

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Monika Zimolzak
Nr albumu: 432583

Determinanty rocznych zarobków w branży IT

Praca zaliczeniowa
na ćwiczenia z Ekonometrii
prowadzone przez mgr Rafała Walaskę

Warszawa, styczeń 2023

Streszczenie

Przedmiotem pracy jest zbadanie determinantów rocznych zarobków w branży IT. W tym celu postawiona została jedna główna hipoteza badawcza oraz cztery poboczne. Weryfikacja ich założeń opiera się na stworzonym modelu ekonometrycznym regresji liniowej. Zmienne do modelu zostały wybrane na podstawie przeglądu literatury i własnych przypuszczeń. Model poddano testom statystycznym na weryfikację jego poprawności. Wyniki modelu pozwoliły potwierdzić założenia postawionych hipotez badawczych.

Słowa kluczowe: zarobki, branża IT, determinanty dochodów, model ekonometryczny, zmienne objaśniające

SPIS TREŚCI

WSTĘP	4
ROZDZIAŁ I. Przegląd literatury i hipotezy badawcze	6
1. HIPOTEZA GŁÓWNA	6
2. HIPOTEZY POBOCZNE.....	7
2.1. Druga hipoteza badawcza.....	7
2.2. Trzecia hipoteza badawcza.....	7
2.3. Czwarta hipoteza badawcza	8
2.4. Piąta hipoteza badawcza.....	8
ROZDZIAŁ II. Opis zbioru danych i wybranych zmiennych.....	10
1. OPIS ZBIORU DANYCH.....	10
2. ZMIENNA OBJAŚNIANA I ZMIENNE OBJAŚNIAJĄCE	10
ROZDZIAŁ III. Strategia empiryczna	14
1. WYBÓR FORMY FUNKCYJNEJ	14
2. WERYFIKACJA KORELACJI	15
3. METODA „GENERAL TO SPECIFIC” – OD OGÓŁU DO SZCZEGÓŁU.	16
ROZDZIAŁ IV. Testy diagnostyczne – weryfikacja założeń KMRL	17
1. LINIOWOŚĆ FORMY FUNKCYJNEJ	17
2. NORMALNOŚĆ SKŁADNIKA LOSOWEGO.....	18
3. HOMOSKEDASTYCZNOŚĆ SKŁADNIKA LOSOWEGO	18
ROZDZIAŁ V. Analiza problemów z danymi.....	21
1. WSPÓŁCZYNNIK DETERMINACJI.....	21
2. OBSERWACJE NIETYPOWE I BŁĘDNE	21
3. WSPÓŁLINIOWOŚĆ	23
ROZDZIAŁ VI. Wyniki modelu.....	24
1. WERYFIKACJA POSTAWIONYCH HIPOTEZ	25
ZAKOŃCZENIE.....	27
BIBLIOGRAFIA.....	28

WSTĘP

Przemysł informatyczny (IT) to szybko rozwijająca się i stale ewoluująca dziedzina, która oferuje szerokie spektrum możliwości zawodowych. Jednym z najważniejszych aspektów pracy w branży IT jest potencjał zarobkowy. Według raportów Unii Europejskiej, sektor IT jest jedną z najlepiej opłacanych w UE, z wynagrodzeniami znacznie wyższymi niż średnia dla wszystkich sektorów. Jak podaje raport Eurostatu¹, średnie zarobki godzinowe brutto w UE dla sektora IT wyniosły 26,1 € w 2020 roku, w porównaniu do średniej 22,4 € dla wszystkich sektorów. Oznacza to, że specjaliści IT w UE mogą liczyć na zarobki około 17% wyższe niż średni pracownik. Dodatkowo należy zauważyć, że poziom wynagrodzeń różni się pomiędzy krajami UE. Najwyższą stawkę godzinową brutto dla sektora IT w 2020 roku osiągnął Luksemburg, gdzie specjaliści IT zarabiali średnio 40,7 € za godzinę. Kolejne miejsce zajmuje Irlandia, gdzie średnie brutto godzinowe zarobki dla sektora IT wyniosły 37,7 €. Inne kraje z wysokim potencjałem zarobkowym w sektorze IT to Holandia, Dania i Austria.

W tym miejscu warto podkreślić kolejną zaletę pracy w tej branży. Wiele firm IT oferuje możliwość pracy zdalnej, co daje pracownikom elastyczność pracy z dowolnego miejsca na świecie i balansowanie swojego życia osobistego i zawodowego. Dzięki takiemu rozwiązaniu wykwalifikowani specjaliści mogą szukać zatrudnienia w innych Państwach, oferujących wyższe wynagrodzenia.

Jednak oprócz możliwości zarobkowych, branża IT oferuje również szeroki wachlarz możliwości kariery i bezpieczeństwa zatrudnienia. Jest to ciągle zmieniający się sektor, umożliwiający nieustanny rozwój i zdobywanie nowych umiejętności. Dodatkowo oferuje on szerokie spektrum ról, od tworzenia oprogramowania i analizy danych po administrację sieci i bezpieczeństwo cybernetyczne. Dzięki temu, każdy chodź trochę zainteresowany tematem, jest w stanie znaleźć tu swoją rolę.

Mimo tak licznych zalet, branża IT jest obecnie jednym z najbardziej pożądanym sektorów na rynku pracy. Szybkie tempo zmian technologicznych i rosnąca zależność od technologii we wszystkich obszarach biznesu stworzyły duże zapotrzebowanie na wykwalifikowanych specjalistów IT. Jednym z kluczowych czynników napędzających popyt w branży IT jest postępująca transformacja cyfrowa przedsiębiorstw. Firmy coraz częściej zwracają się ku technologii, aby poprawić wydajność, obniżyć koszty i zyskać przewagę konkurencyjną. Doprowadziło to do dużego zapotrzebowania na specjalistów IT z

¹ Eurostat, *Structure of earnings survey 2020*,
https://ec.europa.eu/eurostat/databrowser/explore/all/popul?lang=en&subtheme=labour.earn.earn_ses2018&display=list&sort=category&extractionId=EARN_SES18_37, data dostępu 17.01.2023

umiejętnościami w obszarach takich jak tworzenie oprogramowania, analiza danych i cyberbezpieczeństwo.

Innym czynnikiem przyczyniającym się do wysokiego zapotrzebowania na specjalistów IT jest niedobór wykwalifikowanych pracowników w tej dziedzinie. Zdaniem Unii Europejskiej na rynku pracy brakuje informatyków, który w najbliższych latach ma rosnąć. Wynika to częściowo z faktu, że wielu specjalistów IT zbliża się do wieku emerytalnego i nie ma wystarczającej liczby młodych ludzi, którzy mogliby ich zastąpić.

Zapotrzebowanie na specjalistów IT jest również wysokie w określonych podsektorach. Na przykład istnieje duże zapotrzebowanie na specjalistów z umiejętnościami w zakresie sztucznej inteligencji, uczenia maszynowego i przetwarzania w chmurze. Trend pracy zdalnej w czasie pandemii doprowadził również do wzrostu zapotrzebowania na specjalistów IT w administrowaniu sieciami i systemami.

Wysokie zapotrzebowanie na specjalistów IT znajduje odzwierciedlenie w osiągniętych przez nich wynagrodzeniach. Wiele stanowisk IT oferuje konkurencyjne wynagrodzenia i pakiety świadczeń, a specjaliści IT o specjalistycznych umiejętnościach i doświadczeniu mogą uzyskać jeszcze wyższe pensje.

Oczekuje się, że branża IT będzie się nadal rozwijać w przyszłości, tworząc jeszcze więcej możliwości dla specjalistów. Wraz z postępem technologicznym zapotrzebowanie na wyspecjalizowanych pracowników, z dużą wiedzą z szeroko określonego obszaru IT będzie tylko rosło.

Z uwagi na opisane wyżej, liczne zalety pracy w tym sektorze, warto zweryfikować jakie składowe wpływają na poziom wynagrodzeń specjalistów przemysłu informatycznego. Przedmiotem niniejszej pracy jest zbadanie głównych determinantów wysokości rocznych zarobków w branży IT. W tym celu stworzony został model ekonometryczny regresji liniowej, który pozwoli zweryfikować postawione niżej hipotezy badawcze.

ROZDZIAŁ I. Przegląd literatury i hipotezy badawcze

1. Hipoteza główna

Determinanty zarobków w branży IT nie różnią się od ogólnych, przewidywalnych czynników, mających wpływ na wynagrodzenia. Do najczęściej wymienianych w literaturze zaliczane są: wykształcenie, doświadczenie, stanowisko pracy, wielkość firmy oraz płeć.

Jako czynnik tłumaczący wysokość zarobków, intuicyjnie na pierwszym miejscu można wskazać doświadczenie. Jego wpływ może być bardzo znaczący, ze względu na fakt, że wielu pracodawców jest gotowych płacić więcej, za pracowników z większym doświadczeniem w określonej dziedzinie. Doświadczenie może bowiem demonstrować poziom umiejętności, wiedzy i zdolności do wykonywania pracy skutecznie. Ponadto, z większym doświadczeniem, osoba może mieć okazję do przyjęcia większej odpowiedzialności, zarządzania bardziej skomplikowanymi projektami i pełnienia ról kierowniczych. Te odpowiedzialności zwykle wiążą się z wyższymi zarobkami.

Warto zauważyć, że zależność między doświadczeniem a zarobkami nie zawsze jest liniowa, ponieważ niektórzy doświadczeni pracownicy mogli osiągnąć pułap wynagrodzenia lub mieć ograniczone możliwości awansu. Jednak ogólnie rzecz biorąc, im większe doświadczenie ma dana osoba w określonej dziedzinie, tym większy może mieć potencjał zarobkowy.

Według artykułu naukowego pt.: „Human capital at work: The value of experience”¹² napisanego pod kierownictwem McKinsey Global Institute, kapitał ludzki jest znaczącym składnikiem całkowitego bogactwa jednostki i stanowi około 2/3 tej wartości. Umiejętności nabyte przez doświadczenie zawodowe stanowią średnio 46% z tej wartości w trakcie typowego okresu kariery zawodowej. Dodatkowo analiza wykazała, że doświadczenie zawodowe odgrywa kluczową rolę w wysokości zarobków. W krajach wysoce rozwiniętych jest to średnio około 40-43%, natomiast dla Państw zmagających się z problemem dostępu do edukacji, doświadczenie przekłada się nawet na 58% uzyskiwanych dochodów.

Przegląd literatury, obserwacja otoczenia oraz własne przypuszczenia pozwoliły na sformułowanie głównej hipotezy poniższego badania, która brzmi następująco: „*Wraz ze wzrostem lat doświadczenia wzrastają roczne zarobki pracownika*”.

² A. Madgavkar, B. Schaninger, S. Smit, J. Woetzel, H. Samandari, D. Carlin, J. Seong, K. Chockalingam, *Human capital at work: The value of experience*, McKinsey Global Institute, 02.06.2022

2. Hipotezy poboczne

2.1. Druga hipoteza badawcza

Istnieje wiele artykułów naukowych wykazujących różnicę w zarobkach mężczyzn i kobiet. Zjawisko to znane jest pod pojęciem „gender wage gap”. Różnica ta może być spowodowana przez różne czynniki, w tym dyskryminację, selekcję zawodową i różnice w edukacji, doświadczeniu i liczbie godzin pracy.

W pracy naukowej „Determinants Of Earnings In Estonia, Latvia And Lithuania”³, badającej główne determinanty zarobków w krajach nadbałtyckich, zróżnicowanie ze względu na płeć odgrywa jedną z kluczowych ról. Z przeprowadzonego badania wynika, że przeciętne wynagrodzenie kobiet na Łotwie i Litwie wynosiło około 80% średniej płacy mężczyzn. Dodatkowo wykazano brak powiązania płci z wybranymi sektorami pracy. Dla wskazanych dwóch państw zauważono mniejsze zróżnicowanie ze względu na płeć dla prac umysłowych, jednak dla Estonii głównie w zawodach fizycznych. Na podstawie wyników stwierdzono dużą zależność pomiędzy wysokością zarobków a płcią badanych.

W tym miejscu sformułowano drugą hipotezę badawczą: „Roczne zarobki w branży IT są wyższe dla płci męskiej”.

2.2. Trzecia hipoteza badawcza

Kolejnym czynnikiem mającym wpływ na wielkość zarobków jest poziom zajmowanego stanowiska. Stanowiska wyższego szczebla zwykle wiążą się z wyższym wynagrodzeniem i większą odpowiedzialnością. Proces przypisywania odpowiednich stanowisk został wytłumaczony i opisany w pracy naukowej: „Wartościowanie stanowisk pracy a ocena jakości pracy”⁴. Jak podają autorzy, „Metoda wartościowania stanowisk pracy jest narzędziem pomagającym podejmować racjonalne decyzje kadrowe i płacowe. Przeprowadzenie wartościowania dostarcza wiedzy na temat rzeczywistej wartości pracy wykonywanej na poszczególnych stanowiskach oraz jest punktem wyjścia do dokonania oceny jej jakości.” Jako jedna z zalet tego procesu wskazano ustosunkowanie szczebla danego stanowiska do otrzymywanego wynagrodzenia – „równa płaca za pracę o tej samej wartości”.

³ M. Hazans, R. Eamets, J. Earle, *Determinants Of Earnings In Estonia, Latvia And Lithuania*, OECD, 2003

⁴ A. Lis, M. Lis, *Wartościowanie stanowisk pracy a ocena jakości pracy*, Zeszyty Naukowe Wyższej Szkoły Zarządzania Ochroną Pracy W Katowicach, Nr 1(3)/2007, s. 113-118

Na tej podstawie, sformułowano kolejną hipotezę badawczą: „*Wyższe stanowisko wiąże się z wyższymi zarobkami.*”

2.3. Czwarta hipoteza badawcza

Determinując wysokość zarobków należy również rozważyć wielkość firmy. Duże firmy mogą mieć więcej zasobów i stabilności finansowej, aby oferować wyższe wynagrodzenia i świadczenia niż mniejsze firmy. Dodatkowo mogą mieć większe możliwości oferowania premii i pakietów świadczeń, takich jak ubezpieczenie zdrowotne, plany emerytalne i płatne urlopy. Mogą również stwarzać większe możliwości rozwoju zawodowego i awansu zawodowego, co z czasem może prowadzić do zwiększenia potencjału zarobkowego. Z drugiej strony małe firmy mogą mieć bardziej ograniczone zasoby i mogą nie być w stanie zaoferować takiego samego poziomu wynagrodzenia jak większe firmy. Będą w stanie jednak zapewnić pracownikom bardziej elastyczne środowisko pracy oraz możliwości zdobywania praktycznego doświadczenia i uczenia się.

Jak podaje raport PARP⁵: „Poziom wynagrodzenia brutto w przeliczeniu na jednego zatrudnionego jest silnie skorelowany wielkością przedsiębiorstwa – najniższy występuje w mikroprzedsiębiorstwach, najwyższy zaś w przedsiębiorstwach dużych”. Potwierdzenie tych słów można również znaleźć w książce autorstwa Ewy Beckiej-Krała „Wynagrodzenia pracowników w organizacji – teoria i praktyka”⁶. Autorka podkreśla, że porównując wartość wynagrodzeń pracowników, trzeba zwrócić uwagę na kilka aspektów, w tym wielkość firm. Jest oczywiste, że wynagrodzenia w dużych firmach różnią się od tych w małych firmach. Dodatkowo, zakres obowiązków na poszczególnych stanowiskach może być znacznie szerszy lub węższy, co utrudnia porównanie wynagrodzeń i ocenę ich równości.

Obserwacje te pozwoliły na sformułowanie czwartej hipotezy badawczej: „*Wielkość firmy dodatnio koreluje z wysokością zarobków*”.

2.4. Piąta hipoteza badawcza

Ostatnim czynnikiem, który warto rozważyć w kontekście wpływu na wynagrodzenia są dni wolne od pracy. Mogą one potencjalnie różnicować zarobki, w zależności od rodzaju zatrudnienia i warunków umowy o pracę. W przypadku standardowych umów o pracę w

⁵ PARP, *Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce*, Warszawa, 2022

⁶ Ewa Beck-Krała, *Wynagrodzenia pracowników w organizacji – teoria i praktyka*, Wydawnictwa AGH, Kraków, 2013

pełnym etacie, dni wolne zazwyczaj nie wpływają na wynagrodzenie, ponieważ jest ono zwykle stałe i nie zależy od liczby przepracowanych dni. Jednak w niektórych przypadkach pracownicy mogą być uprawnieni do płatnego urlopu, a jeśli wezmą więcej dni wolnych niż mają prawo, może to wpłynąć na ich wynagrodzenie. Natomiast na stanowiskach godzinowych płace są bezpośrednio powiązane z liczbą przepracowanych godzin, więc dni wolne od pracy prawdopodobnie spowodują niższe płace.

Według badania: „Badanie struktury i zmian rozkładu wynagrodzeń w Polsce w latach 2000-2006”⁷ rozkład wynagrodzeń jest ściśle uzależniony od liczby przepracowanych godzin. Dodatkowo warto zauważyć, że praca w sektorze IT jest bardzo elastyczna i może być wykonywana w pełni zdalnie. Według rozprawy doktorskiej: „Model organizacji i zarządzania systemem pracy zdalnej w branży IT”⁸, zwiększanie elastyczności pracy jest obecnie ważnym trendem na rynku pracy. Polega na dostosowywaniu się do zmian potrzeb pracowników, rynku i technologii. To udogodnienie dotyczy nietypowych form zatrudnienia, które są bardziej elastyczne pod względem czasu, miejsca i formy pracy. Rozwiązanie takie przekłada się, nie tylko na treści i zakres pracy, ale również na składniki i wysokość oferowanego wynagrodzenia. Można więc podejrzewać, że większa liczba dni wolnych od pracy wpływa na zmniejszenie wynagrodzenia, jednak pod względem atrakcyjności oferty, bardziej przyciąga potencjalnych kandydatów do pracy. Było to podstawą do sformułowania ostatniej hipotezy badawczej: „*Dni wolne od pracy ujemnie korelują z wysokością zarobków*”.

⁷ I. Marcinkowska, A. Ruzik, P. Strawiński, M. Walewski, *Badanie struktury i zmian rozkładu wynagrodzeń w Polsce w latach 2000-2006*, Departament Analiz Ekonomicznych i Prognoz Ministerstwo Pracy i Polityki Społecznej, Warszawa, 2008

⁸ Michał Trziszka, *Model organizacji i zarządzania systemem pracy zdalnej w branży IT*, Politechnika Poznańska 2019

ROZDZIAŁ II. Opis zbioru danych i wybranych zmiennych

1. Opis zbioru danych

W niniejszej pracy wykorzystano dane z anonimowych badań wynagrodzeń wśród europejskich specjalistów IT, z największym naciskiem na Niemcy. Badanie to jest wykonywane corocznie od 2015 roku, jednak na potrzeby pracy skorzystano z najbardziej aktualnych danych – dla 2020 roku. Autorami wskazanego badania są: Ksenia Legostay przy współpracy z Wiktora Szcherbana i Siergieja Wasiljewa.

Baza danych początkowo zawierała 1238 obserwacji, które zostały ograniczone do liczby 1068, tak by każda z obserwacji zawierała dane dotyczące niżej wymienionych zmiennych (usunięto odpowiedzi, które nie były pełne).

2. Zmienna objaśniana i zmienne objaśniające

Za zmienna objaśniana w modelu przyjęto roczne wynagrodzenie brutto respondenta wyrażone w euro. Jest to zmienna ciągła przyjmująca wartości w zakresie (10001;250000).

$$y = \text{Salary}$$

Zmienne objaśniające:

- Wiek respondenta – jest to zmienna ciągła przyjmująca wartości z zakresu (20;66)

$$x_1 = \text{Age}$$

Umieszczana jest w większości dostępnej literatury poświęconej determinizacji zarobków.

- Płeć respondenta - zmienna binarna, przyjmująca w bazie wartości „Female” i „Male”. W modelu została przekształcona na zmienną zero-jedynkową, przy czym „0” oznacza płeć żeńską, „1” płeć męską.

$$x_2 = \text{Gender}$$

Została wybrana do modelu ze względu na liczne badania dotyczące pojęcia „gender wage gap”.

- Doświadczenie - zmienna ilościowa ciągła, mierzona ilością przepracowanych lat. W modelu przyjmuje wartości z zakresu (1;40)

x₃ = Experience

Umieszczana jest w większości dostępnej literatury poświęconej determinacji zarobków.

- Dni wolne od pracy – jest to zmienna ciągła przyjmująca wartości z zakresu (0;60)

x₄ = Vacation_days

Umieszczona w modelu na podstawie przeglądu literatury.

- Poziom zajmowanego stanowiska - jest to zmienna dyskretna, która w poniższym modelu przyjmuje pięć poziomów:

„Junior” – najniższe stanowisko - osoba, która rozpoczyna swoją przygodę w IT

„Middle” – osoba, która samodzielnie wykonuje powierzone projekty

„Senior” – osoba, mająca rozległą wiedzę z zakresu programowania i odpowiednio duże doświadczenie w pracy programisty (minimum ok. 5 lat)

„Lead” - inżynier oprogramowania, który jest odpowiedzialny za kierowanie zespołem developerów

„Head” – najwyższe z możliwych stanowisk – dyrektor działu IT

Stanowisko „Head” jest poziomem bazowym.

x₅ = Seniority_level

Wybrana do modelowania na podstawie licznych badań nad determinacją zarobków.

- Status zatrudnienia – zmienna binarna, przyjmująca w bazie wartości: „Full-time employee” oraz „Self-employed/Part-time employee”. W modelu została przekształcona na zmienną zero-jedynkową, przy czym „1” oznacza pracownika zatrudnionego w pełnym wymiarze godzin, „0” pozostałe formy zatrudnienia.

x₆ = Employment_status

Umieszczona w modelu na podstawie własnych przypuszczeń o istotności tej zmiennej.

- Okres zatrudnienia – zmienna binarna, przyjmująca w bazie wartości: „Unlimited contract” oraz „Temporary contract”. W modelu została przekształcona na zmienną zero-jedynkową, przy czym „1” oznacza pracownika zatrudnionego na czas nieokreślony, „0” na umowę tymczasową.

x7 = Contract_duration

Umieszczona w modelu na podstawie własnych przypuszczeń o istotności tej zmiennej.

- Wielkość firmy - jest to zmienna dyskretna, która w poniższym modelu przyjmuje pięć poziomów:

„micro” – jednoosobowe działalności lub mikro-przedsiębiorstwa zatrudniające do 10 pracowników

„small” - małe przedsiębiorstwa, zatrudniające od 11 do 50 osób

“medium_size” - średniej wielkości firmy, zatrudniające od 51 do 100 osób

“large” - duże przedsiębiorstwa zatrudniające od 101 do 1000 pracowników

“extra-large” - największe firmy, korporacje, z liczbą ponad 1000 pracowników

Ostatni z poziomów – „extra-large” jest poziomem bazowym.

x8 = Company_size

Umieszczona w modelu na podstawie przeglądu literatury.

- Główny język w pracy - jest to zmienna dyskretna, która w poniższym modelu przyjmuje cztery poziomy:

„English” – język angielski

„German” – język niemiecki

„Italian” – język włoski

„Russian” – język rosyjski

x9 = Language

Umieszczona w modelu na podstawie własnych przypuszczeń o istotności tej zmiennej.

W wyżej opisanych zmiennych objaśniających, brakuje zmiennej mówiącej o poziomie wykształcenia danego pracownika, która jest brana pod uwagę w większości dostępnej literatury poświęconej determinizacji zarobków. Informacja ta nie została zawarta w opisanym badaniu ankietowym, jednak w tym miejscu warto rozważyć, czy wykształcenie faktycznie determinuje zarobki w branży IT. Jak podaje raport Randstad⁹ - globalnego lidera branży HR – sektor IT to jednak nie tylko programiści z dyplomem ukończenia kierunku technicznego na renomowanej uczelni. Obecnie około 10% osób ubiegających się o pracę w tej branży w ogóle nie ma wykształcenia wyższego, 30% to absolwenci kursów stacjonarnych i internetowych, a 46% wśród osób, które wyższe wykształcenie posiadają, to absolwenci kierunków nietechnicznych. Na podstawie takich obserwacji można podejrzewać brak istotnego wpływu wykształcenia na zarobki w sektorze IT.

⁹ Randstad, Employer brand research 2019, https://info.randstad.pl/raport-randstad-employer-brand-research-2019?&referral_domain=www.google.comNaNhttps://www.google.com/, data dostępu: 18.01.2023

ROZDZIAŁ III. Strategia empiryczna

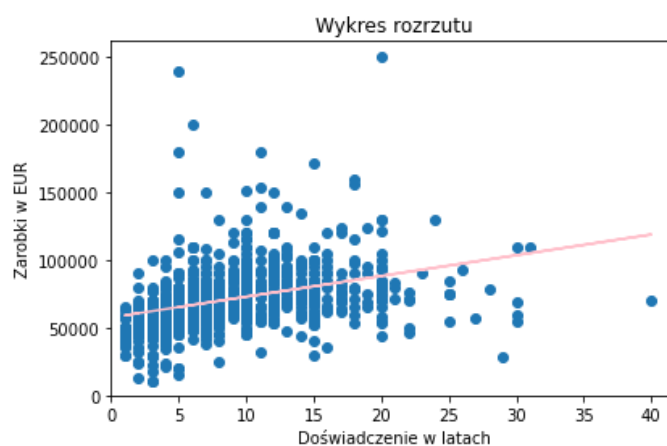
Przeprowadzone badanie ma na celu wskazanie determinantów wysokości rocznych zarobków w branży IT. W celu weryfikacji postawionych hipotez stworzono model regresji liniowej. Proces poniższego badania składał się z kilku etapów, które miały na celu opracowanie formy funkcyjnej modelu, która byłaby najlepiej dostosowana do teoretycznego podłoża dotyczącego tworzenia modeli ekonometrycznych. Przygotowano formę podstawową oraz dokonano kolejnych przekształceń, aby poprawić pierwotną wersję i uzyskać ostateczny model, na którym zweryfikowano hipotezy badawcze.

1. Wybór formy funkcyjnej

W celu określenia podstawowej formy funkcyjnej, zweryfikowano rozkład zmiennej objaśnianej - Salary oraz stworzono wykres rozrzutu tej zmiennej w stosunku do zmiennej objaśniającej – Experience, dla której intuicyjnie spodziewano największego powiązania ze zmienną objaśnianą.



Źródło: Opracowanie własne na podstawie zbioru danych



Źródło: Opracowanie własne na podstawie zbioru danych

Histogram rozkładu zmiennej Salary jest zbliżony do rozkładu normalnego. Wykres rozrzutu może natomiast wskazywać na statystyczną zależność pomiędzy wybranymi zmiennymi. Na podstawie tych wniosków, zarówno zmienną Salary, jak i Experience pozostawiono pierwotnie w formie niezlogarytmizowanej.

2. Weryfikacja korelacji

W kolejnym etapie zweryfikowano korelacje pomiędzy zmiennymi ciągłymi w modelu.



Źródło: Opracowanie własne na podstawie zbioru danych

Na podstawie powyższego diagramu stwierdzono silną dodatnią korelację pomiędzy zmiennymi Age oraz Experience. W związku z czym, w kolejnych etapach modelowania, warto rozważyć wprowadzenie interakcji pomiędzy tymi zmiennymi, co pozwoliłoby na lepsze określenie jak każda z tych zmiennych indywidualnie oddziałuje na zmienną objaśnianą. Pozostałe zmienne ciągłe nie wykazują wzajemnych zależności.

Dla modelu bazowego (m1) przyjęto następującą formę funkcyjną:

$$\begin{aligned} \text{Salary}_i = & \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Gender}_i + \beta_3 \text{Experience}_i + \beta_4 \text{Vacation_days}_i + \beta_5 \text{Contract_duration}_i \\ & + \beta_6 \text{Seniority_level_Junior}_i + \beta_7 \text{Seniority_level_Lead}_i + \beta_8 \text{Seniority_level_Middle}_i + \\ & \beta_9 \text{Seniority_level_Senior}_i + \beta_{10} \text{Employment_status}_i + \beta_{11} \text{Company_size_large}_i + \\ & \beta_{12} \text{Company_size_medium_size}_i + \beta_{13} \text{Company_size_micro}_i + \beta_{14} \text{Company_size_small}_i + \\ & \beta_{15} \text{Language_German}_i + \beta_{16} \text{Language_Italian}_i + \beta_{17} \text{Language_Russian}_i + \epsilon_i \end{aligned}$$

3. Metoda „General to specific” – od ogółu do szczegółu.

Po wprowadzeniu wszystkich wybranych zmiennych do modelu zastosowano metodę „od ogółu do szczegółu”, polegającą na stopniowym usuwaniu zmiennych nie mających istotnego wpływu na zmienną objaśnianą.

Proces ten rozpoczęto od zweryfikowania hipotezy 1. o braku istotności zmiennej „Age”. Na przyjętym poziomie istotności 5% stwierdzono brak podstaw do odrzucenia hipotezy zerowej ($p\text{-value} = 0,39$). Zmienna „Age” została usunięta z modelu.

Forma funkcyjna modelu (m2) przyjęła następującą postać:

$$\text{Salary}_i = \beta_0 + \beta_1 \text{Gender}_i + \beta_2 \text{Experience}_i + \beta_3 \text{Vacation_days}_i + \beta_4 \text{Contract_duration}_i + \beta_5 \text{Seniority_level_Junior}_i + \beta_6 \text{Seniority_level_Lead}_i + \beta_7 \text{Seniority_level_Middle}_i + \beta_8 \text{Seniority_level_Senior}_i + \beta_9 \text{Employment_status}_i + \beta_{10} \text{Company_size_large}_i + \beta_{11} \text{Company_size_medium_size}_i + \beta_{12} \text{Company_size_micro}_i + \beta_{13} \text{Company_size_small}_i + \beta_{14} \text{Language_German}_i + \beta_{15} \text{Language_Italian}_i + \beta_{16} \text{Language_Russian}_i + \varepsilon_i$$

Kolejna zmienna podejrzewana o brak istotności w modelu to „Contract_duration”. W celu weryfikacji tego podejrzenia sprawdzono hipotezę 2. o braku łącznej istotności zmiennej „Age” oraz „Contract_duration” w modelu 1. W tym miejscu również ustalono brak podstaw do odrzucenia hipotezy zerowej ($p\text{-value} = 0,42$). Zmienna „Contract_duration” została usunięta z modelu.

Po wykonaniu tej czynności otrzymano następującą formę funkcyjną (m3):

$$\text{Salary}_i = \beta_0 + \beta_1 \text{Gender}_i + \beta_2 \text{Experience}_i + \beta_3 \text{Vacation_days}_i + \beta_4 \text{Seniority_level_Junior}_i + \beta_5 \text{Seniority_level_Lead}_i + \beta_6 \text{Seniority_level_Middle}_i + \beta_7 \text{Seniority_level_Senior}_i + \beta_8 \text{Employment_status}_i + \beta_9 \text{Company_size_large}_i + \beta_{10} \text{Company_size_medium_size}_i + \beta_{11} \text{Company_size_micro}_i + \beta_{12} \text{Company_size_small}_i + \beta_{13} \text{Language_German}_i + \beta_{14} \text{Language_Italian}_i + \beta_{15} \text{Language_Russian}_i + \varepsilon_i$$

Dla tej formy funkcyjnej, na poziomie istotności 5%, na podstawie wyników modelu, zakończono przeprowadzanie metody „General to specific”, uzyskując przy tym finalny model 3. zawierający tylko te zmienne, które spełniają kryteria istotności statystycznej.

ROZDZIAŁ IV. Testy diagnostyczne – weryfikacja założeń KMRL

W celu uzyskania wiarygodnych i miarodajnych wyników, model musi spełniać założenia klasycznej metody regresji liniowej (KMRL). W przeciwnym wypadku wyniki modelu mogą być błędne lub niepewne, co skutkowałoby błędną interpretacją danych i nieprawidłowymi wnioskami. Weryfikacja tych założeń nastąpiła poprzez przeprowadzenie kilku różnych testów diagnostycznych.

1. Liniowość formy funkcyjnej

Poprawna forma funkcyjna modelu jest jednym z najważniejszych założeń przy stosowaniu modelu regresji liniowej. W celu zbadania liniowości w modelu przeprowadzono test RESET. Na jego postawie uzyskano p-value na poziomie bliskim zeru, a co za tym idzie, na poziomie istotności 5% odrzucono hipotezę o liniowości formy funkcyjnej.

Celem poprawy formy funkcyjnej modelu, zmienna objaśniania „Salary”, oraz zmienna objaśniająca „Experience” zostały zlogarytmizowane. Ten zabieg jest powszechnie stosowany w pracy badawczych, które determinują zarobki.

W rezultacie formę funkcyjną (m4) przyjęła postać:

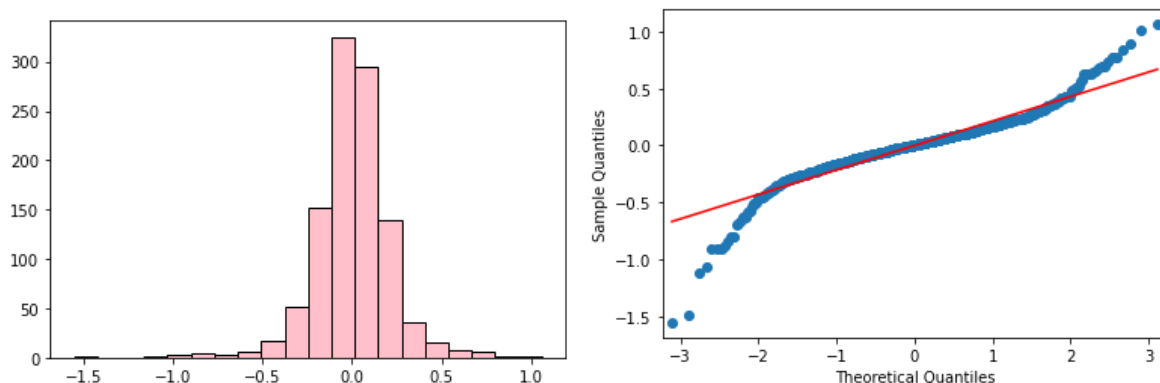
$$\ln(\text{Salary}_i) = \beta_0 + \beta_1 \text{Gender}_i + \beta_2 \ln(\text{Experience}_i) + \beta_3 \text{Vacation_days}_i + \beta_4 \text{Seniority_level_Junior}_i + \beta_5 \text{Seniority_level_Lead}_i + \beta_6 \text{Seniority_level_Middle}_i + \beta_7 \text{Seniority_level_Senior}_i + \beta_8 \text{Employment_status}_i + \beta_9 \text{Company_size_large}_i + \beta_{10} \text{Company_size_medium_size}_i + \beta_{11} \text{Company_size_micro}_i + \beta_{12} \text{Company_size_small}_i + \beta_{13} \text{Language_German}_i + \beta_{14} \text{Language_Italian}_i + \beta_{15} \text{Language_Russian}_i + \varepsilon_i$$

Następnie ponownie zweryfikowano założenie o liniowości formy funkcyjnej testem RESET. Uzyskane p-value równe 0,28, na przyjętym poziomie istotności 5%, wskazało na brak podstaw do odrzucenia hipotezy o liniowości w modelu. Jednocześnie model 4. został uznany za najlepszy na tym etapie.

Kolejnym krokiem było sprawdzenie stabilności parametrów na podstawie testu CHOWA. Procedura ta ma na celu weryfikację czy parametry w modelu będą takie same dla kilku próbek. Test przeprowadzono dla zmiennej binarnej „Gender”, uzyskując przy tym statystykę testową równą 1,12 przy jednoczesnym przedziale krytycznym (1,65 ; +∞). Otrzymana wartość statystyki nie wpada do obszaru krytycznego, w związku z czym brak jest podstaw do odrzucenia hipotezy zerowej. Parametry w próbkach danego modelu są stabilne.

2. Normalność składnika losowego

Kolejnym założeniem KMRL jest normalność składnika losowego. Celem weryfikacji przeprowadzono analizę graficzną oraz wykonano formalny test na normalność składnika losowego - Jargue-Bera.



Źródło: Opracowanie własne na podstawie zbioru danych

Histogram reszt wskazuje wartości na poziomie -1,5, które mogą zaburzać normalność rozkładu składnika losowego. Wykres kwantylowy natomiast obrazuje brak dopasowania reszt do rozkładu normalnego na obu końcach.

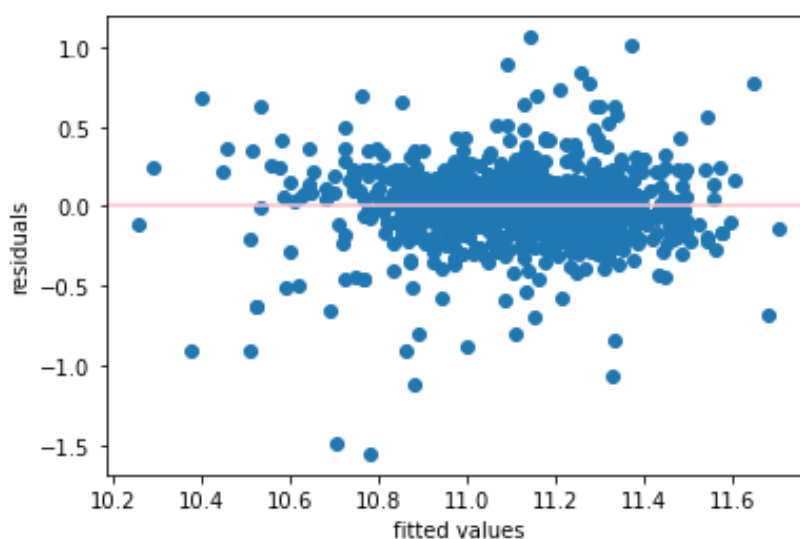
Dodatkowo w teście Jargue-Bera skośność przyjęła wartość -0,69 (przy docelowej wartości 0), kurtoza wyniosła 10,27 (docelowa wartość równa 3), a p-value równe 0, na podstawie którego, odrzucona została hipoteza o normalności rozkładu reszt w modelu.

Niemniej jednak, ilość obserwacji w modelu (1068 – duża próba badawcza) sprawia, że problem braku normalności składnika losowego nie jest istotny, ze względu na fakt, że rozkłady statystyk są bliskie standardowym rozkładom, nawet jeśli założenie to nie jest spełnione.

3. Homoskedastyczność składnika losowego

Homoskedastyczność składnika losowego, to założenie o stałej wariancji reszt w modelu, dla każdego poziomu zmiennej objaśnianej. Odpowiednie założenie o wariancji reszt jest konieczne do prawidłowego obliczenia estymatorów i przedziałów ufności dla parametrów modelu. Jeśli wariancja reszt jest zmienna, oznacza to, że model jest heteroskedastyczny i nie spełnia założenia o homoskedastyczności. To może prowadzić do nieprawidłowych wyników statystycznych i błędnych wniosków.

Założenie o heteroskedastyczności w modelu sprawdzono dwutorowo – w pierwszej kolejności graficznie, na podstawie wykresu rozrzutu reszt:



Źródło: Opracowanie własne na podstawie zbioru danych

Zgodnie z założeniem, wartość oczekiwana reszt powinna być równa 0, w związku z czym, oczekuje się równomiernie rozłożonych punktów wokół wartości 0. Z analizy wykresu można zaobserwować punkty wskazujące na problem z homoskedastycznością – są to np. obserwacje na poziomie residuals równym -1,5.

Formalna weryfikacja założenia o homoskedastyczności opiera się na wykonaniu testu Breuscha – Pagana oraz testu White’a. Pierwszy z wymienionych polega na porównaniu wariancji reszt modelu z wariancją estymatora reszt z modelu złożonego z jedną zmienną objaśniającą. Otrzymane p-value na poziomie bardzo bliskim 0, potwierdziło przypuszczenie o braku homoskedastyczności w zaproponowanym modelu. Hipoteza zerowa o homoskedastyczności została odrzucona na rzecz hipotezy alternatywnej o heteroskedastyczności składnika losowego.

Test White’a, polegający na porównywaniu wariancji reszt modelu, z wariancją estymatora reszt z modelu złożonego z jedną zmienną objaśniającą i komponentami składowymi wariancji, również potwierdził heteroskedastyczność w modelu (p-value ponownie bardzo bliskie 0).

W celu rozwiązania problemu heteroskedastyczności i by móc poprawnie zinterpretować uzyskane wyniki modelu, zastosowano macierz odporną White’a do estymacji macierzy wariancji. Dzięki tej metodzie, nawet jeśli błędy losowe w modelu nie są

homoskedastyczne, rozkłady statystyk uzyskanych są asymptotyczne, co pozwala na poprawne wnioskowanie. Na podstawie tego stworzony został ostatni model (m5), który pozwoli zweryfikować postawione hipotezy badawcze. Wyniki modelu przedstawiono w tabeli poniżej:

Dep. Variable:	ISalary	R-squared:	0.459
Model:	OLS	Adj. R-squared:	0.452
Method:	Least Squares	F-statistic:	47.13
Date:	Wed, 18 Jan 2023	Prob (F-statistic):	2.78e-106
Time:	23:00:45	Log-Likelihood:	78.951
No. Observations:	1068	AIC:	-125.9
Df Residuals:	1052	BIC:	-46.33
Df Model:	15		
Covariance Type:	HC0		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	11.1849	0.230	48.614	0.000	10.734	11.636
Gender	0.0864	0.019	4.563	0.000	0.049	0.123
IExperience	0.0683	0.015	4.580	0.000	0.039	0.098
Vacation_days	0.0102	0.005	2.118	0.034	0.001	0.020
Seniority_level_Junior	-0.5721	0.063	-9.056	0.000	-0.696	-0.448
Seniority_level_Lead	-0.1085	0.044	-2.454	0.014	-0.195	-0.022
Seniority_level_Middle	-0.4245	0.043	-9.916	0.000	-0.508	-0.341
Seniority_level_Senior	-0.2563	0.041	-6.284	0.000	-0.336	-0.176
Employment_status	-0.1770	0.137	-1.289	0.197	-0.446	0.092
Company_size_large	-0.0788	0.015	-5.362	0.000	-0.108	-0.050
Company_size_medium_size	-0.1251	0.022	-5.719	0.000	-0.168	-0.082
Company_size_micro	-0.2173	0.056	-3.862	0.000	-0.328	-0.107
Company_size_small	-0.1362	0.025	-5.401	0.000	-0.186	-0.087
Language_German	-0.0577	0.020	-2.852	0.004	-0.097	-0.018
Language_Italian	-0.5569	0.103	-5.413	0.000	-0.759	-0.355
Language_Russian	-0.4983	0.205	-2.435	0.015	-0.899	-0.097

Omnibus:	233.643	Durbin-Watson:	2.150
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2433.221
Skew:	-0.687	Prob(JB):	0.00
Kurtosis:	10.266	Cond. No.	546.

Źródło: Opracowanie własne na podstawie zbioru danych

ROZDZIAŁ V. Analiza problemów z danymi

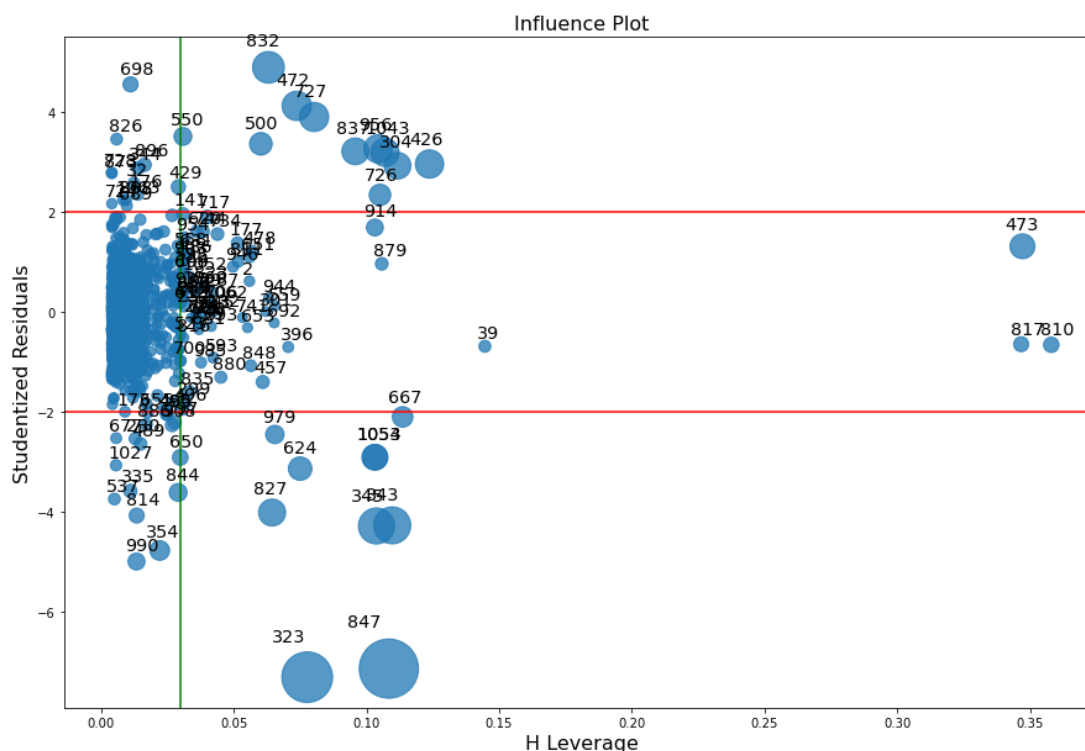
1. Współczynnik determinacji:

Współczynnik determinacji (R^2) to statystyka, która określa, jak dobrze model regresji dopasowuje się do danych, innymi słowy, jak duża część zmienności zmiennej zależnej jest wyjaśniana przez zmienne niezależne w modelu. Dla wersji ostatecznej modelu (m5) wartość R^2 równa jest 0,459. Model tłumaczy więc 45,9% zmienności danych, co jest uważane za dobre, ale nie doskonałe dopasowanie.

2. Obserwacje nietypowe i błędne:

Kolejnym krokiem badania jest przeanalizowanie danych w celu wykrycia ewentualnych nieprawidłowych obserwacji. Jest to na tyle ważne, że tego typu dane mogą mieć wpływ na wyniki modelu. Obserwacje nietypowe różnią się od pozostałych, ale mogą być zgodne z modelem regresji. W takim przypadku poprawiają one dopasowanie modelu i dokładność wyników. Natomiast obserwacje błędne nie pasują do modelu regresji i mają negatywny wpływ na otrzymane wyniki.

Aby zweryfikować czy w danych użytych do modelowania nie ma nietypowych lub błędnych obserwacji, stosowano trzy miary: wartość dźwigni, reszty studentyzowane oraz odległość Cooka. Stworzono wykres obrazujący wartość dźwigni i studentyzowanych reszt:



Źródło: Opracowanie własne na podstawie zbioru danych

Zielona linia pionowa wskazuje wartość graniczną obserwacji nietypowych pod względem x. Czerwone linie poziome wskazują przedział obserwacji nietypowych ze względu na duże y – czyli wysokie reszty. Łącząc te dwa warunki – dźwignię oraz standaryzowane reszty, powstaje możliwość określenia obserwacji nietypowych w modelu. Obszar tych obserwacji to miejsce na wykresie po prawej stronie pionowej linii oraz powyżej pierwszej poziomej i poniżej drugiej poziomej (prawy górny róg i prawy dolny róg). Dodatkowo wielkość pól niebieskich kół odnoszą się do trzeciej miary – odległości Cooka. Większa powierzchnia pola wskazuje na obserwację nietypową.

Na przedstawionym wykresie, obserwacje, które można określić mianem nietypowych, to 9 obserwacji w prawej dolnej części wykresu oraz 10 obserwacji w jego prawej górnej części. Trzy najbardziej zauważalne, znajdujące się we wskazanym obszarze oraz dodatkowo oznaczające się dużym polem powierzchni koła, przedstawiono w tabeli poniżej:

1	Gender	Experience	Seniority_level	Salary	Vacation_days	Employment_status	Language	Company_size
323	Male	9	Senior	60000	28	Full-time employee	English	micro
832	Male	14	Middle	65000	30	Full-time employee	German	large
847	Male	7	Senior	83000	26	Full-time employee	English	large

Obserwacja z numerem 323 to osoba na stanowisku „Senior”. Jej nietypowość odznacza się stosunkowo niskimi zarobkami w odniesieniu do osób zatrudnionych na tym samym stanowisku – średnia wysokość zarobków to 74452 EUR. Uzasadnienia tej różnicy można szukać np. w wieloletniej pracy w jednej firmie, zapewne przy stosunkowo dawnym ustaleniu stawki wynagrodzenia – jest to tzw. stagnacja zawodowa.

Obserwacja z numerem 832 to pracownik dużej firmy, który pomimo wieloletniego doświadczenia w branży IT, w dalszym ciągu obejmuje stanowisko „Middle”. Może być to spowodowane wieloma czynnikami, jak np. brak chęci rozwoju lub polityka firmy niesprzyjająca awansowaniu pracowników.

Ostatnia z opisywanych obserwacji nietypowych to numer 847, która z drugiej strony, mimo stosunkowo niewielkiego doświadczenia w odniesieniu do stanowiska „Senior” osiąga wysokie zarobki. W takim przypadku, wytłumaczenia można szukać w wysoko postawionych zadaniach, zaufaniu pracodawcy, umiejętności skutecznej negocjacji, czy np. otrzymania tego stanowiska na wzgląd powiązań na płaszczyźnie prywatnej.

Każdą, z wyżej wymienionych obserwacji nietypowych, można racjonalnie wyjaśnić powołując się na własne doświadczenia, intuicję czy obserwacje otoczenia. W związku z powyższym, uznaje się brak poważnych argumentów dających prawo do usunięcia tych obserwacji ze zbioru danych.

3. Współliniowość

Następny etap badania poświęcony został analizie współliniowości w modelu. Problem współliniowości może prowadzić do błędów w interpretacji wyników modelu oraz nieprawidłowych wniosków. W celu zbadania związku między zmiennymi objaśniającymi i zmienną objaśnianą, wykonano analizę współliniowości. Wykorzystano w tym celu statystykę VIF dla każdej zmiennej.

Statystyka VIF jest używana do określenia stopnia współliniowości między zmiennymi w modelu regresji. Działa ona poprzez oszacowanie wpływu każdej zmiennej na pozostałe zmienne w modelu. Wartości VIF są liczone dla każdej zmiennej i im większa wartość, tym silniejsze zjawisko współliniowości między zmiennymi. Przyjęta granica dopuszczalności współliniowości to wartość 10. Wyniki przedstawia poniższa tabela:

	VIF	Tolerance
Gender	1.103513	0.906196
IEExperience	1.785069	0.560202
Vacation_days	1.102794	0.906788
Seniority_level_Junior	3.157048	0.316752
Seniority_level_Lead	4.354393	0.229653
Seniority_level_Middle	7.439528	0.134417
Seniority_level_Senior	7.867598	0.127104
Employment_status	1.070032	0.934552
Company_size_large	1.297188	0.770898
Company_size_medium_size	1.179311	0.847953
Company_size_micro	1.158619	0.863096
Company_size_small	1.246767	0.802075
Language_German	1.079687	0.926194
Language_Italian	1.021361	0.979086
Language_Russian	1.013247	0.986926

Źródło: Opracowanie własne na podstawie zbioru danych

Zgodnie z przyjętą wartością graniczną, w modelu nie występują zmienne odpowiedzialne za generowanie problemu współliniowości.

ROZDZIAŁ VI. Wyniki modelu

Interpretacja wyników opiera się na ostatniej otrzymanej formie funkcyjnej – dla modelu 5. Poniższa tabela przedstawia wyniki dla pięciu modeli regresji liniowej:

	(1)	(2)	(3)	(4)	(5)
Age	-120.105 (140.021)				
Company_size_large	-6947.879*** (1233.768)	-7006.682*** (1231.707)	-7000.901*** (1231.695)	-0.079*** (0.017)	-0.079*** (0.015)
Company_size_medium_size	-9779.558*** (1728.715)	-9884.171*** (1724.191)	-9921.031*** (1723.799)	-0.125*** (0.023)	-0.125*** (0.022)
Company_size_micro	-13710.975*** (2812.454)	-13757.721*** (2811.572)	-14024.992*** (2798.868)	-0.217*** (0.038)	-0.217*** (0.056)
Company_size_small	-9727.631*** (1651.344)	-9756.263*** (1650.799)	-9807.435*** (1650.009)	-0.136*** (0.022)	-0.136*** (0.025)
Contract_duration	2638.705 (2649.449)	2651.762 (2649.072)			
Employment_status	-29576.980*** (3908.002)	-29650.595*** (3906.568)	-28483.563*** (3728.552)	-0.177*** (0.050)	-0.177*** (0.137)
Experience	502.486*** (158.474)	412.614*** (118.882)	416.333*** (118.824)		
Gender	6703.095*** (1515.309)	6760.308*** (1513.650)	6726.582*** (1513.277)	0.086*** (0.021)	0.086*** (0.019)
Intercept	109496.950*** (7244.397)	106631.243*** (6427.232)	108171.024*** (6240.454)	11.185*** (0.088)	11.185*** (0.230)
Language_German	-4249.372*** (1526.815)	-4483.185*** (1502.096)	-4472.503*** (1502.060)	-0.058*** (0.020)	-0.058*** (0.020)
Language_Italian	-31043.030*** (9814.657)	-31018.941*** (9813.383)	-31820.048*** (9780.706)	-0.557*** (0.132)	-0.557*** (0.103)
Language_Russian	-17959.348*** (5360.789)	-18220.799*** (5351.444)	-18110.048*** (5350.305)	-0.498*** (0.072)	-0.498*** (0.205)
Seniority_level_Junior	-42118.615*** (3735.432)	-42039.388*** (3733.821)	-42330.878*** (3722.452)	-0.572*** (0.053)	-0.572*** (0.063)
Seniority_level_Lead	-8601.550*** (3098.480)	-8502.673*** (3095.946)	-8637.742*** (3093.007)	-0.108*** (0.042)	-0.108*** (0.044)
Seniority_level_Middle	-33124.761*** (3051.928)	-33015.942*** (3048.907)	-33206.249*** (3042.976)	-0.425*** (0.041)	-0.425*** (0.043)
Seniority_level_Senior	-21816.518*** (2891.619)	-21741.039*** (2889.917)	-21911.991*** (2884.869)	-0.256*** (0.039)	-0.256*** (0.041)
Vacation_days	427.238*** (150.507)	418.640*** (150.154)	419.996*** (150.148)	0.010*** (0.002)	0.010*** (0.005)
lExperience				0.068*** (0.014)	0.068*** (0.015)
Observations	1,068	1,068	1,068	1,068	1,068
R ²	0.417	0.417	0.416	0.459	0.459
Adjusted R ²	0.408	0.408	0.408	0.452	0.452
Residual Std. Error	16735.187 (df=1050)	16733.083 (df=1051)	16733.099 (df=1052)	0.226 (df=1052)	0.226 (df=1052)
F Statistic	44.230*** (df=17; 1050)	46.960*** (df=16; 1051)	50.024*** (df=15; 1052)	59.555*** (df=15; 1052)	47.131*** (df=15; 1052)

Note:

*p<0.1; **p<0.05; ***p<0.01

Źródło: Opracowanie własne na podstawie zbioru danych

Z uwagi na fakt zlogarytmizowania zmiennej objaśnianej, wszystkie wyniki należy interpretować w kontekście semielastyczności cząstkowej. Otrzymane oszacowania zostały więc poddane przekształceniu : $(e^{\beta} - 1) * 100\%$

1. Weryfikacja postawionych hipotez:

Główna hipoteza badawcza brzmi następująco: *„Wraz ze wzrostem lat doświadczenia wzrastają roczne zarobki pracownika”*.

Przystępując do weryfikacji postawionej hipotezy, należy ustosunkować interpretację modelu do dwóch zmiennych zlogarytmizowanych (zmienna „Experience” również występuje w logarytmie”). W takim wypadku, interpretacja wzrostu współczynnika korelacji między zmienną objaśnianą a zmienną objaśniającą jest proporcjonalna wzrostowi logarytmów tych zmiennych, co oznacza, że każdy jednostkowy wzrost zmiennej objaśniającej przynosi proporcjonalnie większy wpływ na zmienną objaśnianą.

Wyniki modelu pokazują, że każdy jednoprocenowy wzrost doświadczenia pracownika niesie za sobą wzrost wynagrodzenia o 0,068%, co wskazuje na zdecydowanie niższą korelację niż przewidywano.

Hipoteza 2: *„Roczne zarobki w branży IT są wyższe dla płci męskiej”*.

Wyniki modelu jednoznacznie wskazują, że wynagrodzenie mężczyzn jest średnio o 8,98% wyższe od wynagrodzenia kobiet, co potwierdza postawioną hipotezę o pozytywnej korelacji wysokości zarobków z byciem mężczyzną.

Hipoteza 3: *„Wyższe stanowisko wiąże się z wyższymi zarobkami”*.

Weryfikując wyniki modelu dla 5 poziomów możliwych stanowisk otrzymano następujące wnioski:

- Pracownik obejmujący posadę inżyniera oprogramowania - „Lead” zarabia średnio o 10,24% mniej w porównaniu do dyrektora działu IT.
- Stanowisko „Senior” odpowiednio 22,59 % mniej.
- Osoba na stanowisku „Middle” - 34,62% mniej
- Natomiast pracownik na najniższym szczeblu – „Junior” otrzymuje wynagrodzenie średnio o 43,56% niższe niż dyrektor działu IT.

Wyżej przedstawione wnioski potwierdzają założenie 3 hipotezy.

Hipoteza 4: „*Wielkość firmy dodatnio koreluje z wysokością zarobków*”

Zmienna odpowiadająca za wielkość firmy również została podzielona na 5 poziomów.

Wynik regresji pozwolił na wyciągnięcie następujących wniosków:

- W przedsiębiorstwach dużych „large” (100-1000 pracowników) zarobki są średnio o 7,6% niższe w porównaniu do dużych korporacji (+1000 zatrudnionych osób)
- Firmy średniego rozmiaru „medium_size” odznaczają się zarobkami niższymi o 11,75%.
- Małe przedsiębiorstwa „small” odpowiednio o 12,72%
- Natomiast mikro przedsiębiorstwa oraz jednoosobowe działalności wykazują 19,5% mniejsze zarobki swoich pracowników w porównaniu do największych korporacji.

Wyniki te potwierdzają założenie hipotezy 4.

Hipoteza 5: „*Dni wolne od pracy ujemnie korelują z wysokością zarobków*”.

Korelacja pomiędzy dniami wolnymi od pracy a wysokością rocznych zarobków okazała się być mało znacząca. Wyniki modelu pokazują, że wzrost dni urlopu o jednostkę, wiąże się ze wzrostem dochodu o jedynie 1%. Jest to mała wartość w stosunku do przewidywanych wyników modelu.

Podsumowując, zbudowany model ekonometryczny pozwolił na jednoznaczną weryfikację 3 z 5 postawionych hipotez badawczych – hipotezy drugiej, trzeciej oraz czwartej. Wynik głównej hipotezy wykazał zgodną z założeniem, jednak niższą niż przewidywano korelację pomiędzy zmiennymi. Taki sam skutek przyniosła weryfikacja hipotezy piątej.

ZAKOŃCZENIE

Podstawowym celem pracy było wskazanie determinantów rocznych zarobków w branży IT, przy pomocy pięciu postawionych hipotez badawczych. Na 3 z nich udało się znaleźć dokładne potwierdzenie, natomiast pozostałe okazały się być równie prawidłowe, jednak zdecydowanie w mniejszej skali niż przewidywano. W celu ich weryfikacji postawionych hipotez został stworzony model ekonometryczny regresji liniowej. Myć móc prawidłowo odczytać jego wyniki, został on poddany kilku testom weryfikującym założenia KMRL i poddany estymacji za pomocą macierzy odpornej White'a.

Uzyskane wyniki pokazują, że największe zarobki w sektorze usług IT mają szansę osiągać mężczyźni, zatrudnieni na najwyższym możliwym stanowisku oraz w dużej korporacji zatrudniającej ponad 1000 pracowników. Dodatkowo osoby z dłuższym stażem pracy oraz mniejszą ilością dni wolnych również mogą spodziewać się uzyskania wyższego wynagrodzenia.

Należy podkreślić, że zbudowany model nie jest idealnym wskaźnikiem determinacji zarobków w branży IT. Literatura opisująca wskazany temat podkreśla również istotność innych zmiennych, jak np. poziomu wykształcenia czy sektor pracy. Zmienne te nie występowały jednak w wybranym do modelu zbiorze danych.

BIBLIOGRAFIA

A. Madgavkar, B. Schaninger, S. Smit, J. Woetzel, H. Samandari, D. Carlin, J. Seong, K. Chockalingam, *Human capital at work: The value of experience*, McKinsey Global Institute, 02.06.2022

M. Hazans, R. Eamets, J. Earle, *Determinants Of Earnings In Estonia, Latvia And Lithuania*, OECD, 2003

A. Lis, M. Lis, *Wartościowanie stanowisk pracy a ocena jakości pracy*, Zeszyty Naukowe Wyższej Szkoły Zarządzania Ochroną Pracy W Katowicach, Nr 1(3)/2007, s. 113-118

Ewa Beck-Krala, *Wynagrodzenia pracowników w organizacji – teoria i praktyka*, Wydawnictwa AGH, Kraków, 2013

I. Marcinkowska, A. Ruzik, P. Strawiński, M. Walewski, *Badanie struktury i zmian rozkładu wynagrodzeń w Polsce w latach 2000-2006*, Departament Analiz Ekonomicznych i Prognoz Ministerstwo Pracy i Polityki Społecznej, Warszawa, 2008

Michał Trziszka, *Model organizacji i zarządzania systemem pracy zdalnej w branży IT*, Politechnika Poznańska 2019

PARP, *Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce*, Warszawa, 2022

Eurostat, *Structure of earnings survey 2020*, https://ec.europa.eu/eurostat/databrowser/explore/all/popul?lang=en&subtheme=labour.earn.earn_ses2018&display=list&sort=category&extractionId=EARN_SES18_37, data dostępu: 17.01.2023

Randstad, Employer brand research 2019, https://info.randstad.pl/raport-randstad-employer-brand-research-2019?&referral_domain=www.google.comNaNhttps://www.google.com/, data dostępu: 18.01.2023