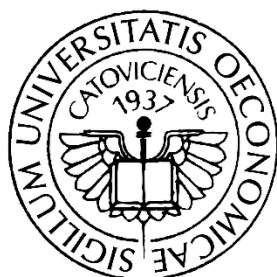


# JAROSŁAW NOCUŃ

Przetwarzanie, analiza i wizualizacja  
danych w systemie SAS

Grupa ćwiczeniowa IE\_cw\_SI



Uniwersytet  
Ekonomiczny  
w Katowicach

## Spis treści

<b>Wprowadzenie .....</b>	<b>2</b>
<b>Cel główny .....</b>	<b>2</b>
<b>Cele szczegółowe .....</b>	<b>2</b>
<b>Dane źródłowe.....</b>	<b>3</b>
<b>Przygotowanie do pracy.....</b>	<b>4</b>
<b>Analiza procesów .....</b>	<b>6</b>
<b>1. Wpływ doświadczenia na wynagrodzenie. Określenie, w jaki sposób poziom doświadczenia wpływa na wysokość wynagrodzeń w branży Data Science.....</b>	<b>6</b>
<b>2. Analiza wynagrodzeń w zależności od zajmowanego stanowiska .....</b>	<b>8</b>
<b>3. Analiza rozkładu wynagrodzeń w branży Data Science w zależności od rozmiaru firmy</b>	<b>11</b>
<b>4. Analiza wynagrodzeń w zależności od rodzaju zatrudnienia .....</b>	<b>15</b>
<b>5. Analiza średniego rocznego wynagrodzenia według poziomu doświadczenia oraz poziomu wiedzy specjalistycznej w latach 2020-2023. ....</b>	<b>17</b>
<b>Podsumowanie .....</b>	<b>21</b>

## Wprowadzenie

Ta praca ma na celu analizę zarobków w branży Data Science w latach 2020-2023 biorąc pod uwagę oferty stanowisk pracy w Stanach Zjednoczonych. Cytując Bluemetrica „Data Science to branża, która zajmuje się przetwarzaniem i analizą danych z wykorzystaniem metod statystycznych, języków programowania oraz zaawansowanych narzędzi komputerowych. Jest to względnie nowa branża, która ma przed sobą świetlaną przyszłość. Do niedawna analizy danych przeprowadzali statystycy.” 29 czerwiec 2022

## Cel główny

Głównym celem jest przeprowadzenie wszechstronnej analizy czynników wpływających na strukturę wynagrodzeń w branży Data Science.

## Cele szczegółowe

1. Wpływ doświadczenia na wynagrodzenie. Określenie, w jaki sposób poziom doświadczenia wpływa na wysokość wynagrodzeń analityków danych w Stanach Zjednoczonych.
2. Analiza wynagrodzeń w zależności od zajmowanego stanowiska.
3. Rozkład wynagrodzeń analityków danych w zależności od rozmiaru firmy.
4. Analiza wynagrodzeń w zależności od rodzaju zatrudnienia
5. Analiza średniego rocznego wynagrodzenia według poziomu doświadczenia oraz poziomu wiedzy specjalistycznej w latach 2020-2023.

Przeanalizowanie tych celów szczegółowych pozwoli na uzyskanie wniosków dotyczących zarobków w Data Science oraz lepsze zrozumienie tego jak kształtuje się rynek w tej branży.

## Dane źródłowe

Dane, które zostały wykorzystane do tej pracy pochodzą ze strony kaggle.com. Na tej stronie dane są udostępniane bezpłatnie jako pliki w formacie .csv. W tym formacie dane zostały wykorzystane do przeprowadzenia analizy.

Dane obejmują okres 2020-2023. Są one podzielone na 11 kolumn w których przedstawione są kolejno:

1. Stanowisko
2. Rodzaj zatrudnienia
3. Poziom doświadczenia
4. Poziom wiedzy specjalistycznej
5. Wynagrodzenie (roczne)
6. Waluta wynagrodzenia
7. Miejsce Firmy
8. Wynagrodzenie w USD
9. Miejsce zamieszkania pracownika
10. Rozmiar firmy
11. Rok

Informacje te zapewniają kompleksowy wgląd w szczegóły związane z wyżej wymienionymi wytycznymi oferując cenne spostrzeżenia do analizy i zrozumienia zbioru danych.

Stanowisko	Rodzaj_zatr...	Poziom_...	Poziom_wiedzy...	Wynagrodz...	Waluta_wynagrodzenia	Miejsce_Firmy	Wynagrodz...	Miejsce...	Rozmiar_firmy	Rok
Business Intelligence Engineer	Full-Time	Senior	Expert	116300	United States Dollar	United States	116300	United		
Business Intelligence Engineer	Full-Time	Senior	Expert	84600	United States Dollar	United States	84600	United		
AI Engineer	Full-Time	Senior	Expert	129700	United States Dollar	United States	129700	United		
AI Engineer	Full-Time	Senior	Expert	86500	United States Dollar	United States	86500	United		
Data Analyst	Full-Time	Senior	Expert	150000	United States Dollar	United States	150000	United		
Data Analyst	Full-Time	Senior	Expert	120000	United States Dollar	United States	120000	United		
Data Analyst	Full-Time	Entry	Junior	60000	British Pound Sterling	United Kingdom	73824	United		
Data Analyst	Full-Time	Entry	Junior	50000	British Pound Sterling	United Kingdom	61520	United		
Machine Learning Engineer	Full-Time	Senior	Expert	225300	United States Dollar	United States	225300	United		
Machine Learning Engineer	Full-Time	Senior	Expert	197400	United States Dollar	United States	197400	United		
Data Scientist	Full-Time	Senior	Expert	212000	United States Dollar	United States	212000	United		
Data Scientist	Full-Time	Senior	Expert	93300	United States Dollar	United States	93300	United		
Data Analyst	Full-Time	Senior	Expert	140130	United States Dollar	United States	140130	United		
Data Analyst	Full-Time	Senior	Expert	84870	United States Dollar	United States	84870	United		
Insight Analyst	Full-Time	Mid	Intermediate	45000	British Pound Sterling	United Kingdom	55368	United		
Insight Analyst	Full-Time	Mid	Intermediate	25000	British Pound Sterling	United Kingdom	30760	United		
Data Engineer	Full-Time	Senior	Expert	266000	United States Dollar	United States	266000	United		
Data Engineer	Full-Time	Senior	Expert	152000	United States Dollar	United States	152000	United		
Data Science Manager	Full-Time	Senior	Expert	224000	United States Dollar	United States	224000	United		
Data Science Manager	Full-Time	Senior	Expert	165000	United States Dollar	United States	165000	United		
Machine Learning Engineer	Full-Time	Senior	Expert	255000	United States Dollar	United States	255000	United		
Machine Learning Engineer	Full-Time	Senior	Expert	169000	United States Dollar	United States	169000	United		
Data Engineer	Full-Time	Mid	Intermediate	169000	United States Dollar	United States	169000	United		
Data Engineer	Full-Time	Mid	Intermediate	91000	United States Dollar	United States	91000	United		
Data Scientist	Full-Time	Senior	Expert	197500	United States Dollar	United States	197500	United		
Data Scientist	Full-Time	Senior	Expert	136500	United States Dollar	United States	136500	United		

Rysunek 1 Tabela wykorzystana jako źródło danych do analiz

## Przygotowanie do pracy

Na początku utworzony został folder „Projekt”. Utworzona została także biblioteka o nazwie ProjektB do której zaimportowano dane pobrane ze strony Kaggle.com

```
9 FILENAME REFFILE '/home/u63619470/Projekt/Pliki/v6_Latest_Data_Science_Salaries.csv';
10
11 PROC IMPORT DATAFILE=REFFILE
12     DBMS=CSV
13     OUT=PROJEKTB.DaneDoProjektu;
14     GETNAMES=YES;
15 RUN;
16
17 PROC CONTENTS DATA=PROJEKTB.DaneDoProjektu; RUN;
18
19
20 %web_open_table(PROJEKTB.DaneDoProjektu);
```

Rysunek 2 Zaimportowanie danych do biblioteki ProjektB

Następnie utworzono tabelę za pomocą tego kodu:

```
1 PROC SQL;
2 CREATE TABLE WORK.query AS
3 SELECT 'Job Title'n , 'Employment Type'n ,
4 'Experience Level'n , 'Expertise Level'n ,
5 Salary , 'Salary Currency'n , 'Company Location'n ,
6 'Salary in USD'n , 'Employee Residence'n , 'Company Size'n ,
7 'Year'n FROM _TEMP3.danedoprojektu;
8 RUN;
9 QUIT;
10
11 PROC DATASETS NOLIST NODETAILS;
12 CONTENTS DATA=WORK.query OUT=WORK.details;
13 RUN;
14
15 PROC PRINT DATA=WORK.details;
16 RUN;
```

Rysunek 3 Stworzenie tabeli z danymi

Kolejno za pomocą Data Stepu przetłumaczono nazwy kolumn na język polski.

```
14 data;
15 set _temp3.danedoprojektu;
16 rename 'Job Title'n = Stanowisko;
17 rename 'Employment Type'n = Rodzaj_zatrudnienia;
18 rename 'Experience Level'n = Poziom_doswiadczenia;
19 rename 'Expertise Level'n = Poziom_wiedzy_specjalistycznej;
20 rename Salary = Wynagrodzenie;
21 rename 'Salary Currency'n = Waluta_wynagrodzenia;
22 rename 'Company Location'n = Miejsce_Firmy;
23 rename 'Salary in USD'n = Wynagrodzenie_w_USD;
24 rename 'Employee Residence'n = Miejsce_zamieszkania_pracownika;
25 rename 'Company Size'n = Rozmiar_firmy;
26 rename 'Year'n = Rok;
27 keep 'Job Title'n 'Employment Type'n 'Experience Level'n 'Expertise Level'n
28 Salary 'Salary Currency'n 'Company Location'n 'Salary in USD'n 'Employee Residence'n 'Company Size'n 'Year'n;
29 run;
```

Rysunek 4 Przetłumaczenie nazw kolumn na polski

Firmy które są analizowane w tej pracy znajdują się na całym świecie, jednak większość z nich znajduje się w Stanach Zjednoczonych. Aby to pokazać skorzystano z zadania „Listowanie danych”.

```
18 title "Miejsca w których znajdują się analizowane firmy";
19
20 proc freq data=_TEMP0.TABELAPOLSKI;
21     tables Miejsce_Firmy;
22 run;
```

Rysunek 5 Kod wywołujący listę miejsc w których znajdują się analizowane firmy

Dodano także filtr wybierający jedynie rekordy zawierające firmy znajdujące się w Stanach Zjednoczonych.

Miejsce_Firmy	Liczebność
United States	3888

Rysunek 6 Liczebność firm znajdujących się w Stanach Zjednoczonych

Wszystkich rekordów jest 4976, więc firmy znajdujące się w stanach zjednoczonych to 78,14% wszystkich rekordów.

## Analiza procesów

### 1. Wpływ doświadczenia na wynagrodzenie. Określenie, w jaki sposób poziom doświadczenia wpływa na wysokość wynagrodzeń w branży Data Science

#### 1. Wyjaśnienie celu szczegółowego:

Celem analizy jest zbadanie tego jakie są zarobki w branży Data Science w zależności od posiadanego poziomu doświadczenia w Stanach Zjednoczonych.

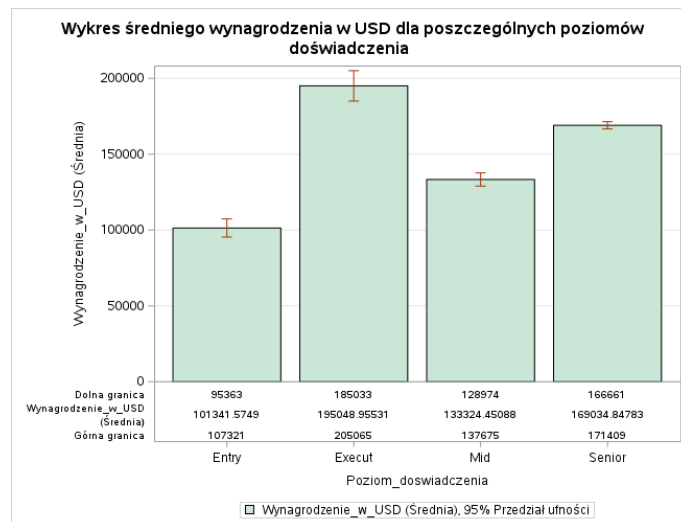
#### 2. Opis procesu uzyskania danych wynikowych:

Tę analizę rozpoczęto od przygotowania wykresu słupkowego w którym na osi X znajdują się poziomy doświadczenia pracowników, a na osi Y średnia wartość wynagrodzenia w USD wartość wynagrodzenia została uśredniona. Na wykresie dodano także słupki odpowiadające za prezentację 95% przedziału ufności. Spodziewać się można, że wraz ze wzrostem poziomu doświadczenia wzrośnie także średnie wynagrodzenie.

```
15 ods graphics / reset width=6.8in height=5.2in imagemap;  
16  
17 proc sgplot data=_TEMP0.TABELAPOLSKI;  
18     vbar Poziom_doswiadczenia / response=Wynagrodzenie_w_USD  
19         fillattrs=(color=CXcae6d6) datalabel limits=both limitstat=clm stat=mean;  
20     yaxis grid;  
21 run;  
22  
23 ods graphics / reset;
```

Rysunek 7 Kod tworzący wykres słupkowy

Na wykresie poniżej zostały przedstawione średnie roczne poziomy wynagrodzeń w zależności od poziomu doświadczenia pracowników.



Rysunek 8 Wykres średniego poziomu wynagrodzenia w USD dla poszczególnych poziomów doświadczenia w Stanach Zjednoczonych

### 3. Wnioski

Zgodnie z oczekiwaniami możemy zaobserwować, że wraz ze wzrostem poziomu doświadczenia wynagrodzenie także wzrasta.

Interpretacja wyników w rosnącej kolejności wynagrodzeń:

1. Najmniej zarabia osoba z doświadczeniem podstawowym (entry). Średnie wynagrodzenie z tym poziomem doświadczenia w zaokrągleniu do dolara wynosi 101342\$. Oprócz tego można stwierdzić, że z 95% poziomem ufności wynagrodzenie osoby z tym poziomem doświadczenia będzie mieściło się w przedziale od 95363\$ do 107321\$.
2. Więcej zarabia pracownik z doświadczeniem średniego poziomu (mid). Średnie wynagrodzenie z tym poziomem doświadczenia w zaokrągleniu do dolara wynosi 133324\$. Przedział wynagrodzenia takiego pracownika z 95% poziomem ufności to 128974\$ do 137675\$.
3. Średnie wynagrodzenie pracownika z doświadczeniem wyższego poziomu (senior) w zaokrągleniu do dolara wynosi 169035\$. Przedział wynagrodzenia takiego pracownika z 95% poziomem ufności to od 166661\$ do 171409\$.
4. Najwięcej zarabia pracownik z doświadczeniem poziomu kierowniczego (executive). Średnie wynagrodzenie pracownika z takim poziomem doświadczenia w zaokrągleniu do dolara wynosi 195049\$. Przedział wynagrodzenia takiego pracownika z 95% przedziałem ufności to od 185033\$ do 205065\$.



## 2. Analiza wynagrodzeń w zależności od zajmowanego stanowiska

### 1. Wyjaśnienie celu szczegółowego:

Celem analizy jest zbadanie tego na jakich stanowiskach w branży Data Science zarabia się najwięcej, a na jakich najmniej.

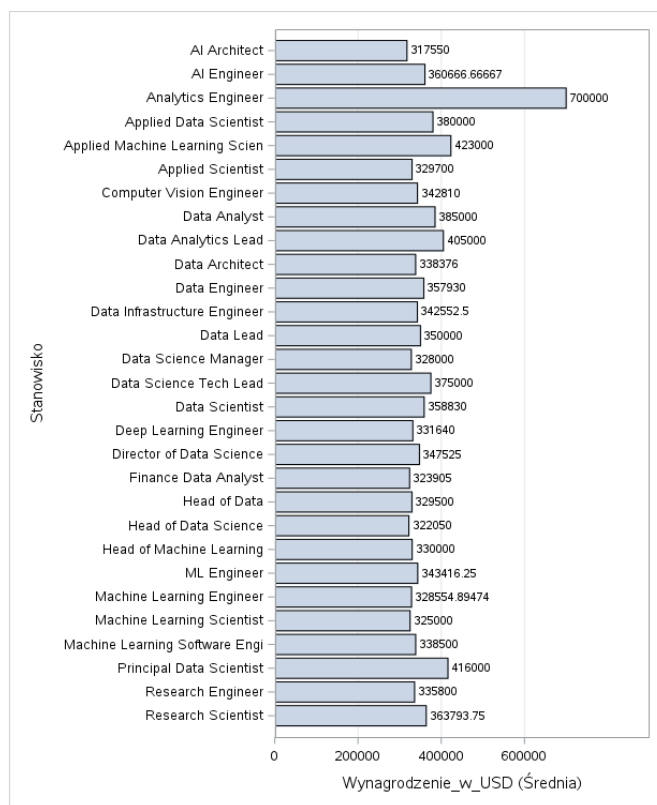
### 2. Opis procesu uzyskania danych wynikowych:

Analizę rozpoczęto od przygotowania wykresu słupkowy w którym na osi X znajduje się uśredniony poziom wynagrodzenia a na osi Y nazwa zajmowanego stanowiska.

Dodano także warunek: „Wynagrodzenie\_w\_USD > 300000” aby pośród wyświetlanych wyników były tylko stanowiska na których pracownik zarabia średnio rocznie więcej niż 300000\$.

```
15 ods graphics / reset width=6.4in height=7.8in imagemap;
16
17 proc sgplot data=_TEMP0.TABELAPOLSKI (where=(Miejsce_Firmy='United States' and
18 Wynagrodzenie_w_USD > 300000));
19 hbar Stanowisko / response=Wynagrodzenie_w_USD datalabel stat=mean;
20 xaxis grid;
21 run;
22
23 ods graphics / reset;
```

Rysunek 9 Kod tworzący wykres słupkowy



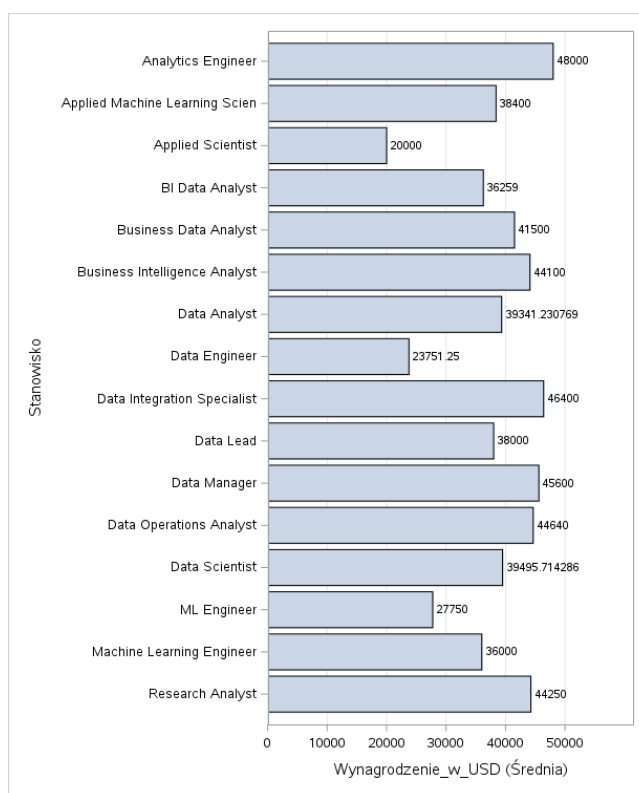
Rysunek 10 Wykres średniego poziomu wynagrodzenia w USD w zależności od zajmowanego stanowiska

Następnie przygotowano kolejny wykres słupkowy. Zmieniono jednak warunek wybierający tylko stanowiska z wynagrodzeniem większym niż 300000\$ na warunek:

Wynagrodzenie\_w\_USD < 50000. W ten sposób wyświetlą się tylko wyniki w których średnie roczne wynagrodzenie jest mniejsze niż 50000\$.

```
15 ods graphics / reset width=6.4in height=7.8in imagemap;
16
17 proc sgplot data=_TEMP0.TABELAPOLSKI (where=(Miejsce_Firmy='United States' and
18     Wynagrodzenie_w_USD < 50000));
19     hbar Stanowisko / response=Wynagrodzenie_w_USD datalabel stat=mean;
20     xaxis grid;
21 run;
22
23 ods graphics / reset;
```

Rysunek 11 Kod tworzący wykres słupkowy



Rysunek 12 Wykres średniego poziomu wynagrodzenia w USD w zależności od zajmowanego stanowiska

### **3. Wnioski:**

#### **5 stanowisk z najwyższym średnim rocznym wynagrodzeniem:**

1. Stanowisko na którym zarabia się najwięcej to Analytics Engineer. Osoba na takim stanowisku średnio rocznie zarabia aż 700000\$. Jest to wynik zdecydowanie większy niż pozostałe.
2. Na drugim miejscu jest stanowisko Applied Machine Learning Scientist. Na tym stanowisku średnie roczne wynagrodzenie wynosi 423000\$. Mimo tego, że jest to druga pozycja, to różnica między średnim wynagrodzeniem Analytics Engineer a Applied Machine Learning Scientist jest bardzo duża. Wynosi aż 277000\$.
3. Na trzecim miejscu plasuje się stanowisko Principal Data Scientist ze średnim rocznym wynagrodzeniem na poziomie 416000\$. Tutaj już nie ma tak dużej różnicy w porównaniu z wynagrodzeniem Applied Machine Learning Scientist.
4. Na czwartym miejscu stanowisko Data Analytics Lead. Na tym stanowisku średnie roczne wynagrodzenie wynosi 405000\$. Jest to także ostatnie stanowisko na którym średnie roczne wynagrodzenie przekracza 400000\$.
5. Na piątym miejscu stanowisko Data Analyst. Osoba na tym stanowisku średnio rocznie zarabia 385000\$.

#### **5 stanowisk z najniższym średnim rocznym wynagrodzeniem:**

1. Najmniej zarabia osoba na stanowisku Applied Scientist. Średnie roczne wynagrodzenie na tym stanowisku wynosi zaledwie 20000\$.
2. Nieco więcej zarabia Data Engineer. Jego średnie roczne wynagrodzenie to 23751,25\$.
3. Trzecim stanowiskiem o najniższym średnim rocznym wynagrodzeniu jest ML Engineer. Osoba na takim stanowisku może liczyć na 27750\$. Jest to także ostatnie stanowisko na którym średnie roczne wynagrodzenie nie przekracza 30000\$.
4. Więcej wynosi średnie roczne wynagrodzenie na stanowisku Machine Learning Engineer. Jest to 36000\$.
5. Na piątym miejscu stanowisk o najniższym średnim rocznym wynagrodzeniu jest BI Data Analyst. Wynagrodzenie na tym stanowisku wynosi 36259\$.

### 3. Analiza rozkładu wynagrodzeń w branży Data Science w zależności od rozmiaru firmy

#### 1. Wyjaśnienie celu szczegółowego:

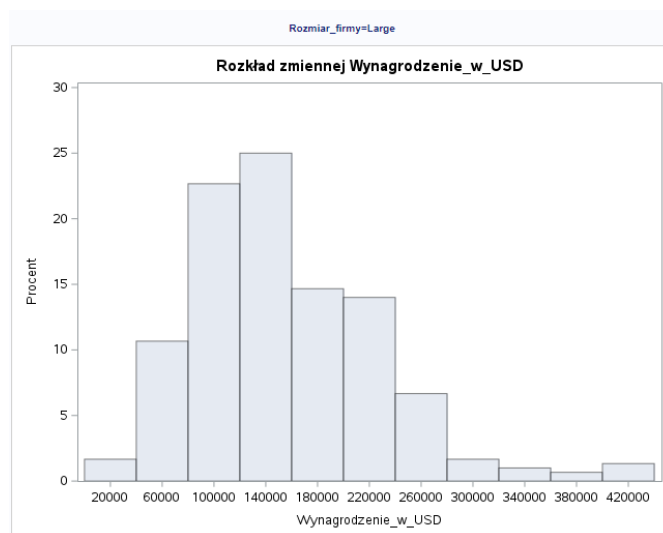
Celem analizy jest zbadanie rozkładu wynagrodzeń w branży Data Science oraz przeprowadzenie testu na rozkład normalny dla poszczególnych wykresów z podziałem na rozmiar firmy.

#### 2. Opis procesu uzyskania danych wynikowych:

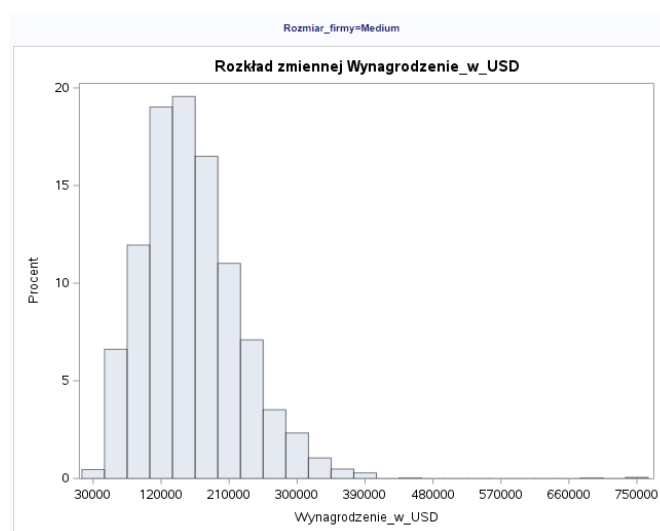
Analizę rozpoczęto od przygotowania wykresu analizy rozkładu. Na osi X znajduje się roczne wynagrodzenie, na osi Y procent stanowisk oferujących taki poziom wynagrodzenia. Wykresy zostały także podzielone ze względu na rozmiar firmy: Large, Medium oraz Small. Oprócz tego chciałem także test na rozkład normalny dla każdego z wykresów. Pod uwagę będzie jednak wzięty tylko wykres dla firm średniego rozmiaru, ponieważ rekordów dotyczących tych firm jest aż 3522. Natomiast rekordów dla firm dużych i małych jest kolejno jedynie 300 i 66.

```
15 ods noproctitle;
16 ods graphics / imagemap=on;
17
18 proc sort data=_TEMP0.TABELAPOLSKI out=Work.SortTempTableSorted;
19   where Miejsce_Firmy='United States';
20   by Rozmiar_firmy;
21 run;
22
23 /* EKSPLOKACJA DANYCH */
24 proc univariate data=Work.SortTempTableSorted;
25   ods select Histogram;
26   var Wynagrodzenie_w_USD;
27   histogram Wynagrodzenie_w_USD;
28   by Rozmiar_firmy;
29 run;
30
31 proc univariate data=Work.SortTempTableSorted;
32   ods select Histogram GoodnessOfFit;
33   var Wynagrodzenie_w_USD;
34
35   /* SPRAWDZANIE NORMALNOŚCI */
36   histogram Wynagrodzenie_w_USD / normal(mu=est sigma=est);
37   by Rozmiar_firmy;
38 run;
39
40 proc delete data=Work.SortTempTableSorted;
41 run;
```

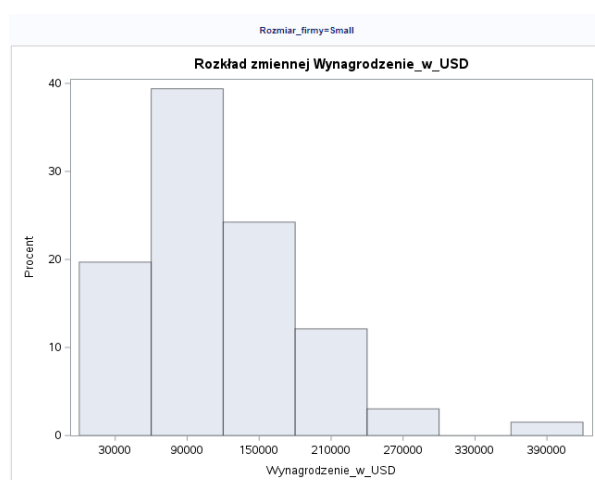
Rysunek 13 Kod tworzący wykresy analizy rozkładu oraz testy na rozkład normalny



Rysunek 14 Wykres rozkład wynagrodzeń w USD dla firmy rozmiaru Large



Rysunek 15 Wykres rozkład wynagrodzeń w USD dla firmy rozmiaru Medium



Rysunek 16 Wykres rozkład wynagrodzeń w USD dla firmy rozmiaru Small

### **3. Wnioski:**

#### **Dla firmy rozmiaru Large:**

Rozkład wynagrodzeń jest asymetryczny prawostronny, dodatni. Oznacza to, że większa część stanowisk pracy oferuje wynagrodzenie poniżej wartości przeciętnej.

Aż 25% stanowisk oferuje roczne wynagrodzenie na poziomie 140000\$.

Następnie 22,667% stanowisk oferuje roczne wynagrodzenie na poziomie 100000\$.

14,667% stanowisk ma roczne wynagrodzenie na poziomie 180000\$.

Oznacza to, że 62,334% stanowisk w dużych firmach oferuje roczne wynagrodzenie nie mniejsze niż 100000\$.

#### **Dla firmy rozmiaru Medium:**

Ponownie rozkład wynagrodzeń jest asymetryczny prawostronny, dodatni.

Dla 19,563% stanowisk roczne wynagrodzenie jest na poziomie 150000\$.

19,023% stanowisk oferuje roczne wynagrodzenie na poziomie 120000\$.

16,496% stanowisk ma roczne wynagrodzenie na poziomie 180000\$.

Oznacza to, że w średnich firmach 55,055% stanowisk oferuje roczne wynagrodzenie nie mniejsze niż 120000\$.

#### **Dla firm rozmiaru Small:**

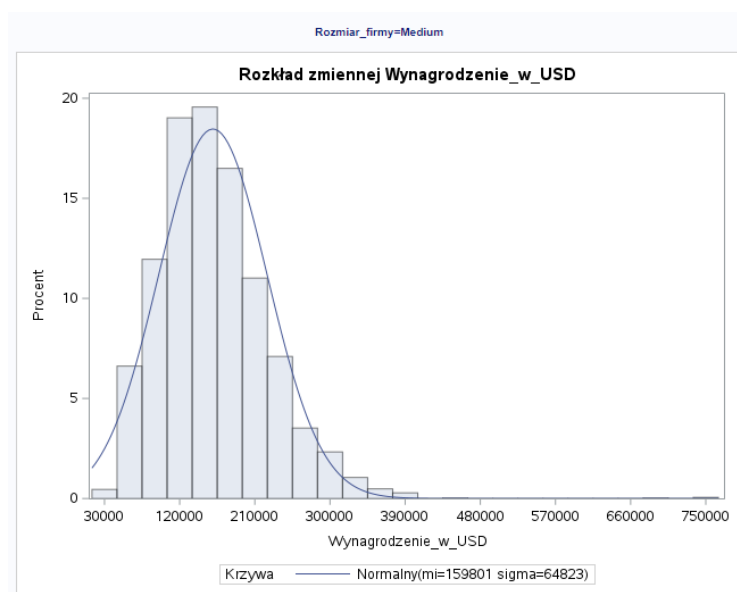
39,394% stanowisk oferuje roczne wynagrodzenie na poziomie 90000\$.

24,242% stanowisk oferuje roczne wynagrodzenie na poziomie 150000\$.

19,697% stanowisk oferuje roczne wynagrodzenie na poziomie 30000\$.

W małych firmach 59,091% stanowisk oferuje roczne wynagrodzenie nie mniejsze niż 90000\$. (Nie wzięto pod uwagę trzeciej największej grupy, gdyż już suma procentów dwóch największych grup przekracza 50%.)

## Test na rozkład normalny dla firm średniego rozmiaru



Rysunek 17 Wykres analiza rozkładu wynagrodzenia z dopasowanym wykresem rozkładu normalnego

Dopasowany rozkład normalny dla Wynagrodzenie_w_USD				
Rozmiar_firmy=Medium				
Testy dopasowania dla rozkładu Normalny				
Testowanie	Statystyka		Wartość p	
Kolmogorow-Smirnow	D	0.0590429	Pr. > D	<0.010
Cramer-von Mises	W-kwadr.	3.6604830	Pr. > W-kwadr.	<0.005
Anderson-Darling	A-kwadr.	23.2509509	Pr. > A-kwadr.	<0.005

Rysunek 18 Test na rozkład normalny

Wartość p jest  $<0.05$  dlatego odrzucamy  $H_0$  mówiącą, że zmienna ma rozkład normalny. Wynagrodzenie w USD dla firm średniego rozmiaru nie ma rozkładu normalnego.

#### 4. Analiza wynagrodzeń w zależności od rodzaju zatrudnienia

##### 1. Wyjaśnienie celu szczegółowego:

Celem analizy jest przeanalizowanie poziomu wynagrodzeń w zależności od rodzaju zatrudnienia.

##### 2. Opis procesu uzyskania danych wynikowych:

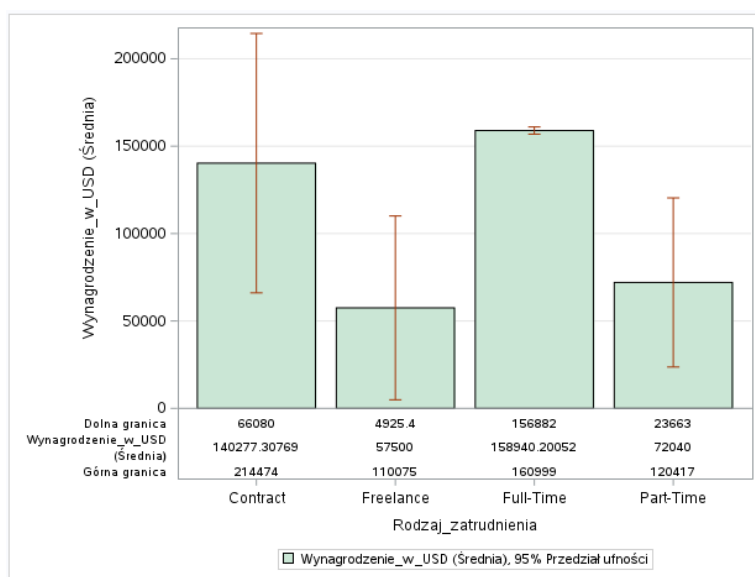
Analizę rozpoczęto od przygotowanie wykresu słupkowego. Na osi X wypisane są rodzaje zatrudnienia:

- Contract
- Freelance
- Full-Time
- Part-Time

Na osi Y natomiast uśrednione roczne wynagrodzenie w USD. Dodatkowo dodano 95% przedziały ufności dla poziomu wynagrodzeń.

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;  
16  
17 proc sgplot data=_TEMP0.TABELAPOLSKI (where=(Miejsce_Firmy='United States'));  
18     vbar Rodzaj_zatrudnienia / response=Wynagrodzenie_w_USD  
19         fillattrs=(color=CXcae6d5) datalabel limits=both limitstat=c|n stat=mean;  
20     yaxis grid;  
21 run;  
22  
23 ods graphics / reset;
```

Rysunek 19 Kod tworzący wykres słupkowy



Rysunek 20 Wykres słupkowy średnie roczne wynagrodzenie w zależności od rodzaju zatrudnienia



### **3. Wnioski:**

#### **Full-Time**

Najwyższe średnie zarobki oferują stanowiska z zatrudnieniem w pełnym wymiarze czasu pracy. Średnie zarobki osoby z takim rodzajem zatrudnienia wynoszą 158940,20\$. Z 95% poziomem ufności można stwierdzić, że średnie roczne zarobki będą należeć do przedziału od 156882\$ do 160999\$. Jest to bardzo mały przedział w porównaniu z tymi które zostaną przedstawione następnie. Różnica między górną a dolną granicą tego przedziału wynosi zaledwie 4117\$. Oznacza to, że można naprawdę z dużym prawdopodobieństwem ocenić średnie roczne wynagrodzenie osoby zatrudnionej w pełnym wymiarze czasu pracy.

#### **Part-Time**

Stanowiska z zatrudnieniem w niepełnym wymiarze czasu pracy oferują średnie roczne zarobki na poziomie 72040\$. Z 95% poziomem ufności można stwierdzić, że średnie roczne zarobki będą mieściły się w przedziale od 23663\$ do 120417\$. W przeciwieństwie do stanowisk z rodzajem zatrudnienia Full-Time, różnica między górną a dolną granicą przedziału jest znacząca i wynosi 96754\$.

#### **Contract**

Na drugim miejscu pod względem średniej rocznej wysokości wynagrodzenia znajdują się stanowiska z rodzajem zatrudnienia contract. Średnie wynagrodzenie wynosi 140277,31\$. 95% poziom ufności pozwala stwierdzić, że średnie roczne zarobki będą mieściły się w przedziale od 66080\$ do 214474\$. Różnica między górną a dolną granicą tego przedziału wynosi 148394\$. Jest to największa różnica spośród analizowanych wyników.

#### **Freelance**

Najniższych średnich rocznych wynagrodzeń mogą się spodziewać freelancerzy. Średnie wynagrodzenie dla stanowisk z rodzajem zatrudnienia freelance wynosi 57500\$. Z 95% poziomem ufności można stwierdzić, że wynagrodzenie będzie należało do przedziału od 4925,40\$ do 110075\$. Różnica między górną a dolną granicą przedziału jest bardzo duża w porównaniu do średnich zarobków, ponieważ stanowi ona aż 182,87% średniego wynagrodzenia.

## 5. Analiza średniego rocznego wynagrodzenia według poziomu doświadczenia oraz poziomu wiedzy specjalistycznej w latach 2020-2023.

### 1. Wyjaśnienie celu szczegółowego:

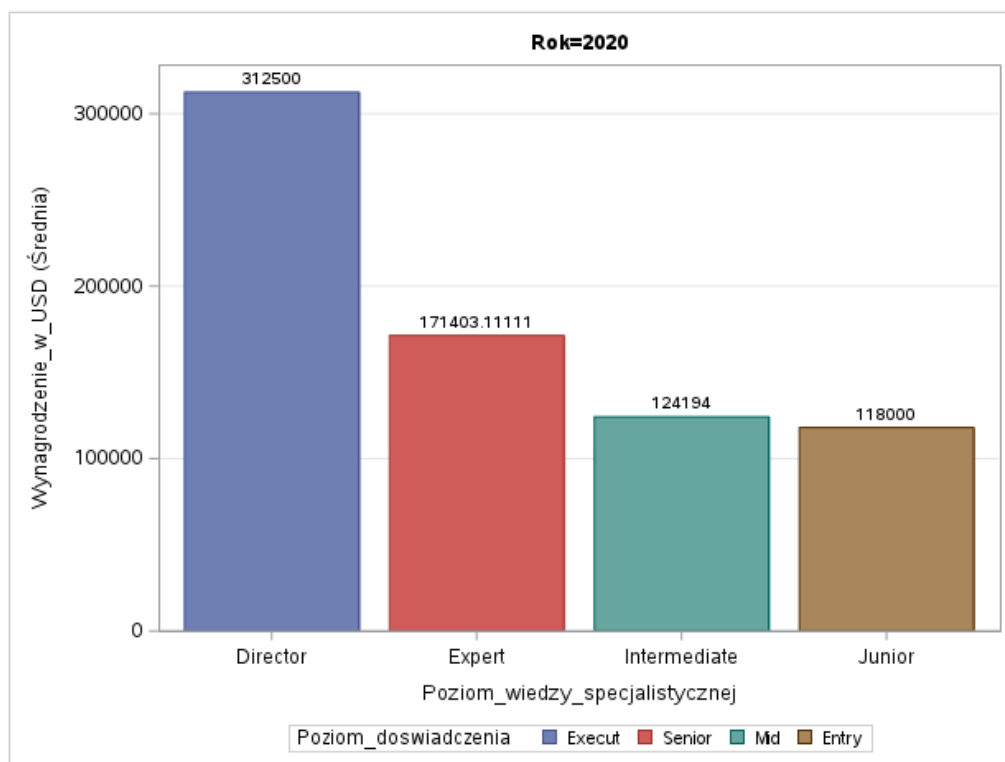
Celem analizy jest przeanalizowanie poziomu wynagrodzeń w zależności od poziomu doświadczenia oraz poziomu wiedzy specjalistycznej na przestrzeni lat.

### 2. Opis procesu uzyskania danych wynikowych:

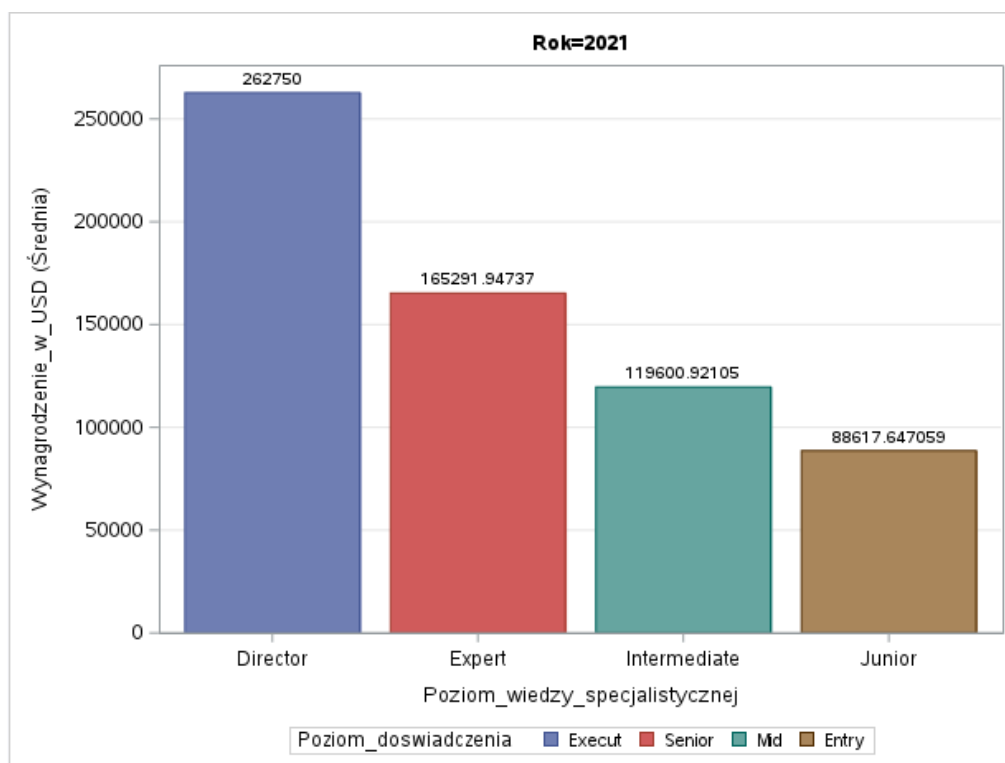
Analizę rozpoczęto od przygotowania wykresu słupkowego. Na osi X wypisane są: poziom wiedzy specjalistycznej oraz poziom doświadczenia, natomiast na osi Y średnie roczne wynagrodzenie w USD.

```
15 ods graphics / reset width=6.4in height=4.8in imagemap;
16
17 proc sort data=_TEMP2.TABELAPOLSKI out=_BarChartTaskData;
18     by Rok;
19 run;
20
21 proc sgplot data=_BarChartTaskData (where=(Miejsce_Firmy='United States'));
22     by Rok;
23     vbar Poziom_wiedzy_specjalistycznej / response=Wynagrodzenie_w_USD
24         group=Poziom_doswiadczenia groupdisplay=cluster datalabel stat=mean;
25     yaxis grid;
26 run;
27
28 ods graphics / reset;
29
30 proc datasets library=WORK noprint;
31     delete _BarChartTaskData;
32 run;
```

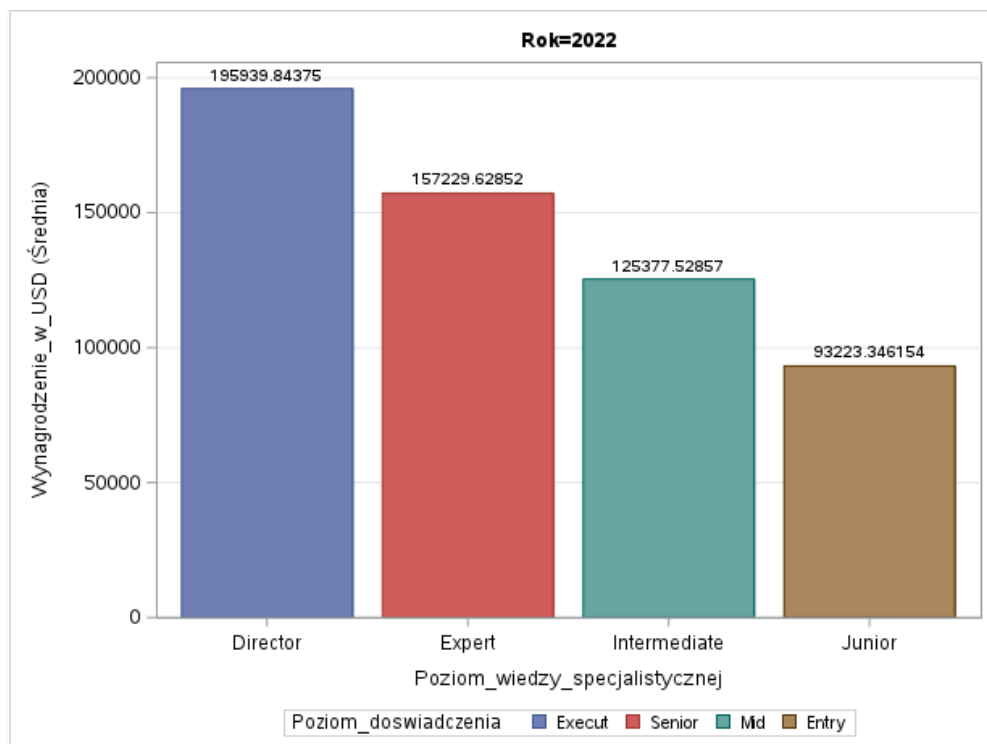
*Rysunek 21 Kod tworzący wykres słupkowy*



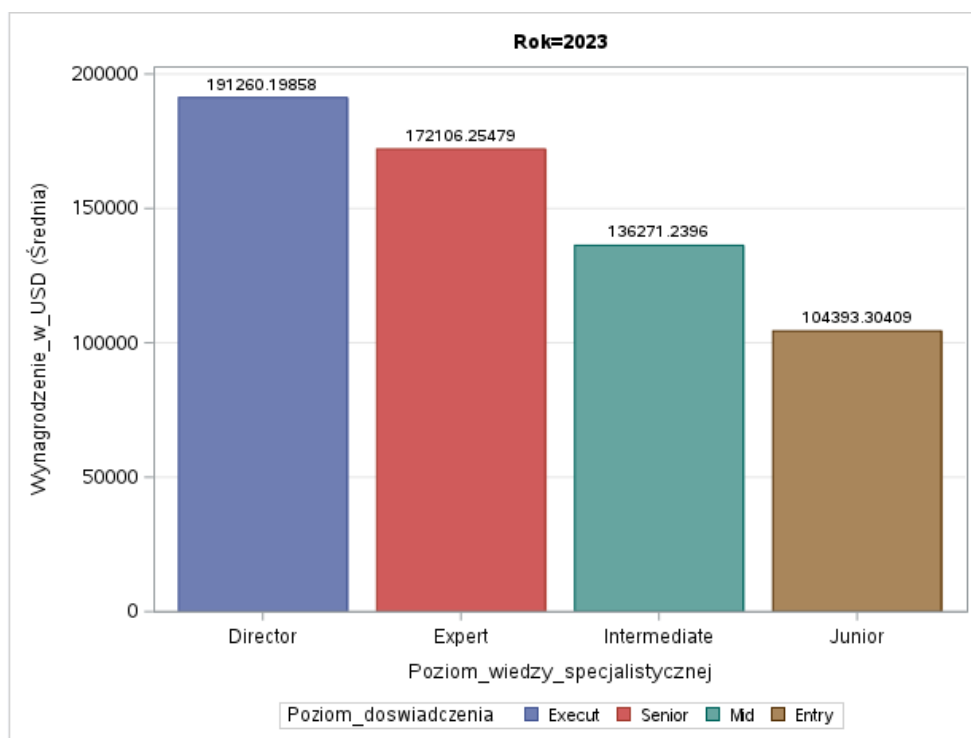
Rysunek 22 Wykres słupkowy średnie wynagrodzenie w zależności od poziomu wiedzy specjalistycznej i doświadczenia



Rysunek 23 Wykres słupkowy średnie wynagrodzenie w zależności od poziomu wiedzy specjalistycznej i doświadczenia



Rysunek 24 Wykres słupkowy średnie wynagrodzenie w zależności od poziomu wiedzy specjalistycznej i doświadczenia



Rysunek 25 Wykres słupkowy średnie wynagrodzenie w zależności od poziomu wiedzy specjalistycznej i doświadczenia

### **3. Wnioski:**

#### **Executive Director**

W każdym roku najwięcej zarabiają pracownicy z wiedzą specjalistyczną na poziomie dyrektorskim (Director) oraz doświadczeniem poziomu kierowniczego (Executive).

Najwyższe średnie zarobki pracownicy z takimi kwalifikacjami osiągnęli w roku 2020. Wyniosły one wtedy 312500\$.

Najniższe średnie wynagrodzenie pracownicy z takimi kwalifikacjami otrzymali w 2023 roku. Wyniosło ono wtedy 191260,20\$.

#### **Senior Expert**

Na drugim miejscu pod względem średniego wynagrodzenia są pracownicy z wiedzą specjalistyczną na poziomie eksperckim (Expert) oraz z wyższym poziomem doświadczenia (Senior).

W przeciwieństwie do poprzednio analizowanej grupy, pracownicy z takimi kwalifikacjami osiągnęli najwyższe średnie wynagrodzenie w roku 2023. Wyniosło ono wtedy 172106,25\$.

Najniższe średnie wynagrodzenie dla tej grupy pracowników przypadło na rok 2022 i wyniosło 157229,63\$.

#### **Mid Intermediate**

Na trzecim miejscu znajdują się pracownicy z wiedzą specjalistyczną na poziomie średniozaawansowanym (Intermediate) oraz z średnim poziomem doświadczenia (Mid).

Najwyższe średnie wynagrodzenie Ci pracownicy osiągnęli w roku 2023. Wyniosło ono wtedy 136271,24\$.

Najniższe średnie wynagrodzenie przypadło na rok 2021 i wyniosło 119600,92\$.

#### **Entry Junior**

Na ostatnim miejscu w tym zestawieniu znajdują się pracownicy z wiedzą specjalistyczną na poziomie juniorskim oraz doświadczeniu na poziomie początkującym (Entry)

Najwyższe średnie wynagrodzenie pracownicy z tymi kwalifikacjami otrzymali w roku 2020. Wyniosło ono wtedy 11800\$. Natomiast najniższe, czyli 88617,65\$ w 2021r.

## **Podsumowanie**

### **Wpływ Doświadczenia na Wynagrodzenie w Branży Data Science:**

Wraz ze wzrostem poziomu doświadczenia, średnie wynagrodzenie również rośnie.

Najniższe zarobki notowane są dla osób z doświadczeniem entry, a najwyższe dla pracowników na stanowiskach executives.

### **Analiza Wynagrodzeń według Stanowiska:**

Najwyższe roczne wynagrodzenia osiągają stanowiska typu Analytics Engineer, Applied Machine Learning Scientist, oraz Principal Data Scientist.

Najniższe wynagrodzenia przypisane są stanowiskom takim jak Applied Scientist czy Data Engineer.

### **Analiza Rozkładu Wynagrodzeń w Zależności od Rozmiaru Firmy:**

Firmy o różnych rozmiarach wykazują różnice w rozkładzie wynagrodzeń.

W dużych firmach większość stanowisk oferuje wynagrodzenia przekraczające 100000\$, natomiast w mniejszych firmach ta granica jest niższa.

### **Analiza Wynagrodzeń według Rodzaju Zatrudnienia:**

Średnie wynagrodzenie dla zatrudnienia Full-Time jest znacznie wyższe niż dla innych rodzajów zatrudnienia.

Freelancerzy otrzymują najniższe wynagrodzenia w porównaniu z innymi rodzajami zatrudnienia.

### **Analiza Średniego Roczego Wynagrodzenia według Poziomu Doświadczenia oraz Wiedzy Specjalistycznej w Latach 2020-2023:**

Pracownicy na stanowiskach executives oraz senior experts osiągają najwyższe wynagrodzenia.

Trend wzrostowy w średnich wynagrodzeniach jest zauważalny od 2020 do 2023 roku.