

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



Metody i Systemy Decyzyjne

Raport - "Co wpływa na wynagrodzenie w branży IT?"

AUTOR

Michał Puchyr

nr albumu: **272733**

kierunek: **Informatyka Stosowana**

4 czerwca 2024

1 Problem badawczy

Problem badawczy: Jakie czynniki najbardziej wpływają na zarobki w branży IT w Polsce?

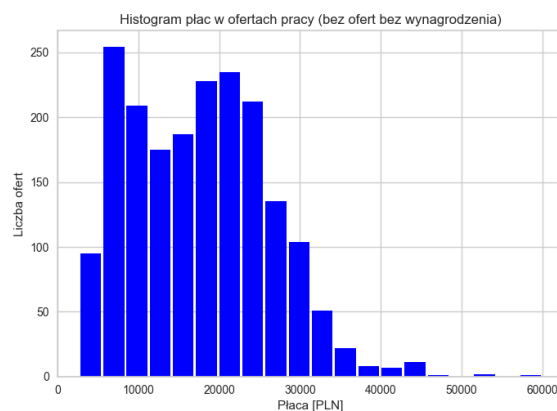
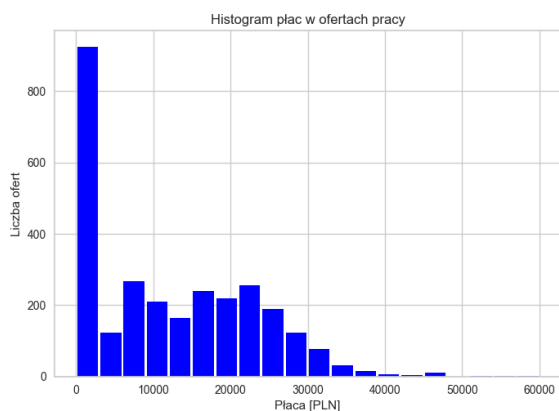
Celem niniejszego raportu jest zbadanie czynników wpływających na zarobki w branży IT. Rozwój technologii informatycznych sprawia, że specjaliści IT są jednymi z najbardziej poszukiwanych pracowników na rynku pracy. W związku z tym że, zarobki informatyków są bardzo zróżnicowane i zależne od wielu czynników, w niniejszym raporcie zostaną przedstawione wyniki badań dotyczące wynagrodzenia informatyków w Polsce oraz czynników, które wpływają na ich wysokość.

2 Dane badawcze

Dane potrzebne do przeprowadzenia analizy zostały pobrane z portali NoFluffJobs oraz Pracuj.pl.

- Nazwa oferty
- Nazwa firmy (pracodawcy)
- Technologie (języki programowania i narzędzia)
- Poziom doświadczenia (trainee, junior, mid, senior, expert)
- Lokalizacja
- Zarobki (widełki płacowe - minimalne i maksymalne)
- Czy praca jest zdalna

Do celów analitycznych zostało pobranych 2859 ofert pracy. Niestety nie wszystkie oferty zawierały informacje o zarobkach, dlatego zostały one odfiltrowane.



W procesie czyszczenia danych, oferty pracy, które nie zawierały informacji o wynagrodzeniu, miały wartości zerowe w kolumnie zarobków. **Brak danych o zarobkach występuje w około 32% ofertach pracy.**

Lokalizacje ofert pracy, których liczba wystąpień była mniejsza bądź równa 5, zostały zgrupowane do jednej kategorii - 'Other' tak aby nie wpłynęły na czytelność wykresów i analizy danych. Lokalizacje z za granicę zostały zgrupowane do jednej kategorii - 'Abroad'.

Poziomy doświadczenia z ofert pracy z portalu Pracuj.pl zostały przekształcone na poziomy zgodne z ofertami z portalu NoFluffJobs w celu ujednolicenia danych. Poziomy doświadczenia zostały przekształcone na: *trainee*, *junior*, *mid*, *senior*, *expert*.

Nie każda oferta zawierała informacje o języku naturalnym oraz o technologii wymaganej do pracy. Wartości takie w bazie danych posiadają wartość null. W przypadku użycia kodowania one-hot, wartości null są ignorowane.

Oferty pracy z brakiem informacji o wynagrodzeniu zostały odzucone w dalszej analizie ze względu na brak możliwości ich uzupełnienia i wykorzystania do modelowania.

ID	Name	SalaryFrom	SalaryTo	ExpLevel	Category	JobLanguage	Location	IsRemote	Technology	Company
125	Ework Group	18480.00	29400.00	Senior	erp		Kraków	false		Ework Group
126	ROCKWOOL	0.00	0.00	Mid	support		Poznan	false		ROCKWOOL
127	Omni Calculator	7300.00	9600.00	Mid	marketing		Remote	true		Omni Calculator
150	Devire Sp. z o.o.	28560.00	35280.00	Mid	backend	pl, en	Warszawa	false	Java	Devire Sp. z o.o.
151	ITDS	14700.00	23100.00	Senior	fullstack	pl	Warszawa	false	.NET	ITDS

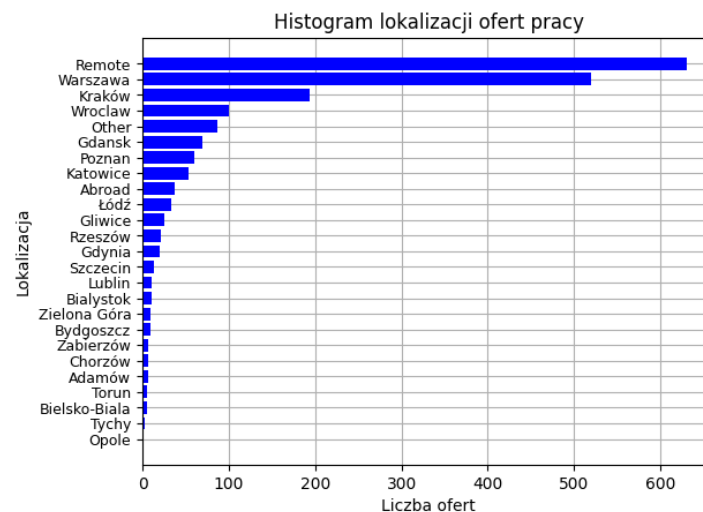
Tabela 1: Przykładowy wycinek danych na podstawie których przeprowadzono analizę

3 Analiza danych

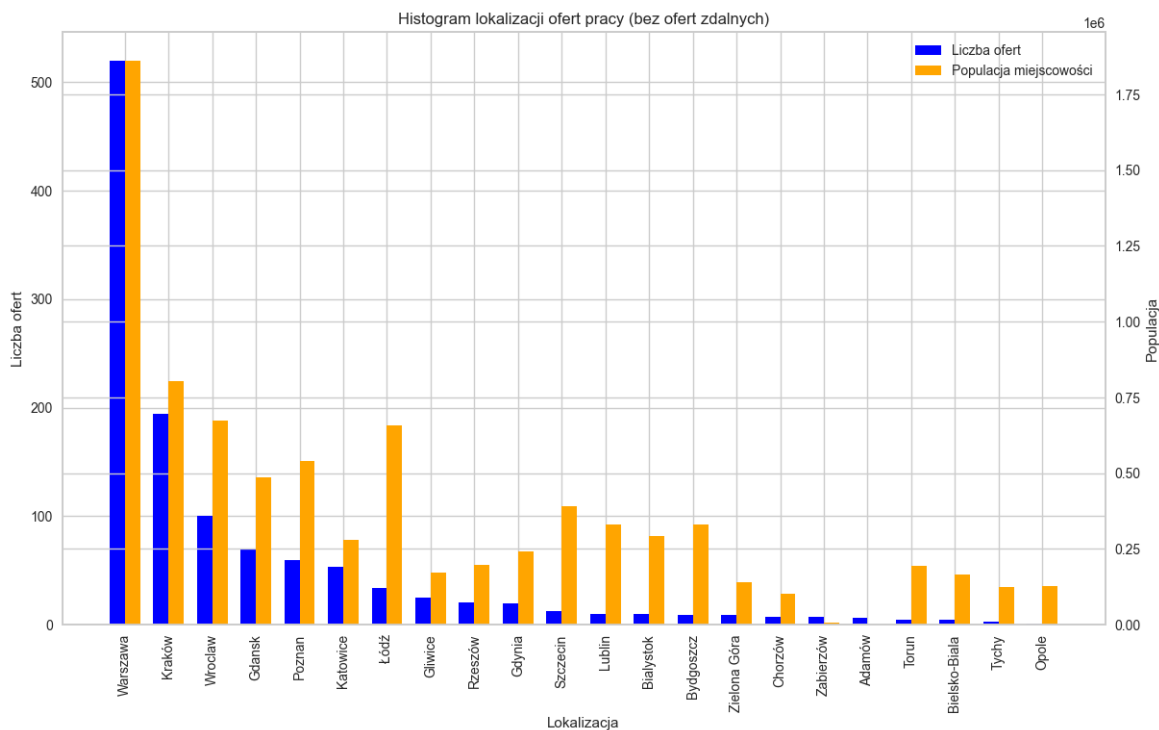
	Średnia	Mediana	Odchylenie std.	Min	25%	50%	75%	Maks
Zarobki [PLN]	17637.34	17500.00	8578.94	2664.00	10000.00	17500.00	23520.00	60000.00

Tabela 2: Statystyki opisowe zarobków programistów

Mediana wysokości płacy analizowanych ofert pracy wynosi 17 500 PLN, a średnia 17 637.34 PLN. Odchylenie standardowe wynosi 8 578.94 PLN, co oznacza, że zarobki w branży IT są bardzo zróżnicowane. Najwięcej ofert pracy możemy znaleźć w przedziale od 7 500 PLN do 23 520 PLN. Z wykresu możemy również odczytać, że oferty pracy z wynagrodzeniem powyżej 35 000 PLN są rzadkością. Możemy zatem przyjąć, że zarobki w branży IT mieszczą się w przedziale od 3 000 PLN - 40 000 PLN.



Najpopularniejszą "lokalizacją" pracy dla programistów jest praca zdalna. Praca zdalnych jako lokalizacja występuje w 32% ofert pracy. Jej wpływ na wynagrodzenie zostanie zbadany w dalszej części raportu. W przypadku lokalizacji pracy stacjonarnej najwięcej ofert pracy pochodzi z odpowiednio z Warszawy, Krakowa, Wrocławia, Gdańska i Poznania co pokrywa się z wielkością tych miast pod względem liczby mieszkańców. Do raportu zostały również uwzględnione oferty pracy zza granicy, liczba takich ofert w porównaniu do ofert z Polski jest znikoma, stanowią one zaledwie 2% wszystkich ofert, w związku z czym zostały one wrzucone do jednej kategorii - "Abroad".

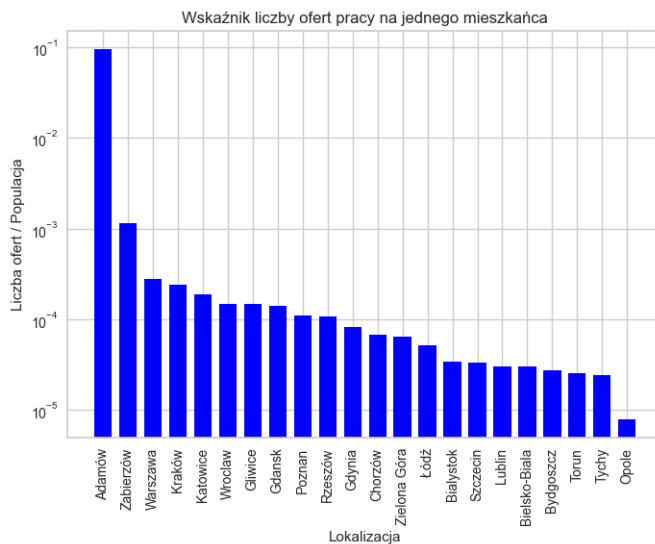


Jeśli weźmiemy pod uwagę do lokalizacji ofert liczbę mieszkańców w danym mieście możemy zauważyć, że liczba mieszkańców nie zawsze będzie oznaczać większą liczbę ofert pracy.

Punktem odniesienia jest Warszawa, gdzie słupek z liczbą ofert pracy jest na równi ze słupkiem liczby mieszkańców. Proporcja wysokości słupków liczby ofert pracy i ludności dla reszty miast jest odniesiona bezpośrednio do Warszawy.

Wiedząc to, możemy wyczytać, że Wrocław jest miastem, które jak na swoją liczbę mieszkańców powinno mieć więcej ofert pracy tak aby dorównać Warszawie przy swojej liczbie mieszkańców.

Aby ułatwić analizę ludności i liczby ofert możemy wyliczyć proporcje liczby ofert pracy do liczby mieszkańców. Wartość ta pozwoli na zobrazowanie, w którym mieście jest największe zagęszczenie ofert pracy.

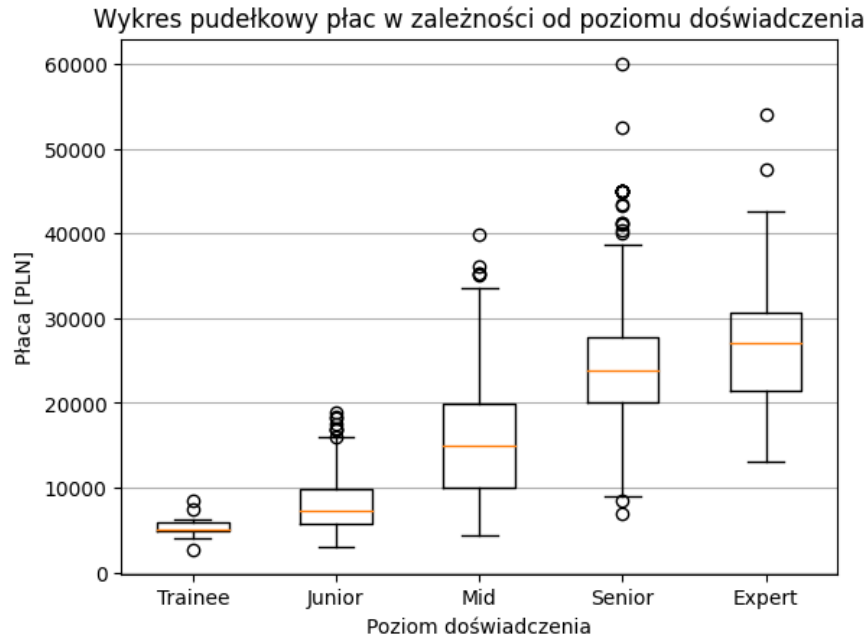


Najlepiej mają programiści z Adamowa. Należy jednak wziąć pod uwagę, że Adamów jak i Zabierzów są bardzo małymi miejscowościami z ulokowanymi w nich firmami informatycznymi ze względu na bliskość do metropolii odpowiednio warszawskiej i krakowskiej.

Do wizualizacji wskaźnika zastosowano skalę logarytmiczną tak aby przypadek Adamowa i Zabierzowa nie zaburzał czytelności wykresu.

3.1 Wstępne oszacowanie

Wstępnie możemy dokonać analizy zarobków w zależności od potencjanych czynników.



Poziom doświadczenia	Średnie wynagrodzenie [PLN]
Expert	26890.65
Senior	24242.49
Mid	15429.12
Junior	8154.89
Trainee	5232.71

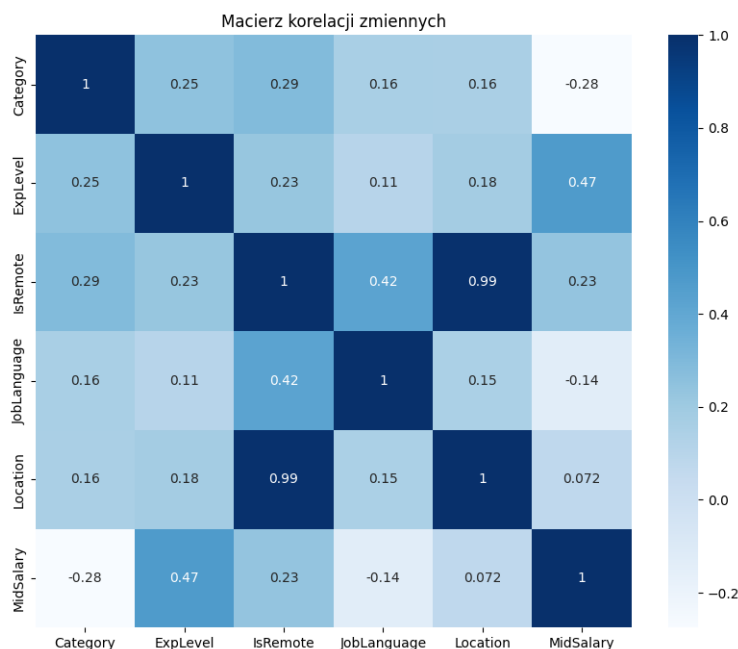
Z wykresu wynika, że poziom doświadczenia jest wyraźnie skorelowany z wysokością wynagrodzenia co jest zgodne z intuicją. Im bardziej doświadczony pracownik, tym większą wartość ma jego praca. Na największe zarobki mogą liczyć programiści klasyfikujący się jako eksperci zaś najmniej jako stażyści i juniorzy.

Kategoria	Średnie wynagrodzenie [PLN]
Sztuczna inteligencja	27135.55
Architektura kodu	25766.50
Analiza danych	23632.85
Agile / Scrum	22725.00
Backend	21734.63
...	...
HR	10400.39
Marketing	8365.63
Prawo	8186.00
Administracja biurowa	7807.85
Obsługa klienta	7166.08

Zarobki od 5-tego miejsca od góry do 5-tego miejsca od dołu zostały pominięte w tabeli dla czytelności.

Na najwyższe zarobki przeciętnie mogą liczyć programiści związani z dziedziną sztucznej inteligencji, architekturą oraz danymi. Najniższe zarobki przeciętnie otrzymują pracownicy związani z obszarami obsługą klienta, administracją biurową oraz prawem.

3.2 Korelacja między zmiennymi



Przy pomocy macierzy korelacji możemy zidentyfikować zmienne, między którymi istnieje pewna zależność. Z mapy wynika, że istnieje korelacja między poziomem doświadczenia i wynagrodzeniem. Jest to zgodne z ogólnie przyjętą intuicją, że im wyższy poziom doświadczenia, tym wyższe zarobki.

Zauważalna jest również korelacja języka (naturalnego) a tym czy praca jest zdalna. Można spekulować, że wynika to z tego, że język angielski jest wymagany przy współpracy z firmami zagranicznymi, które nie mając siedziby w Polsce oferują pracę zdalną.

Korelacja między lokalizacją a pracą zdalną wynika z założenia analizy, że jeśli lokalizacja jest 'Remote' to praca jest zdalna.

Ewentualnymi słabszymi korelacjami, na które można zwrócić uwagę są również:

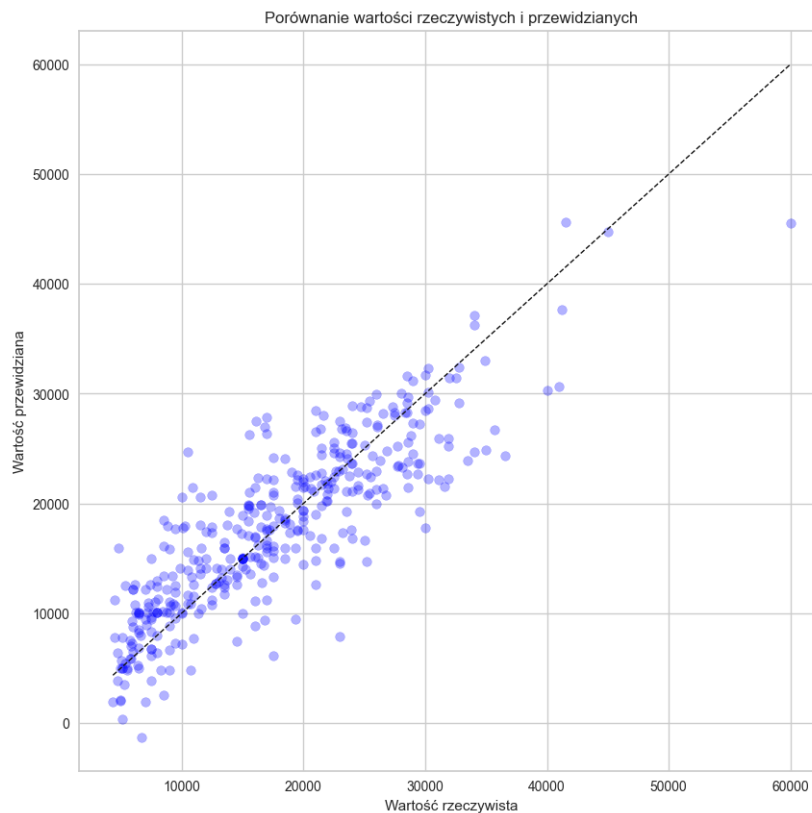
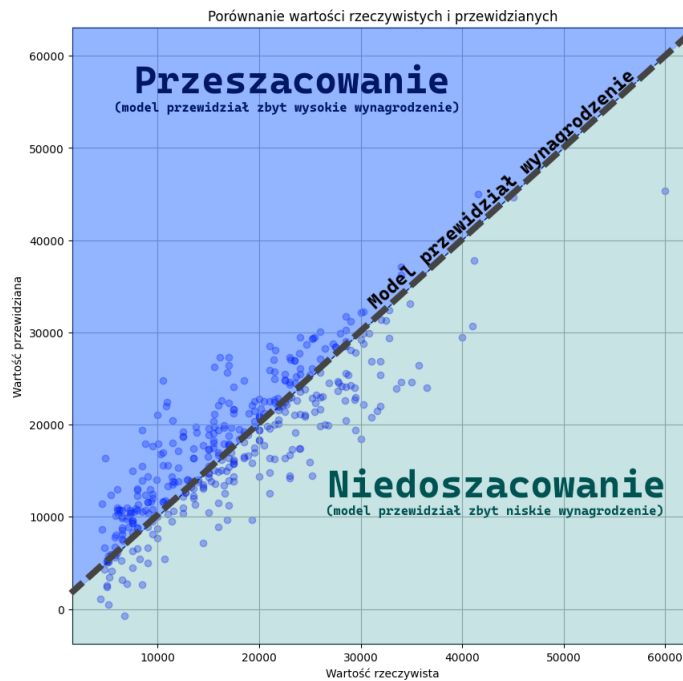
- Praca zdalna a zarobki
- Kategoria a praca zdalna
- Poziom doświadczenia a praca zdalna
- Kategoria a zarobki

4 Opracowanie modelu

Aby móc odpowiedzieć na pytanie o wpływ czynników na zarobki programistów, należy zbadać wagę poszczególnych zmiennych. W tym celu najlepiej sprawdzi się model regresji liniowej. Model regresji liniowej pozwala na przewidywanie wartości zmiennej zależnej na podstawie wartości innych zmiennych.

Poprzez analizę współczynników regresji, można określić, które zmienne mają największy wpływ na zarobki programistów. Dane zostały podzielone na zbiór treningowy i testowy w stosunku 80:20. Cechy katagoryczne zostały zakodowane za pomocą kodowania one-hot tak aby model mógł je zinterpretować jako zmienne

numeryczne. Wybór tego kodowania jest podyktowany tym, że zmienne kategoryczne używane w modelu nie mają charakteru porządkowego, model musi traktować je jako osobne kategorie. Aby ułatwić zrozumienie przewidywań modelu, niżej przedstawiono grafikę jak należy odczytywać wyniki modelu regresji liniowej na wykresie.



Pod uwagę przy tworzeniu modelu regresji liniowej wzięto następujące zmienne:

- Poziom doświadczenia
- Technologie
- Kategorię branży
- Czy praca jest zdalna
- Lokalizację miejsca pracy

Współczynnik determinacji modelu wynosi 0.73, oznacza to, że model, jest w stanie przewidzieć 73% zmienności wynagrodzenia przez zmienne objaśniające.

Z wykresu również można zauważyć, że dużo jest ofert pracy, których zarobki nie udało się w pełni przewidzieć. Różnica między wartościami przewidywanymi a rzeczywistymi w pewnych przypadkach jest bardzo duża - rzędu 10 000 złotych. Jedną z możliwych przyczyn takiego stanu rzeczy jest brak większej liczby danych w zbiorze bądź atrybutów, które pozwoliłyby wyjaśnić większą część zmienności zarobków programistów.

4.1 Ważność zmiennych

Chcąc rozwiązać problem przewidywania wynagrodzenia programistów należy zidentyfikować zmienne mające największy wpływ. W przypadku zidentyfikowania tych zmiennych ważną informacją będzie jej kontekst, czyli analiza na podstawie obserwacji świata rzeczywistego.

Czynnik	Znaczenie
(Firma) Snowflake	23646.26
(Firma) Technosource	17071.15
(Firma) KZ INSPIRE	16624.71
(Firma) Chorus One	15690.84
(Firma) Infopulse	15651.57
(Firma) Plenti	14813.50
(Firma) Harvey Nash Technology Sp. z o.o.	14320.10
(Firma) RunBit	12440.59
(Firma) vonRoll Infratec.com	12315.75
(Firma) Be in IT	12208.10
(Firma) Transition Technologies PSC S.A.	11571.51
(Firma) Directio Sp. z o.o.	11391.51
(Firma) First Derivative	11139.73
(Firma) Varwise	10754.25
(Firma) Wipro IT Services	10427.30
(Firma) int2code GmbH	10255.82
(Firma) Consult Red	10077.89
(Firma) Tesco Technology	10012.09
(Firma) 7N Sp. z o.o.	9839.15
(Poziom doświadczenia) Expert	9802.32

Tabela 3: 20 najważniejszych zmiennych w modelu regresji liniowej

Z tabeli wynika, że największy wpływ na zarobki programistów mają zmienne związane z firmą, z której pochodzi oferta pracy. Ważny przy analizie jest kontekst, gdyż firma z której pochodzi oferta nie zawsze implikuje wyższych zarobków z faktu, że jest akurat tą firmą.

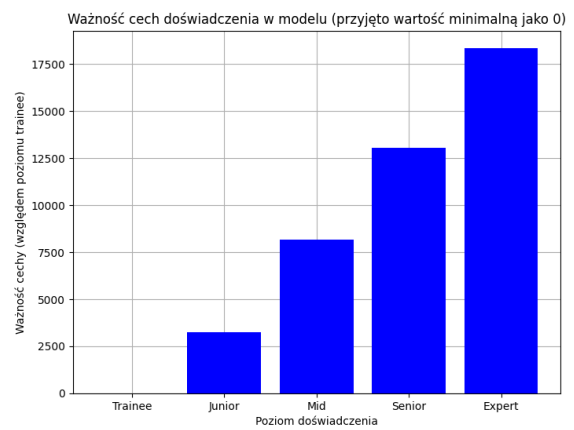
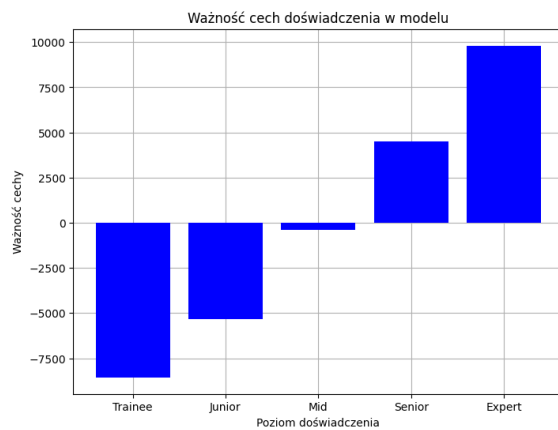
Duże firmy informatyczne zazwyczaj będą oferowały wysokie zarobki ze względu na swoją renomę, a co za tym idzie będą chciały przyciągnąć najlepszych pracowników z jak największym doświadczeniem. Warto mieć to na uwadze przy analizie wyników, chcąc zasugerować się modelem i zdobyć jak najlepiej płatną pracę osoba powinna się skupić na czynnikach, które są w większym stopniu od niej zależne.

W celu lepszego zrozumienia czynników wpływających na wygradzenie, atrybuty powinny zostać rozdzielone na podkategorie tak, aby móc zobaczyć najważniejsze z nich.

Poziom doświadczenia	Znaczenie
Expert	9802.32
Senior	4481.91
Mid	-394.09
Junior	-5332.76
Trainee	-8557.38

Tabela 4: Wpływ poziomu doświadczenia na zarobki programistów

Poziom doświadczenia jest cechą, która obejmuje każdego programistę. Jest cechą, na którą mamy największy wpływ i to ona zawsze będzie czynnikiem, która będzie wpływać na nasze zarobki. Aby móc lepiej wyciągnąć wnioski z tej cechy można ją zwizualizować w postaci wykresu.



Im wyższy poziom doświadczenia, tym na wyższe zarobki może liczyć programista. Warty zauważenia jest fakt, że wzrost wynagrodzenia względem poziomu doświadczenia jest liniowy.

Branża	Znaczenie
businessIntelligence	5867.84
security	4827.58
artificialIntelligence	4479.77
architecture	4054.83
projectManager	3925.77
devops	3510.55
erp	3001.52
embedded	2976.36
agile	2909.15
data	2818.46
frontend	2655.47
fullstack	2378.75
telecommunication	2152.71
backend	2098.76
businessAnalyst	1874.30
None	956.40
productManagement	846.45
gameDev	446.57
sysAdministrator	303.16
testing	50.51

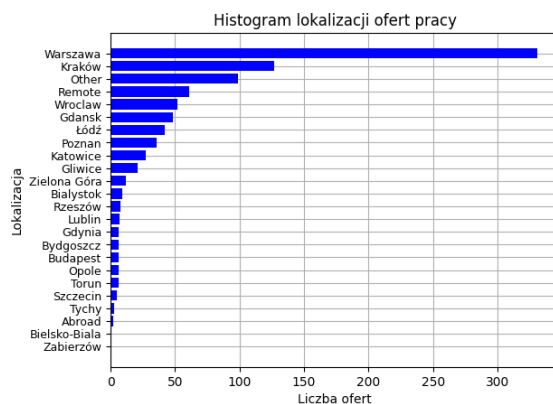
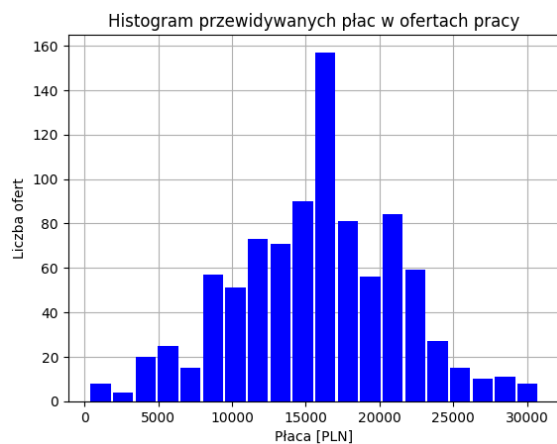
Tabela 5: Wpływ branży na zarobki programistów

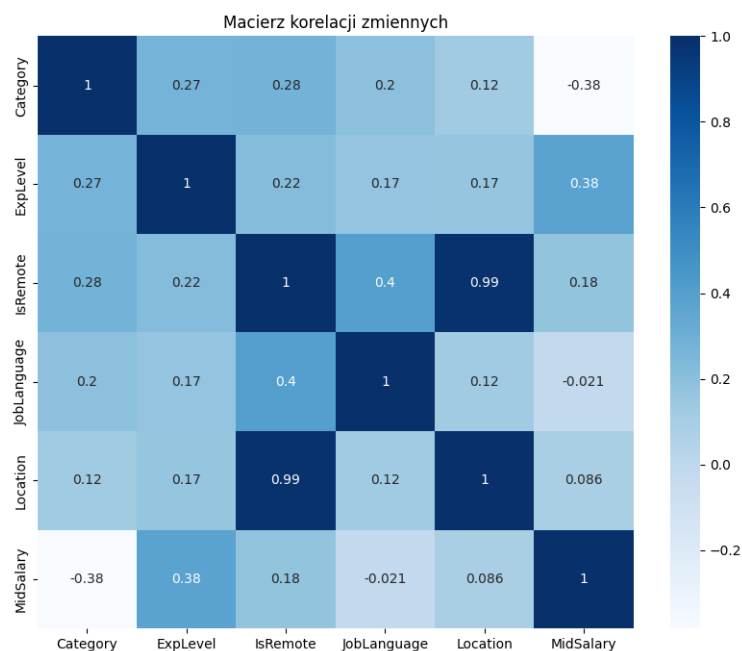
Technologia	Znaczenie
Pentesting	8680.04
CI/CD	8624.90
Android	7992.38
iOS	7205.09
React native	7102.48
ORACLE	6280.74
BeyondTrust	6213.83
Staking	5159.74
TPRM	5087.54
Selenium	4618.72
Flutter	4395.73
k8	4079.92
Oracle	3586.46
CASE	3366.19
PixiJS	2996.42
JasperReports	2504.53
Kubernetes	1832.98
Flink	1758.12
Informatica	1651.12
GCP	1639.86

Tabela 6: Wpływ technologii na zarobki programistów

5 Oferty bez widełek płacowych

Ciekawym zagadnieniem jest analiza ofert pracy, które nie zawierają informacji o zarobkach. Przy pomocy modelu opracowanego na podstawie danych z widełkami płacowymi, można spróbować przewidzieć zarobki programistów, których oferty nie zawierają tej informacji. **Czy oferty pracy bez podanego wynagrodzenia wyróżniają się czymś względem tych, które je posiadają?**





Oferty pracy bez widełek płacowych nie posiadają cech szczególnych w porównaniu do ofert z widełkami płacowymi. Nie można zauważyć żadnych zależności, dzięki którym moglibyśmy wskazać "w ciemno" czy oferta miała podane wynagrodzenie czy nie. Przyczyny tego zjawiska mogą być indywidualne dla każdej oferty.

6 Wnioski

Przy pomocy modelu regresji liniowej i macierzy korelacji udało się zidentyfikować czynniki, które mają największy wpływ na zarobki w branży IT. Z przeprowadzonej analizy wynika, że największy wpływ na zarobki mają kolejno:

- Poziom doświadczenia
- Firma
- Technologia
- Branża

Zauważona została wyraźna korelacja między poziomem doświadczenia a zarobkami, im wyższy poziom doświadczenia, tym na większe zarobki może liczyć specjalista IT.

Firma z której pochodzi oferta pracy również ma duży wpływ na zarobki programistów. Im bardziej znana i renomowana firma, tym wyższe zarobki mogą oferować. Należy jednak pamiętać, że firma sama w sobie nie gwarantuje wysokich zarobków, trzeba również wziąć pod uwagę inne czynniki, takie jak poziom doświadczenia czy branża.

W przypadku technologii, największe zarobki związane są z testami penetracyjnymi, CI/CD (tematyka de-vops), Androidem, iOS oraz React Native.

Jeśli chcemy zarabiać jak najwięcej powinniśmy zainteresować się tematem analizy biznesowej, bezpieczeństwa lub sztucznej inteligencji. Najniższe zarobki przeciętnie otrzymują programiści związani z testowaniem, administracją systemów czy tworzeniem gier.

Warto zauważyć, że zarobki w branży informatycznej są zróżnicowane i są zależne od wielu czynników, również takich, które mogły nie zostać uwzględnione w analizie. Podjęta została również próba zbadania zjawiska ofert pracy bez widełek płacowych, nie udało się jednak zidentyfikować cech wspólnych wyróżniające te oferty względem ofert z widełkami płacowymi.