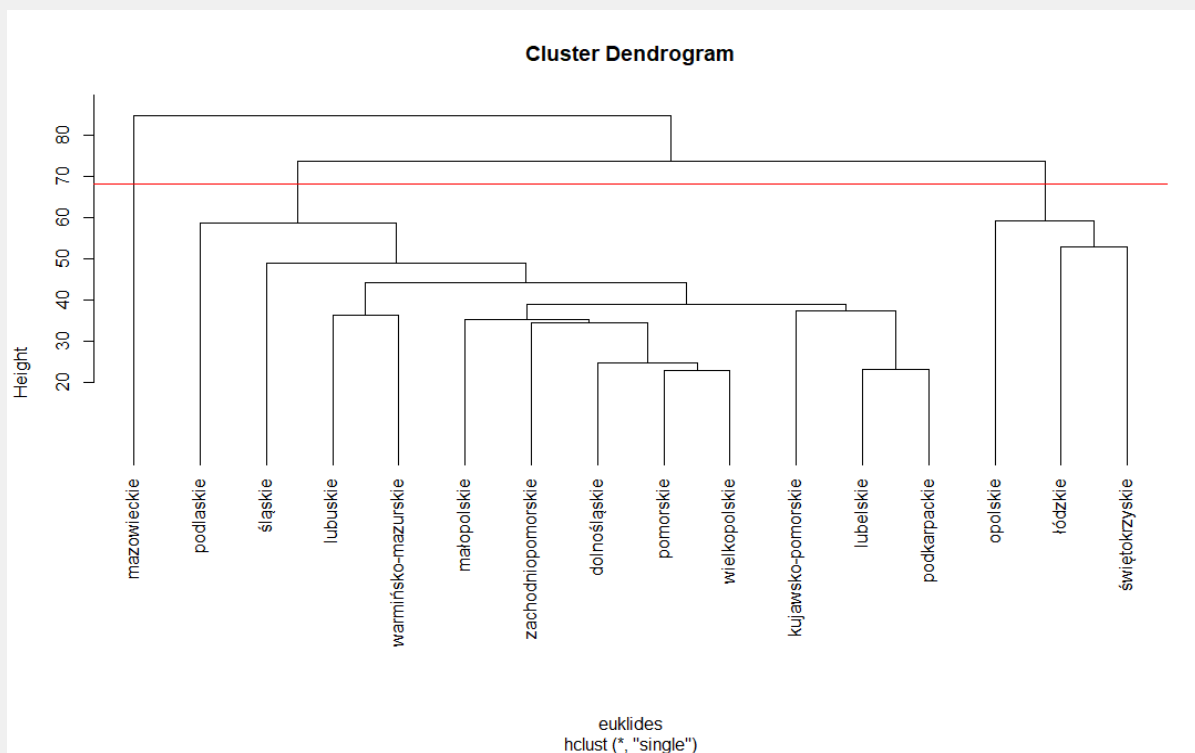
Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

## Metoda najbliższego sąsiada

Wysoki współczynnik korelacji kofenetycznej (0,87) otrzymano również dla metody najbliższego sąsiada (z wykorzystaniem odległości euklidesowej).

W metodzie tej przelicza się odległości między jednostkami jednego skupienia a jednostkami innego skupienia, zgodnie z kryterium najmniejszej odległości. Odległości pomiędzy nowoutworzonym skupieniem a innymi skupieniami są określane przez najbliższe obiekty w nowym skupieniu w stosunku do pozostałych skupień<sup>18</sup>.

Rysunek 2. Dendrogram metody najbliższego sąsiada z odległością euklidesową



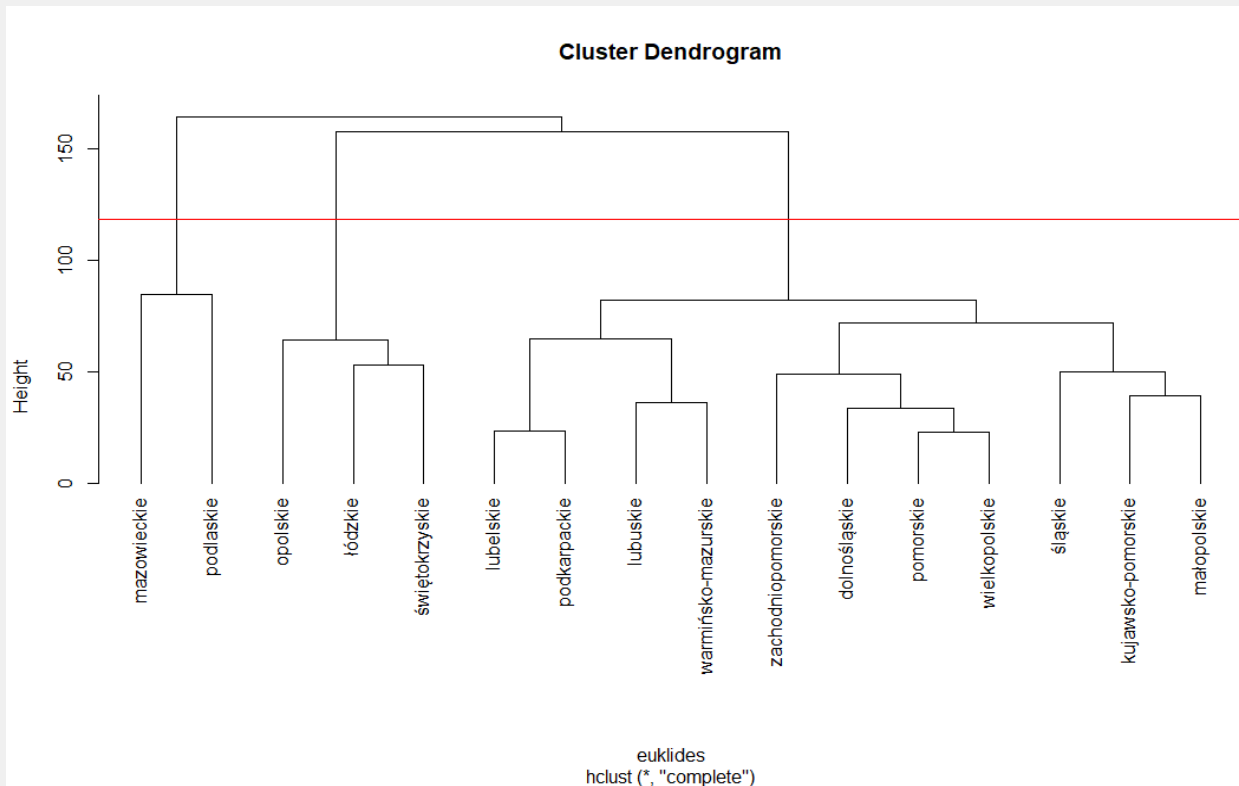
Źródło: opracowanie własne na podstawie obliczeń w programie R.

Ponownie położenie czerwonej linii na dendrogramie (rysunek 2) wyznacza podział województw na skupienia. Dla metody pojedynczego wiązania z odległością euklidesową wartość poziomu łączenia skupień wynikająca z sugestii Mojeny wyniosła 68,27. Była zatem niższa niż w przypadku metody średnich połączeń. Mimo to województwa zostały pogrupowane w trzy takie same skupienia.

## Metoda najdalszego sąsiada

W przypadku metody najdalszego sąsiada z odległością euklidesową współczynnik korelacji kofenetycznej również był stosunkowo wysoki (0,828). Tym razem o możliwym połączeniu skupień decydują odległości między najbardziej oddalonymi obiektami w skupieniach. Na poszczególnych etapach łączy się takie dwa skupienia, które mają najmniejszą wśród największych odległości<sup>19</sup>.

Rysunek 3. Dendrogram metody najdalszego sąsiada z odległością euklidesową



Źródło: opracowanie własne na podstawie obliczeń w programie R.

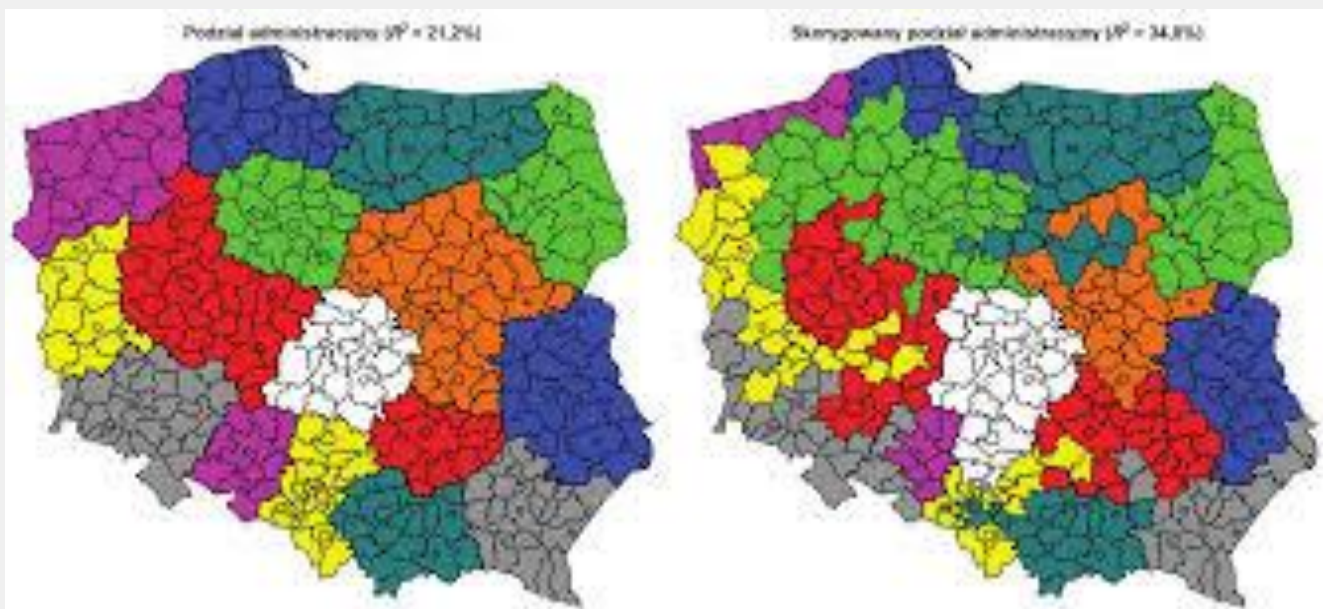
Na rysunku 3 przedstawiono dendrogram opisywanej metody. Tym razem wartość wynikająca z kryterium Mojeny wyniosła 118,25, a więc była wyższa niż przy wcześniej wspomnianych metodach z wykorzystaniem odległości euklidesowej.

Dodatkowo zauważa się różnicę w podziale województw. Pierwsze skupienie nie jest już jednoelementowe. Województwo mazowieckie tworzy skupienie razem z podlaskim. W związku z tym podlaskie nie należy już do najbardziej licznego skupienia, jak było to w przypadku wcześniej analizowanych metod. Poza tym, opierając się cały czas na kryterium Mojeny, województwa są już przyporządkowane do takich samych skupień jak wcześniej.

# Metody niehierarchiczne

Innym rodzajem metod analizy skupień są metody niehierarchiczne, w których poszukuje się takiego podziału, który jest optymalny ze względu na pewne przyjęte kryterium grupowania. Do grupy metod niehierarchicznych zaliczane są metody, w wyniku stosowania, których następuje wyodrębnienie skupień niezawierających innych skupień czyli skupień rozłącznych oraz prowadzące do otrzymania skupień nierozłącznych. W metodach klasyfikacji nierozłącznej dopuszcza się przynależność klasyfikowanych jednostek do więcej niż jednego skupienia, natomiast w klasyfikacji rozłącznej takie postępowanie jest niemożliwe.

W projekcie omówiono metodę k-średnich, której celem jest tworzenie skupień rozłącznych oraz przetestowaną ją dla kilku skupień. Następnie wykonano wykresy średnich dla każdego skupienia i przedstawiono ostateczny podział na skupienia.



Zdjęcie obrazujące proces zachodzący podczas analizy metodą k-średnich.

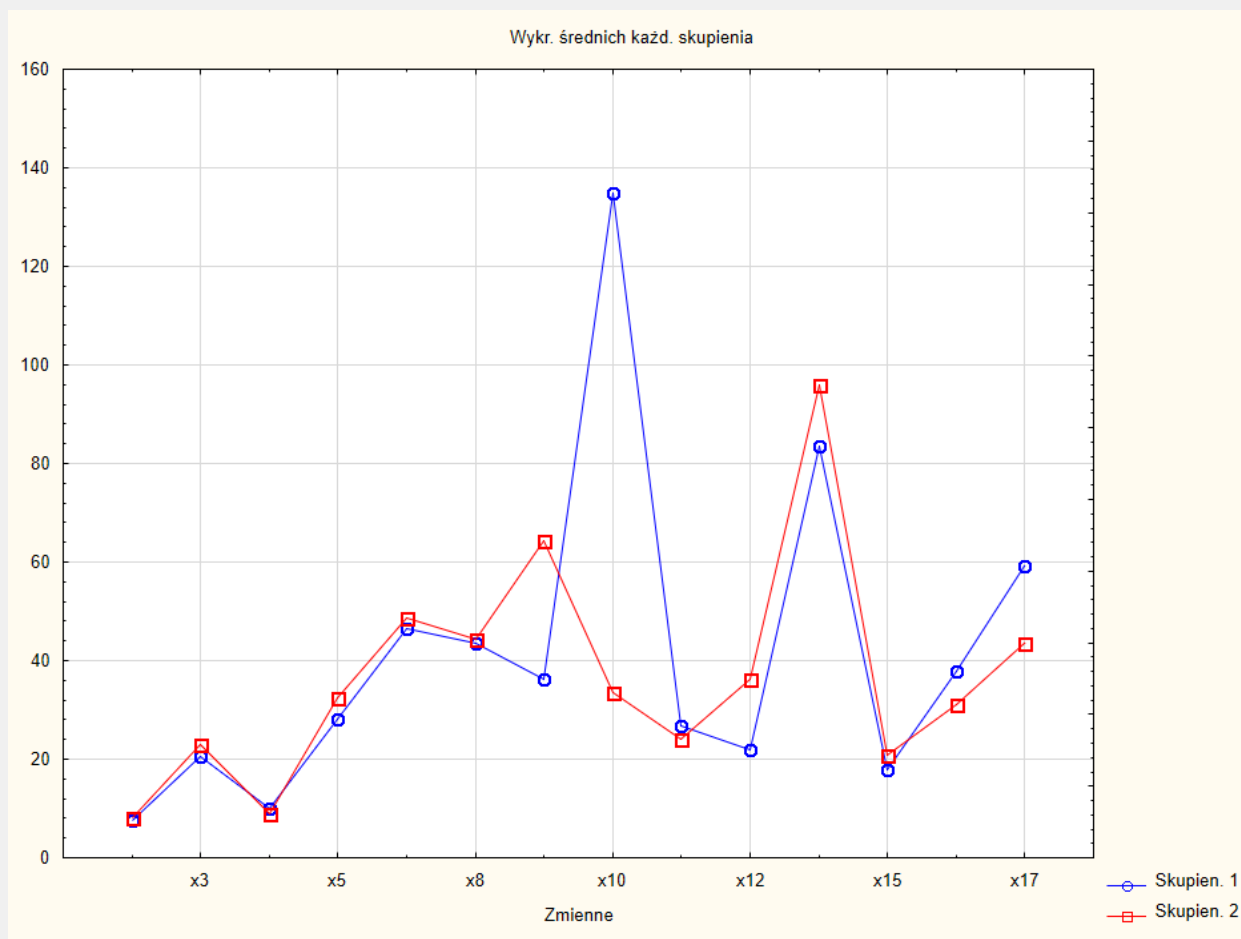
## Metoda k-średnich

Metoda k-średnich znacznie różni się od metod przedstawionych wcześniej. Analizę rozpoczynamy od wyboru liczby skupień czyli liczby grup, na które zostaną podzielone nasze dane. Celem tej metody jest osiągnięcie maksymalnego podobieństwa wewnątrz skupienia, a minimalnego podobieństwa między członkami danej grupy a pozostałymi obiektami.

W projekcie wykonano analizę dla 2, 4 oraz 8 skupień. W dalszej części projektu przedstawione są wyniki dla 2 i 4 skupień. Liczba 8 skupień okazała się zbyt duża, ponieważ powstały skupienia jednoelementowe, dlatego w tym przypadku nie kontynuowaliśmy obliczeń.

### Wyniki dla 2 skupień

Wykres średnich dla 2 skupień



Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

Poniżej zaprezentowano dwie tabele przedstawiające utworzone skupienia.

**Tabela 33. Skupienie 1**

województwo	Odległ.
łódzkie	9,039
opolskie	9,942
świętokrzyskie	8,219

**Tabela 34. Skupienie 2**

województwo	Odległ.
dolnośląskie	9,350
kujawsko-pomorskie	9,155
lubelskie	11,263
lubuskie	13,776
małopolskie	6,786
mazowieckie	28,554
podkarpackie	11,876
podlaskie	16,062
pomorskie	6,070
śląskie	13,563
warmińsko-mazurskie	15,579
wielkopolskie	6,824
zachodniopomorskie	12,993

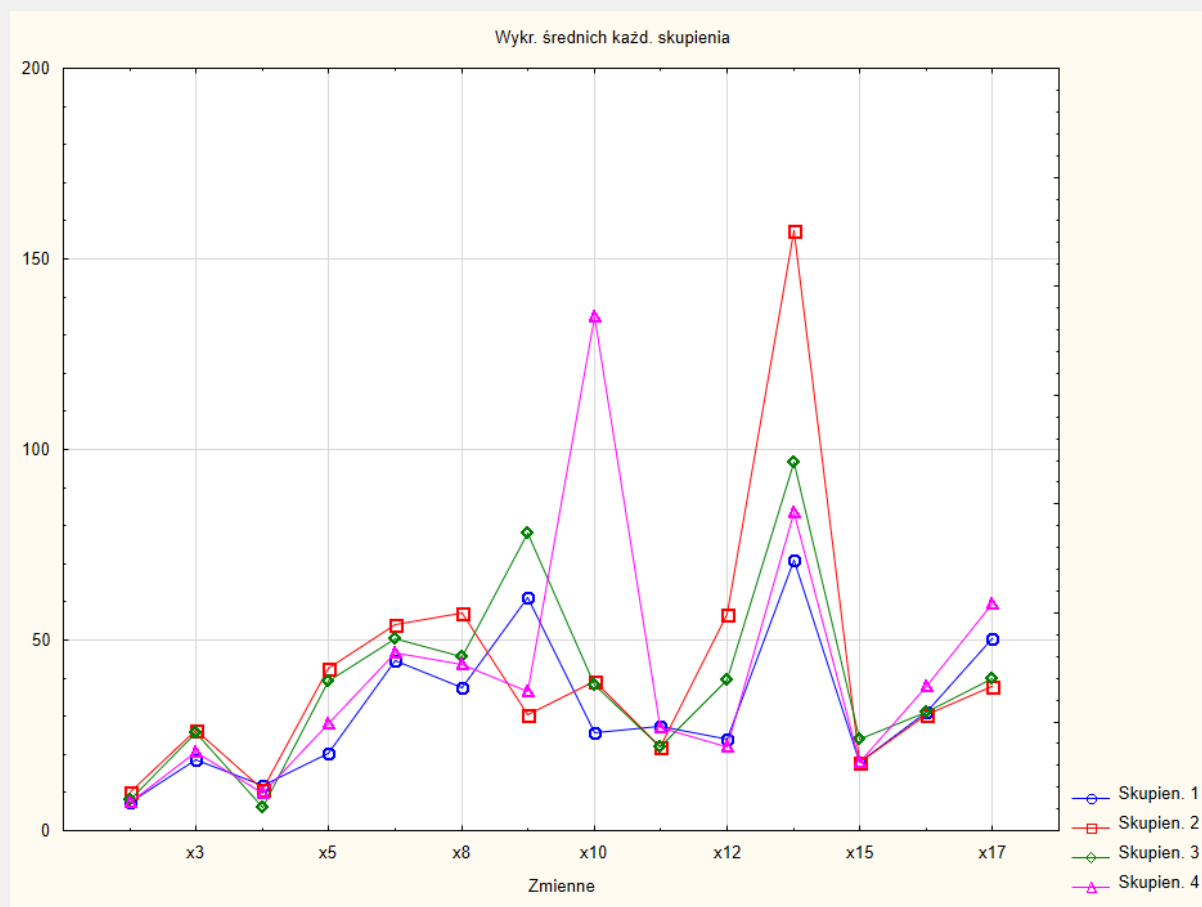
Źródło: Opracowanie własne na podstawie obliczeń w programie *Statistica*

Powstało jedno skupienie trzelementowe i jedno trzynastoelementowe. Tak duża różnica w liczebności grup może świadczyć o tym, że liczba skupień jest niewystarczająca. Z powyżej przedstawionych tabel wynika, iż najbardziej podobne do siebie są województwa łódzkie, opolskie i świętokrzyskie oraz można stwierdzić, że wymienione województwa najbardziej różnią się od pozostałych województw z drugiego skupienia.



## Wyniki dla 4 skupień

Wykres średnich dla 4 skupień



Źródło: opracowanie własne na podstawie obliczeń w programie *Statistica*.

**Tabela 35. Skupienie 1**

województwo	Odległ.
kujawsko-pomorskie	9,274
lubelskie	7,137
lubuskie	10,677
podkarpackie	7,045
warmińsko-mazurskie	8,413

**Tabela 36. Skupienie 2**

województwo	Odległ.
mazowieckie	11,319
podlaskie	11,319

**Tabela 37. Skupienie 3**

województwo	Odległ.
dolnośląskie	6,481
małopolskie	8,566
pomorskie	5,247
śląskie	12,257
wielkopolskie	5,689
zachodniopomorskie	10,428

**Tabela 38. Skupienie 4**

województwo	Odległ.
łódzkie	9,039
opolskie	9,942
świętokrzyskie	8,219

Po wykonaniu analizy metodą k-średnich dla 4 skupień, można stwierdzić, że najbardziej podobne są do siebie województwa kujawsko-pomorskie, lubelskie, podkarpackie i warmińsko-mazurskie. Województwa mazowieckie i podlaskie otrzymały identyczny wynik. Trzecie skupienie wskazuje na to, że województwa dolnośląskie, małopolskie, pomorskie, śląskie, wielkopolskie i zachodniopomorskie są do siebie nawzajem najbardziej podobne. Pozostałe województwa tj. łódzkie, opolskie i świętokrzyskie znalazły się w 4 skupieniu.



## Podsumowanie

Celem analizy było wytypowanie województwa o najwyższej jakości życia, na podstawie wcześniej wybranych dziedzin jakości życia.

Metoda głównych składowych wyraźnie wskazała, że najlepszym pod względem jakości życia jest województwo mazowieckie. Na drugim miejscu znalazło się województwo łódzkie. Wnioskujemy tak na podstawie wysokich wartości PCA1 i PCA2 w obu przypadkach. Świadczy to o wysokich wartościach zmiennych uważanych za stymulanty jednocześnie przy niskich wartościach destymulant. Przykładowe stymulanty to liczba pracujących w sektorze publicznym czy wysokość wynagrodzeń.

W porządkowaniu liniowym w przypadku większości metod również mazowieckie było typowane jako najlepsze, natomiast następne w kolejności były wielkopolskie i dolnośląskie.

# Bibliografia

1. <https://www.money.pl/gielda/gus-inflacja-srednioroczna-wyniosla-1-6-w-2018-r-wobec-2-w-2017-r-6338736306026625a.html> [25.11.2020].
2. GUS - Produkt krajowy brutto w 2019 roku - szacunek wstępny.
3. <https://www.money.pl/gielda/polska-gospodarka-na-10-wykresach-jaki-byl-2019-rok-6459785796986497a.html> [10.10.2020].
4. *Jakość życia w Polsce w 2015 r.*, GUS, 2017.
5. *Regionalne zróżnicowanie jakości życia w 2018 r. Wyniki Badania spójności społecznej 2018*, GUS, 2019.
6. T. Panek, *Jakość życia gospodarstw domowych w Polsce w układzie wojewódzkim*, „Zeszyty Naukowe - Instytut Statystyki i Demografii”, 2015, nr 46.
7. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 317, 318.
8. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 319.
9. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 319.
10. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 318-319.
11. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 321.
12. A. Bąk, *Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS. - analiza porównawcza*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu”, 2016, nr 426, s. 26.
13. A. Bąk, *Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS. - analiza porównawcza*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu”, 2016, nr 426, s. 26.
14. A. Balicki, dz. cyt., s. 322; A. Bąk, dz. cyt., s. 26.
15. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, dz. cyt., s. 324.
16. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 259-260.
17. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 270.
18. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 261.
19. A. Balicki, *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2009, s. 267.