

*Prezentacja pracy magisterskiej pt. Obsadzone czy
nieobsadzone? Statystyczna analiza wyników naborów w służbie
cywilnej z wykorzystaniem metod analizy regresji i uczenia
maszynowego (Wasiak i Walczak)*

Maciej Beręsewicz^{1,2}

¹Katedra Statystyki, Uniwersytet Ekonomiczny w Poznaniu

²Ośrodek Metodologii Badań Ludnościowych, Urząd Statystyczny w Poznaniu

Warszawa, 25.09.2025



POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS

Plan prezentacji

1 Wprowadzenie

- Praca magisterska
- Najważniejsze wnioski
- Dekompozycja predykcji

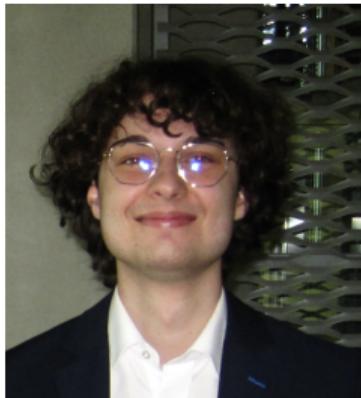
2 Dane – pozyskanie i przygotowanie

3 Zaproponowane podeście

4 Wyniki (wybrane)

5 Podsumowanie

Praca magisterska



mgr Konrad Wasiak i mgr Manuela Walczak

- Praca mgr pt. Obsadzone czy nieobsadzone? Statystyczna analiza wyników naborów w służbie cywilnej z wykorzystaniem metod analizy regresji i uczenia maszynowego
- Przygotowana i obroniona na Uniwersytecie Ekonomicznym w Poznaniu
- Obrona: 14 lipca 2025
- Promotor: dr Maciej Beręsewicz, prof. UEP
- Recenzent: dr hab. Tomasz Klimanek, prof. UEP

Wprowadzenie

- Celem pracy było wskazanie czynników mających wpływ na rekrutację (w służbie cywilnej) i pogłębioną interpretację skali ich oddziaływania.
- Główne źródło danych stanowiły archiwalne ogłoszenia naborów z lat 2017-2023, pochodzące z oficjalnego portalu rekrutacyjnego <https://nabory.kprm.gov.pl>.
- Pobrano 107 260 ogłoszeń z lat 2017-2023, pochodzących z ponad 1 660 instytucji i 347 powiatów (w tym miast na prawach powiatu).
- Zbiór danych wzbogacono o ogólnodostępne dane dla powiatów oraz dodatkowe informacje o instytucjach.

Najważniejsze wnioski

Najważniejsze wnioski (1) – wyjaśnienie sukcesu naboru

Ranking	Czynnik	Ważność	% Ważność
1	Instytucja	0.35	32.21
2	Wykształcenie	0.26	23.63
3	Powiat (lokalny rynek pracy)	0.22	19.89
4	Stanowisko (własna klasyfikacja)	0.18	16.24
5	Ogłoszenie sprzed 2020	0.02	2.10
6	Stanowisko (wg rozporządzenia)	0.02	1.82
7	Liczba wymagań	0.02	1.82
8	Wskaźnik migracji (w powiecie)	0.01	0.82
9	Odsetek zatrudnionych na umowę (w instytucji)	0.01	0.82
10	Ogłoszenie po 2020	0.01	0.64

Wyjaśnienia: Zastosowany model w około 75% poprawnie przewidział wyniki naboru. Ważność badana z wykorzystaniem metod permutacyjnych (cf. Fisher, Rudin i Dominici 2019).

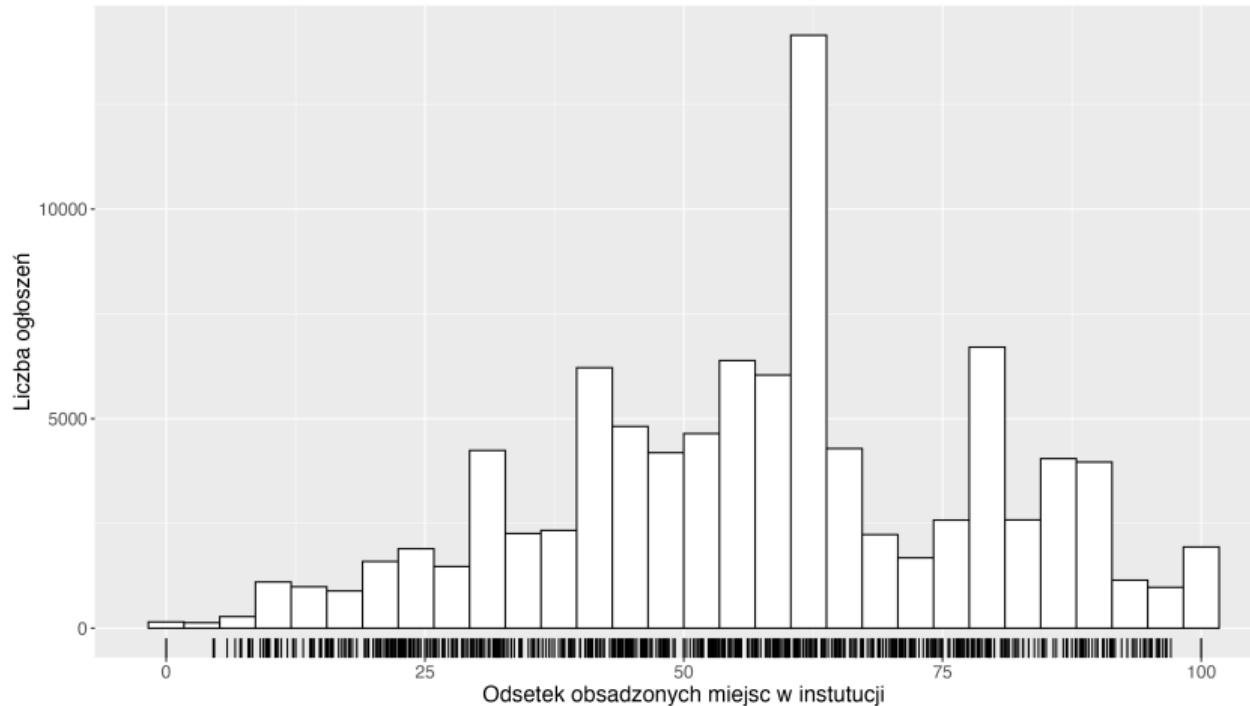
Najważniejsze wnioski

Najważniejsze wnioski (2)

- Najważniejszy czynnik wyjaśniający sukces lub porażkę rekrutacji: **efekt instytucji**.
- Co to znaczy? **Oznacza to, że o sukcesie rekrutacji w największym stopniu odpowiadają nieobserwowane charakterystyki podmiotu** (np. konkretny dział/departament, renoma, zespół, kierownik).
- Kolejne czynniki: wymagany poziom wykształcenia, kategoria zawodu (nasza własna klasyfikacja) oraz powiat (lokalny rynek pracy).
- Wynikiem jest również model klasyfikacji i dekompozycji prawdopodobieństwa obsadzenia stanowiska.

Najważniejsze wnioski

Odsetek obsadzonych stanowisk w instytucjach



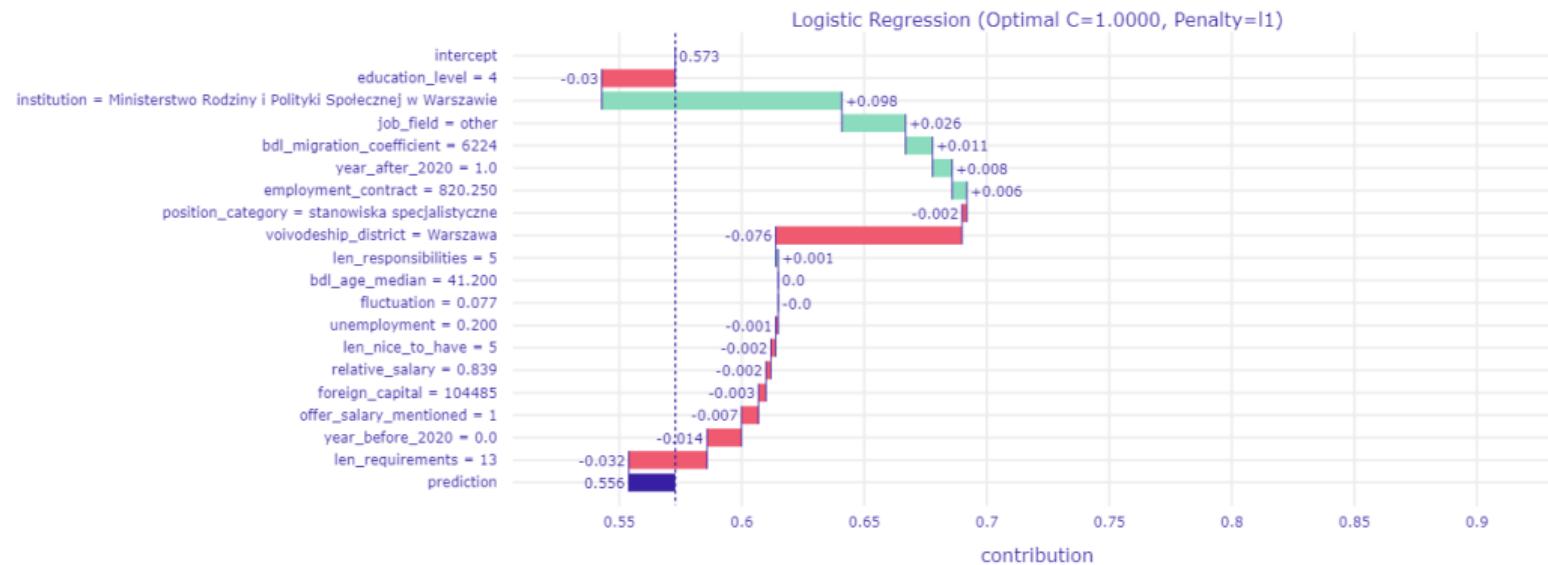
Najważniejsze wnioski

Top 10 instytucji wg liczby ogłoszeń

Lp.	Instytucja	# Ogłoszenia	% obsadzonych
1	Izba Administracji Skarbowej w Warszawie	5 962	62.09
2	Generalna Dyrekcja Dróg Krajowych i Autostrad	3 503	63.49
3	Izba Administracji Skarbowej w Poznaniu	1 651	82.13
4	Dolnośląski Urząd Wojewódzki we Wrocławiu	1 523	32.17
5	Ministerstwo Finansów w Warszawie	1 522	54.01
6	Izba Administracji Skarbowej w Gdańsku	1 519	78.87
7	Izba Administracji Skarbowej w Katowicach	1 400	78.50
8	Izba Administracji Skarbowej we Wrocławiu	1 378	87.95
9	Mazowiecki Urząd Wojewódzki w Warszawie	1 323	60.09
10	Łódzki Urząd Wojewódzki w Łodzi	1 304	39.95

Dekompozycja predykcji

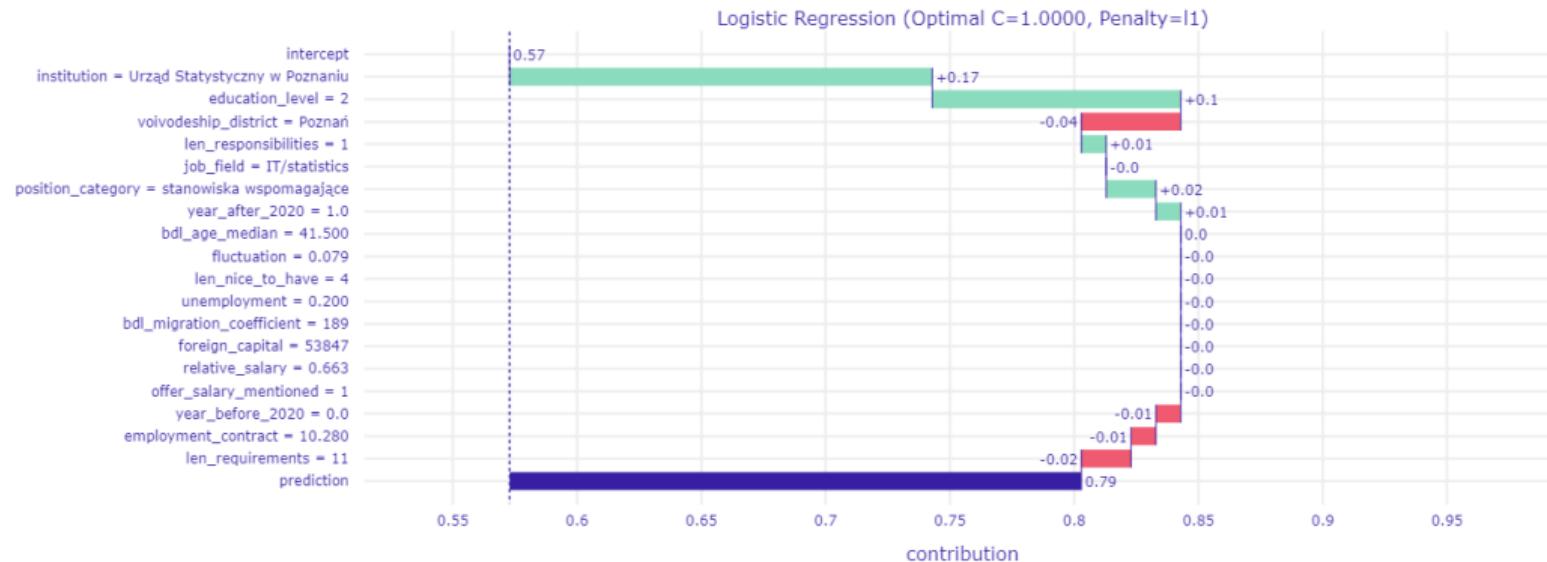
Dekompozycja – Ministerstwo Rodziny, Pracy i Polityki Społecznej



Objaśnienia: kolor czerwony – spadek prawdopodobieństwa obsadzenia miejsca pracy (w p.p.), zielony – wzrost, niebieski – ostateczna predykcja. Najwyższy wzrost to instytucja (9.8 p.p.), potem spadek o 7.6 p.p. przez Warszawę.

Dekompozycja predykcji

Dekompozycja – Urząd Statystyczny w Poznaniu



Objaśnienia: Najwyższy wzrost to instytucja (17 p.p), potem wykształcenie 10 p.p i spadek przez Poznań o 4 p.p.

Dekompozycja predykcji

Najważniejsze wyniki

- Zbiór danych – dostępny na RepOD:
<https://repol.icm.edu.pl/dataset.xhtml?persistentId=doi:10.18150/Q23MGT>.
- Klasyfikator – skuteczność około 75% poprawnych klasyfikacji.
- Wyjaśnianie klasyfikacji – dekompozycja.

Plan prezentacji

1 Wprowadzenie

2 Dane – pozyskanie i przygotowanie

3 Zaproponowane podeście

4 Wyniki (wybrane)

5 Podsumowanie

Dane – web-scraping

Znalezionych wyników: **146080**

ILOŚĆ WYNIKÓW NA STRONIE 5 10 15 20

SORTUJ:

OD NAJNOWSZYCH

OD NAJSTARSZYCH

NR 153343

młodzy specjalista

URZĄD: GENERALNA DYREKCJA OCHRONY ŚRODOWISKA W WARSZAWIE
MIEJSKOŚĆ: WARSZAWA
PUBLIKACJA OGŁOSZENIA: 27.06.2025

NR 155215

specjalista

URZĄD: BIURO RZECZNIKA PRAW PACJENTA W WARSZAWIE
MIEJSKOŚĆ: WARSZAWA
PUBLIKACJA OGŁOSZENIA: 24.06.2025

- Skrypt autorstwa Pana mgr Marcina Kostki – absolwenta Uniwersytetu Ekonomicznego w Poznaniu oraz byłego członka koła naukowego SKN Estymator.
- Skrypt przeszedł przez wszystkie podstrony, oferty pracy i wyniki naboru.
- Powstał zbiór danych, który w sposób ujednolicony gromadził wszystkie informacje z ofert pracy.

Przygotowanie danych do analizy

- Dane zostały poddane czyszczeniu: usunięto błędne ogłoszenia, ujednolicono nazwy podmiotów, miejsc pracy, wymiaru etatu (np. 0.5, 1/2).
- Przypisano ogłoszenia do konkretnych powiatów (identyfikatorów TERYT), a podmiotom przypisano identyfikatory NIP/REGON. Wykorzystano również informacje z o strukturze i hierarchii stanowisk w Służbie Cywilnej.
- Treści ogłoszeń zostały odpowiednio sklasyfikowane i utworzono wiele dodatkowych zmiennych, m.in. czy podano wynagrodzenie, liczbę wymagań, kategoria stanowiska czy poziom wykształcenia.
- Wykorzystano dodatkowe informacje na poziomie powiatów (m.in. mediana wieku, relacja pensji w instytucji do przeciętnego wynagrodzenia w powiecie, saldo migracji) oraz pozyskane z KPRM (m.in. rotacja zatrudnienia, liczba pracowników na umowę o pracę).
- Zdecydowano o usunięciu z danych przekazywanych modelowi podmioty, które w latach 2018-2023, opublikowały mniej niż 10 ogłoszeń.

Plan prezentacji

1 Wprowadzenie

2 Dane – pozyskanie i przygotowanie

3 Zaproponowane podeście

4 Wyniki (wybrane)

5 Podsumowanie

Zaproponowane podejście

- Zmienna celu:

$$y_{it} = \begin{cases} 1, & \text{jeśli ogłoszenie zakończyło się obsadzeniem stanowiska,} \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

gdzie obsadzenie: *informacja o zatrudnieniu kandydatki/kandydata, nabór zakończony wyborem kandydatki/kandydata oraz nabór zakończony zatrudnieniem kandydatki/kandydata.*

- Czynniki wyjaśniające: instytucja, powiat, wykształcenie, stanowisko, charakterystyki powiatu, informacje z ogłoszenia.
- Zastosowano regresję LASSO wraz z badaniem ważności (tzw. permutacyjne badanie ważności).
- Wykorzystano wyłącznie narzędzia dostępne w otwartej licencji (w języku R i Python).

Plan prezentacji

1 Wprowadzenie

2 Dane – pozyskanie i przygotowanie

3 Zaproponowane podeście

4 Wyniki (wybrane)

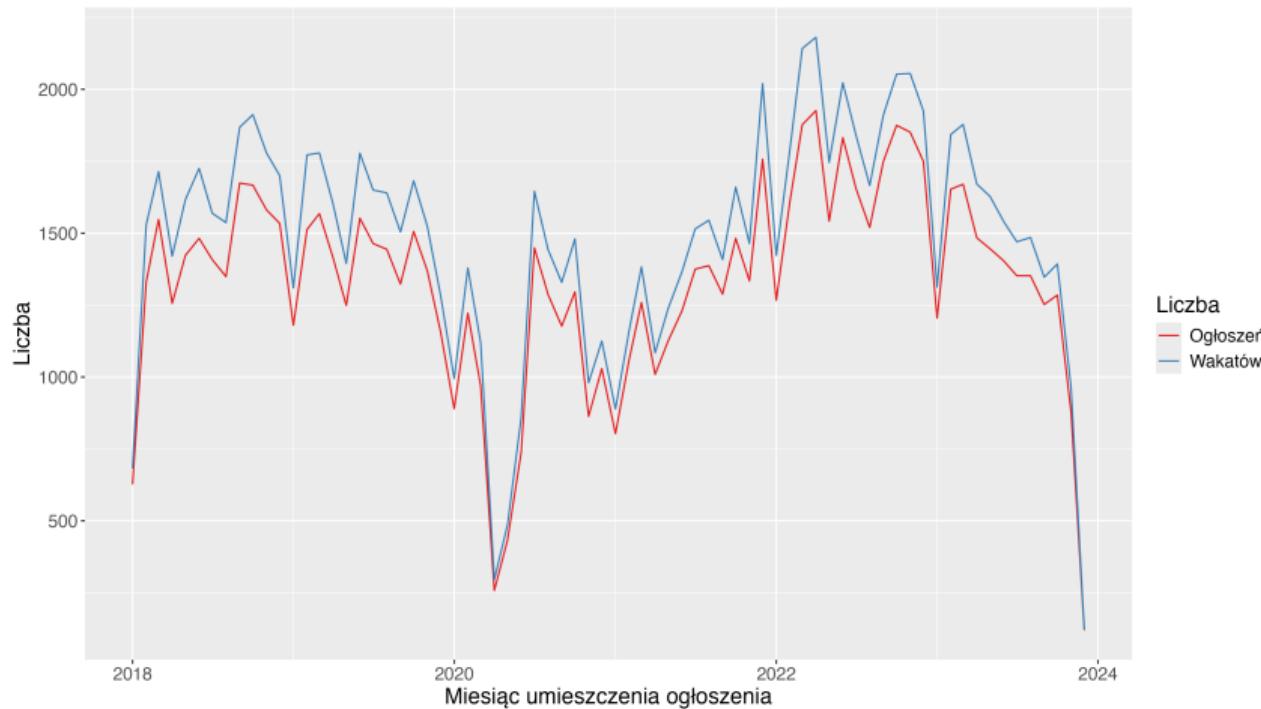
5 Podsumowanie

Liczba ogłoszeń i wakatów według wyniku naboru

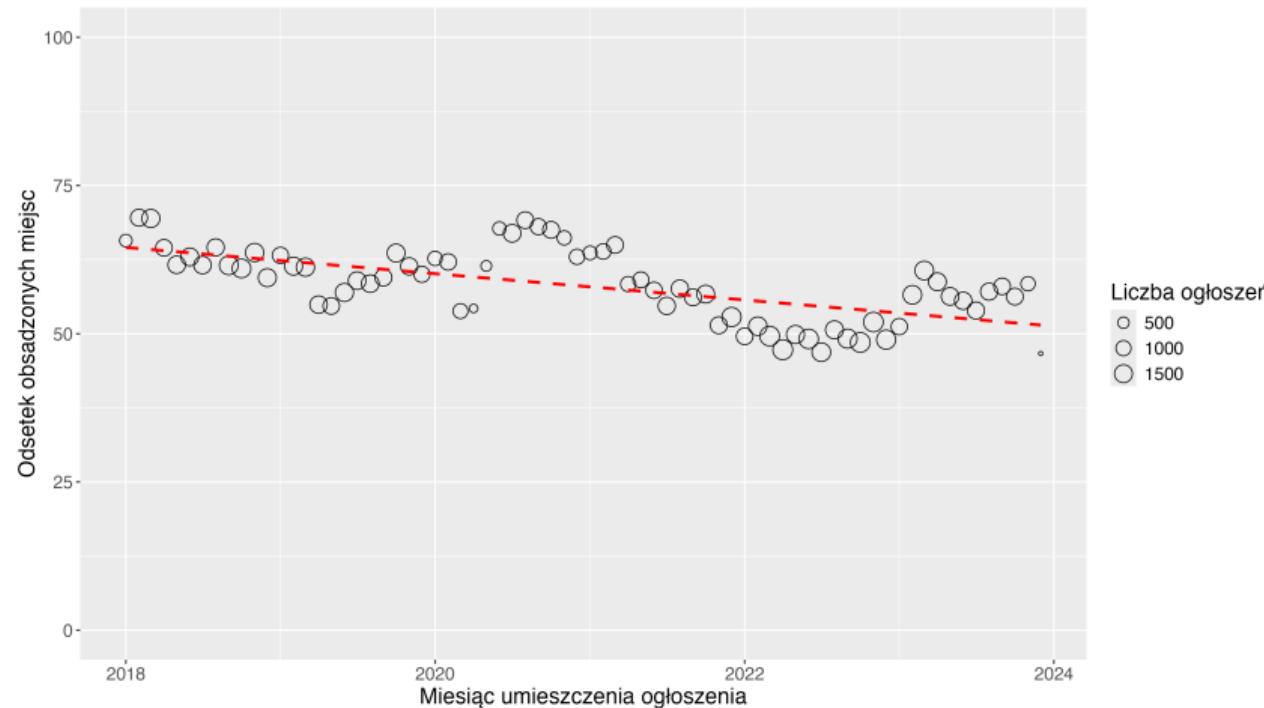
Wynik naboru	Ogłoszenia	Wakaty
Zatrudniono kandydata	58 658	67 013
Nie zatrudniono kandydata	41 920	45 284
Nabór anulowano	5 813	6 262
Rezygnacja kandydata	868	957
inne	1	2

Objaśnienia: *zatrudniono* – nabór zakończony wyborem / informacja o zatrudnieniu / nabór zakończony zatrudnieniem; *nie zatrudniono* – nabór zakończony bez wyboru / nabór zakończony bez zatrudnienia / nie zatrudniono / brak ofert / nie wyłoniono najlepszych / nie spełniały wymagań formalnych; *rezygnacja kandydata* – decyzja kandydatki/kandydata, rezygnacja kandydata.

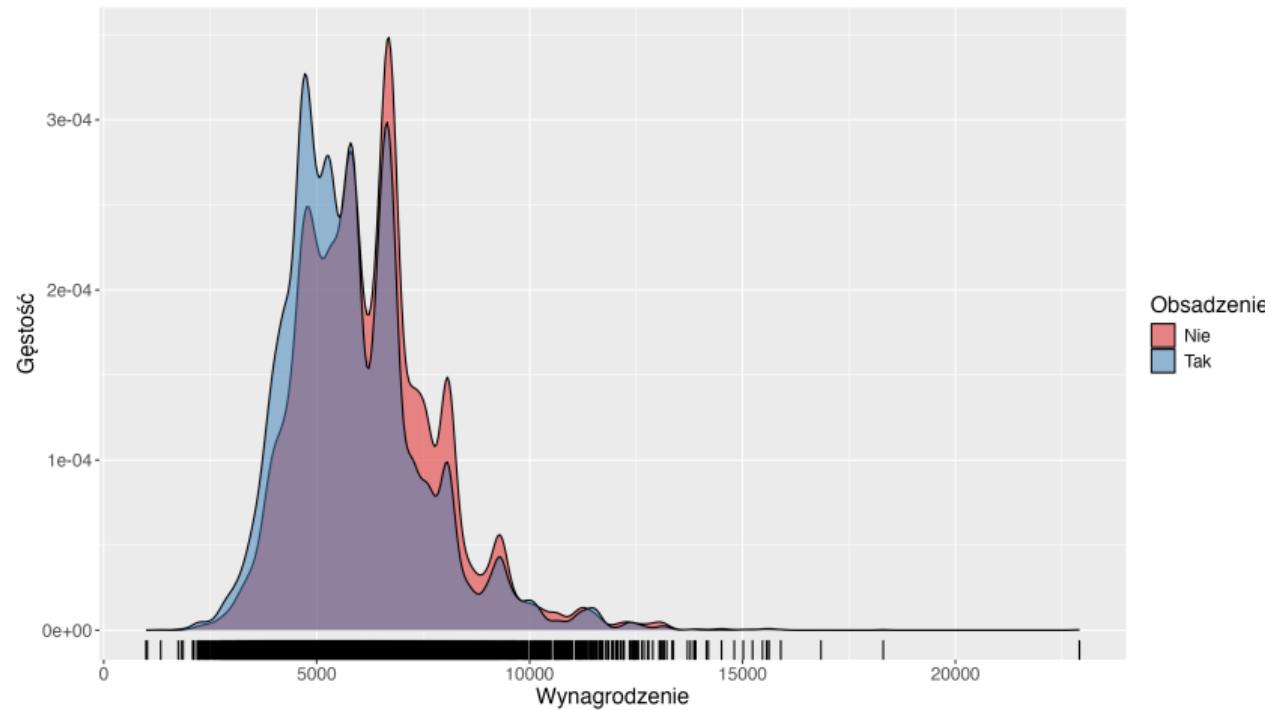
Liczba ogłoszeń i wakatów



Odsetek obsadzonych stanowisk



Odsetek obsadzonych stanowisk w wynagrodzeniu



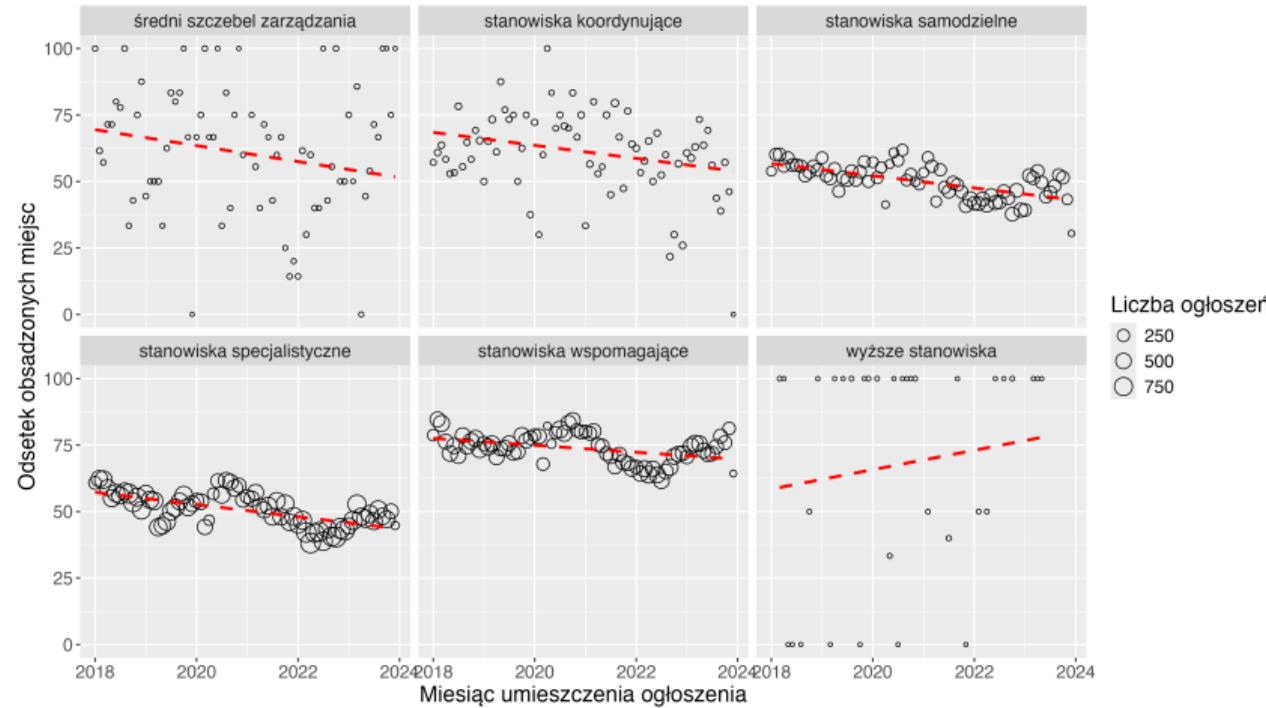
Stanowiska według podanej nazwy (top 10)

Lp.	Stanowisko	# Ogłoszenia	% obsadzonych
1	referent	14026	79.58
2	specjalista	11474	53.51
3	inspektor weterynaryjny	9058	24.61
4	starszy specjalista	8527	45.78
5	inspektor	8464	71.43
6	główny specjalista	5575	47.80
7	starszy inspektor	4512	63.90
8	inspektor wojewódzki	4179	43.79
9	starszy referent	3969	82.69
10	referendarz	2568	64.33

Stanowiska według grup określonych w rozporządzeniu

Kategoria stanowiska	# Ogłoszenia	% obsadzonych
Stanowiska specjalistyczne	44 618	50.30
Stanowiska wspomagające	30 820	73.76
Stanowiska samodzielne	18 656	49.68
Stanowiska koordynujące	1 309	60.89
Średni szczebel zarządzania	446	60.99
Wyższe stanowiska	50	68.00

Odsetek obsadzonych stanowisk wg kategorii

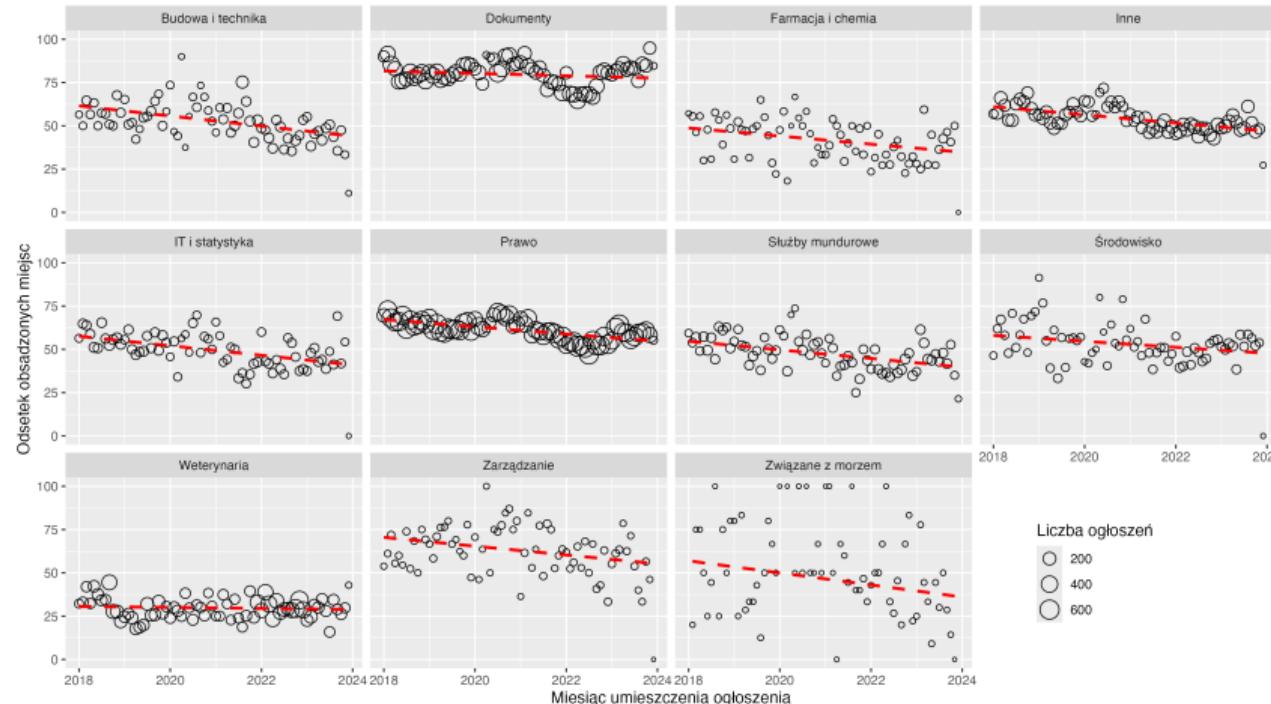


Stanowiska według słów nazw i słów kluczowych

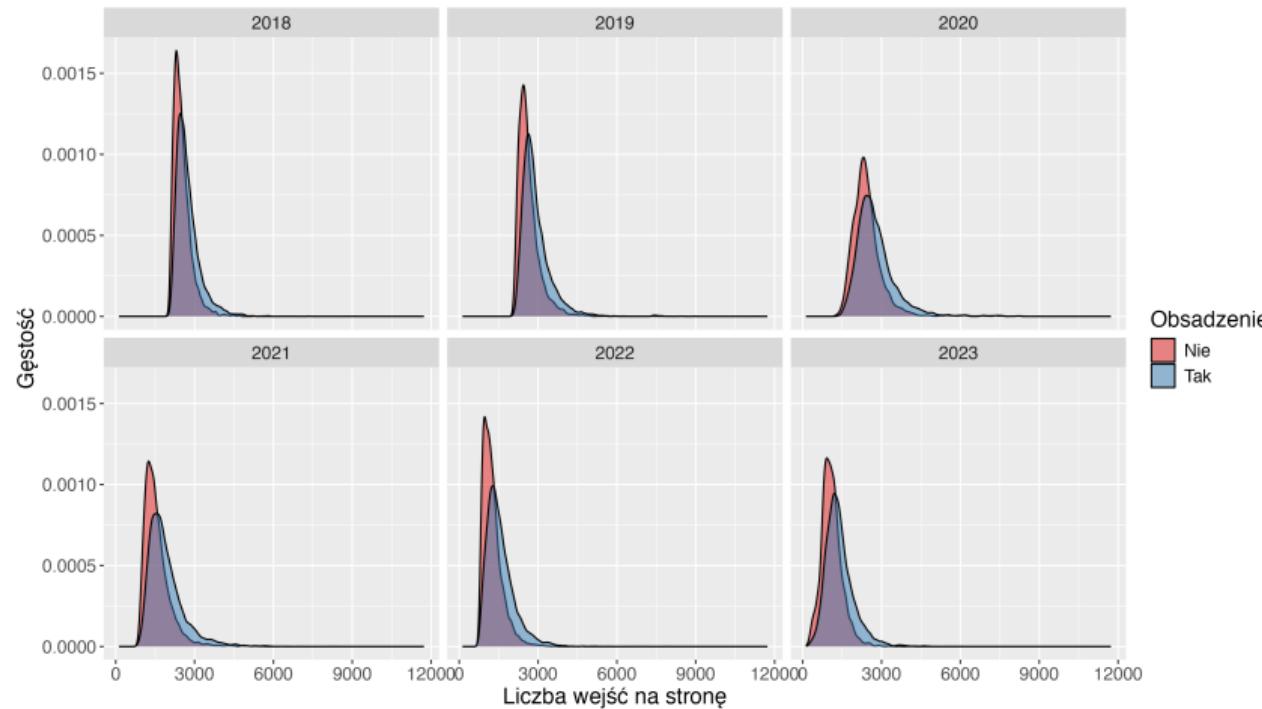
Kategoria	# Ogłoszenia	% obsadzonych
Prawo	33 279	61.05
Dokumenty	19 528	79.70
Weterynaria	11 711	29.70
Służby mundurowe	4 583	47.24
IT i statystyka	4 366	49.95
Budowa i technika	4 091	52.16
Środowisko	3 234	52.44
Farmacja i chemia	1 641	41.26
Zarządzanie	1 399	62.76
Związane z morzem	388	45.36
<i>Inne</i>	11 679	53.79

Objaśnienia: kategorie przypisane na podstawie słów kluczowych.

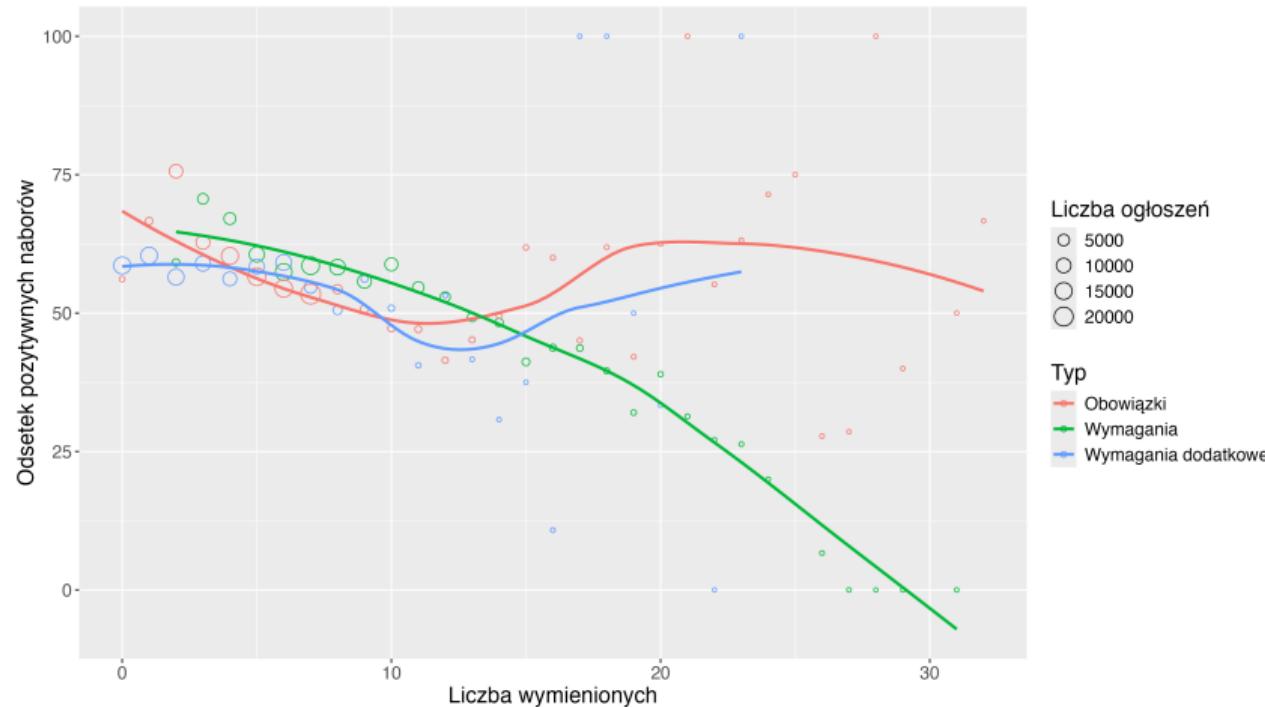
Odsetek obsadzonych stanowisk wg kategorii



Liczba ogłoszeń i wakatów



Liczba wymagań, a sukces naboru



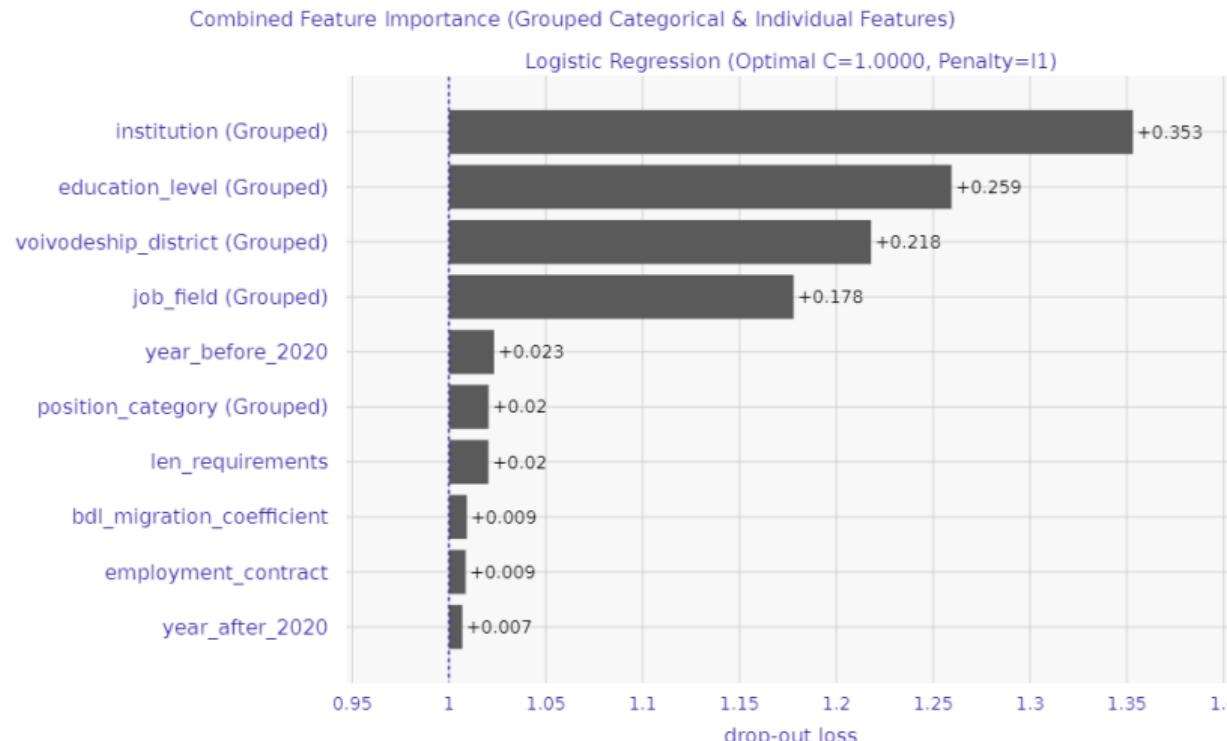
Odsetek obsadzonych według poziomów wykształcenia

Poziom wykształcenia	# Ogłoszenia	% obsadzonych
Średnie	7 190	57.52
Średnie specjalistyczne	34 903	77.91
Wyższe	31 807	37.89
Wyższe specjalistyczne	21 999	55.30

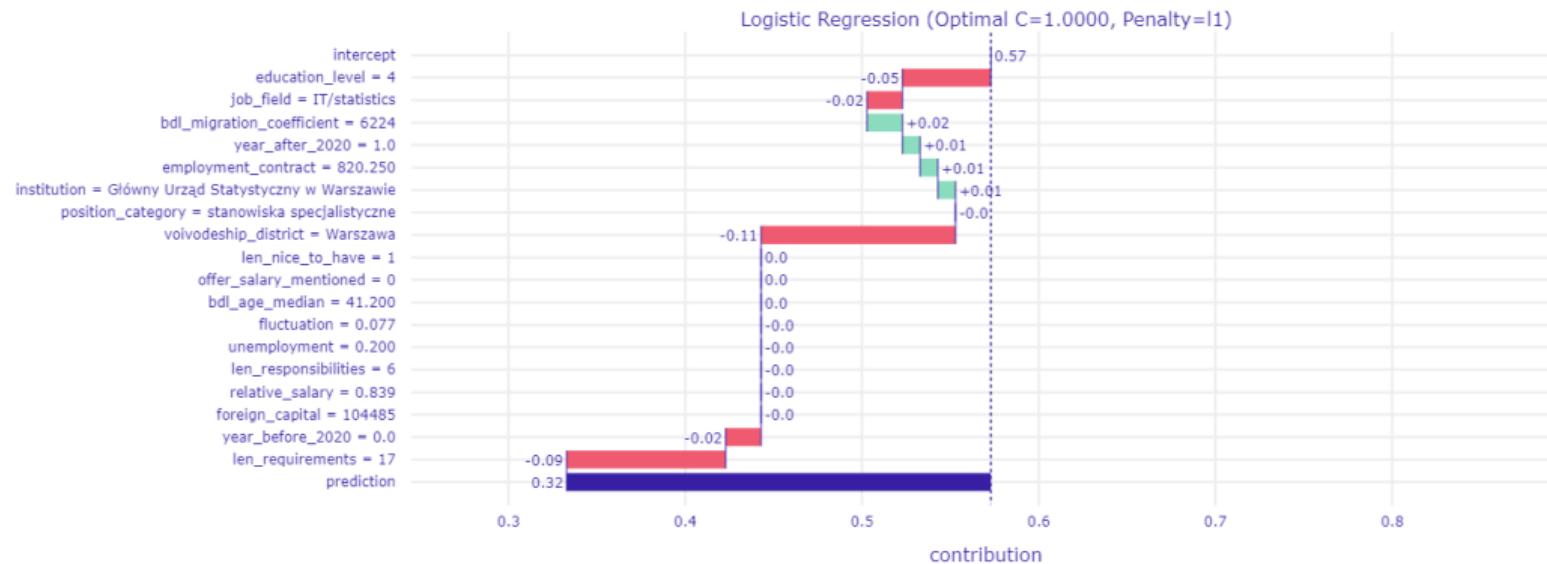
Odsetek obsadzonych według powiatów (top 10)

Lp.	TERYT	Powiat	# Ogłoszenia	% obsadzonych
1	1465	Warszawa	30845	52.23
2	0264	Wrocław	3841	54.96
3	1061	Łódź	3533	55.22
4	2261	Gdańsk	3291	55.21
5	3064	Poznań	3168	64.43
6	3262	Szczecin	2723	54.87
7	1261	Kraków	2436	74.47
8	2469	Katowice	2309	59.12
9	1661	Opole	1878	54.74
10	0461	Bydgoszcz	1788	62.36

Permutacyjna ważność zmiennych wg regresji LASSO



Dekompozycja – Główny Urząd Statystyczny



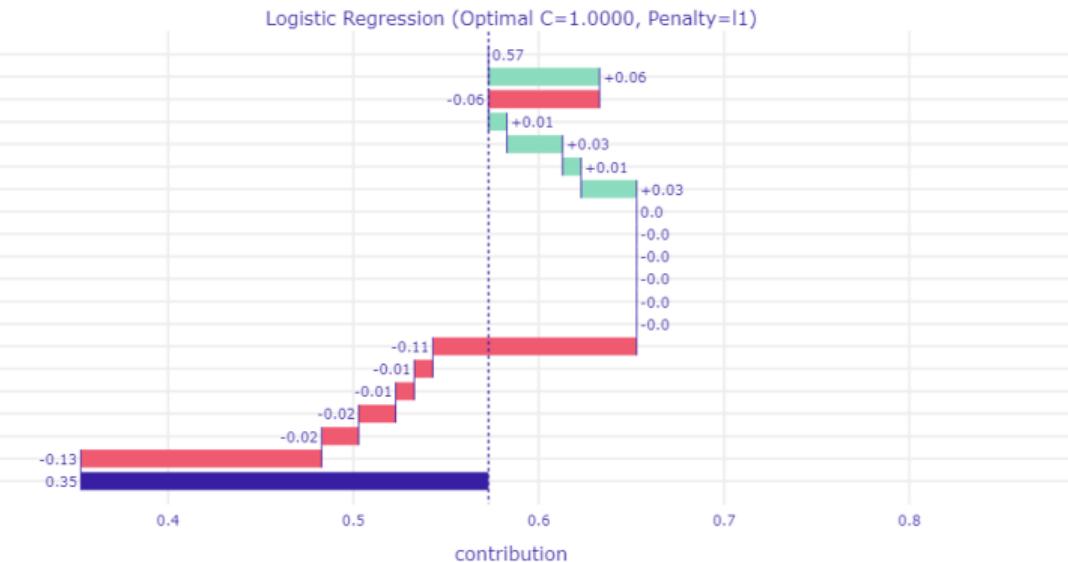
Objaśnienia: spadek o 5 p.p. ze względu na wykształcenie (wyższe specjalistyczne), o 9 p.p. przez liczbę wymagań i o 11 p.p. przez Warszawę.

Dekompozycja – Wielkopolski Urząd Wojewódzki w Poznaniu

```

intercept
job_field = other_manager
voivodeship_district = Poznań
relative_salary = 1.366
position_category = stanowiska koordynujące
year_after_2020 = 1.0
institution = Wielkopolski Urząd Wojewódzki w Poznaniu
bdl_age_median = 41.500
fluctuation = 0.062
unemployment = 0.200
bdl_migration_coefficient = 189
foreign_capital = 53847
employment_contract = 91.730
education_level = 3
len_nice_to_have = 7
offer_salary_mentioned = 1
year_before_2020 = 0.0
len_responsibilities = 10
len_requirements = 18
prediction

```



Objaśnienia: spadek o 6 p.p. przez Poznań, o 11 p.p. ze względu na wykształcenie (wyższe) i o 13 p.p. przez liczbę wymagań.

Plan prezentacji

1 Wprowadzenie

2 Dane – pozyskanie i przygotowanie

3 Zaproponowane podeście

4 Wyniki (wybrane)

5 Podsumowanie

Podsumowanie

- Zgodnie z moją wiedzą była to pierwsza praca magisterska w Polsce analizująca oferty z pracy ze Służby Cywilnej.
- W tej pracy nie podejmowano "wyciągania" z treści ogłoszeń informacji o konkretnych umiejętnościach czy wymaganiach (np. zgodnych z klasyfikacją ISCO).
- Również nie badano ponownego umieszczania tego samego ogłoszenia (tj. każde ogłoszenie traktowano jako oddzielne).
- Wynikiem pracy jest model umożliwiający klasyfikację naborów do udanych i zakończonych niepowodzeniem, a także wskazanie czynników, które wyjaśniają predykcję dla poszczególnych ogłoszeń.
- Prace dalej będą kontynuowane przez innych magistrantów.
- Moją pracę w ramach tej pracy magisterskiej wspierało Narodowe Centrum Nauki (OPUS 27 grant no. 2024/53/B/HS4/01580).

Program „Nauka dla Rozwoju Społeczeństwa” – propozycja

Informacje o programie:

- Budżet: 180 mln zł.
- Dofinansowanie: do 2 mln zł na projekt.
- Okres: do 36 miesięcy.
- Nabór: tryb ciągły od czerwca 2025.
- Status: program uruchomiony.

Potencjał dla badań rynku pracy:

- Obszar „Nauka dla innowacyjności” - współpraca nauki z otoczeniem gospodarczym.
- Obszar „Nowe horyzonty nauki” - przełomowe badania społeczno-gospodarcze.
- Obowiązkowa współpraca z partnerem społeczno-gospodarczym.
- Nacisk na implementację rozwiązań aplikacyjnych.

Dziękuję za uwagę!

- maciej.beresewicz@ue.poznan.pl
- Katedra Statystyki, Uniwersytet Ekonomiczny w Poznaniu
- Wykresy i tabele: <https://github.com/OJALAB/2025-kprm-talk>