

# ANOVA dwuczynnikowa – wpływ doświadczenia oraz pozycji w firmie na roczne wynagrodzenie programisty.

## 1.Opis zmiennych

Przed przystąpieniem do analizy należy opisać badane zmienne oraz przybliżyć kim są frontend developera oraz backend developera.

Programista frontend zajmuje się tworzeniem stron internetowych i aplikacji mobilnych. Jest on odpowiedzialny za wygląd strony internetowej i jej interaktywną funkcjonalność. Developerzy w tej dziedzinie są odpowiedzialni za wszystko, co widoczne podczas nawigacji na stronie internetowej lub aplikacji. Programiści front-end używają przede wszystkim trzech języków: HTML , CSS , Javascript.

Backend Developer to programista, który jest odpowiedzialny za to, czego na pierwszy rzut oka nie widzi użytkownik danej strony czy aplikacji. Odpowiada on za to, aby dany system działał poprawie, zapewniając jego funkcjonalność techniczną. Sprawuje nadzór nad serwerowym zapleczem strony czy aplikacji, a często także bezpieczeństwem i rozwojem zastosowanych rozwiązań. Jego praca polega głównie na tworzeniu API oraz komunikacji z bazami danych. Programiści backend powinni być biegli w językach programowania, które działają po stronie serwera strony internetowej lub aplikacji takich jak: Java, PHP , języki na platformie .NET i Python.

Podsumowując, frontend developer buduje wygląd aplikacji, a backend developer buduje silnik działania aplikacji.

## 2.Opis badania i obserwacja danych

Poniższe badanie dotyczyć będzie rocznych zarobków programistów pracujących w branży IT. Za pomocą modelu ANOVA zbadane zostanie czy czynniki klasyfikujące takie jak lata doświadczenia oraz zajmowane stanowisko wpływają na wynagrodzenie programisty.

Dane do poniższego badania zostały pobrane ze strony: <https://www.kaggle.com/parulpandey/2020-it-salary-survey-for-eu-region>

Badana cecha: **Roczne wynagrodzenie programistów w EUR**

Czynniki klasyfikujące:

- **Zajmowane stanowisko w firmie:**
  - Frontend Developer
  - Backend Developer
- **Lata doświadczenia w branży IT:**
  - Od 0 (brak doświadczenia) do 4 lat
  - Od 5 do 9 lat
  - Powyżej 9 lat

Hipotezy badawcze testu ANOVA:

$H_{A0}$ : Zajmowane stanowisko w firmie nie różnicuje zarobków programistów

$H_{B0}$ : Doświadczenie w branży IT nie różnicuje zarobków programistów

$H_{AB0}$ : Wzajemna interakcja zmiennych doświadczenie oraz stanowisko nie różnicuje zarobków programistów

W dalszej części zostanie przybliżona struktura danych oraz ich statystyki opisowe.

Dane składają się z wybranych losowo 120 obserwacji na temat wynagrodzeń w branży IT pochodzących z roku 2020, a ich podział został przedstawiony w tabeli 1.

Tabela 1. Liczba zmiennych w zależności od podgrupy

	0-4	5-9	9+
<b>Backend Developer</b>	20	20	20
<b>Frontend Developer</b>	20	20	20

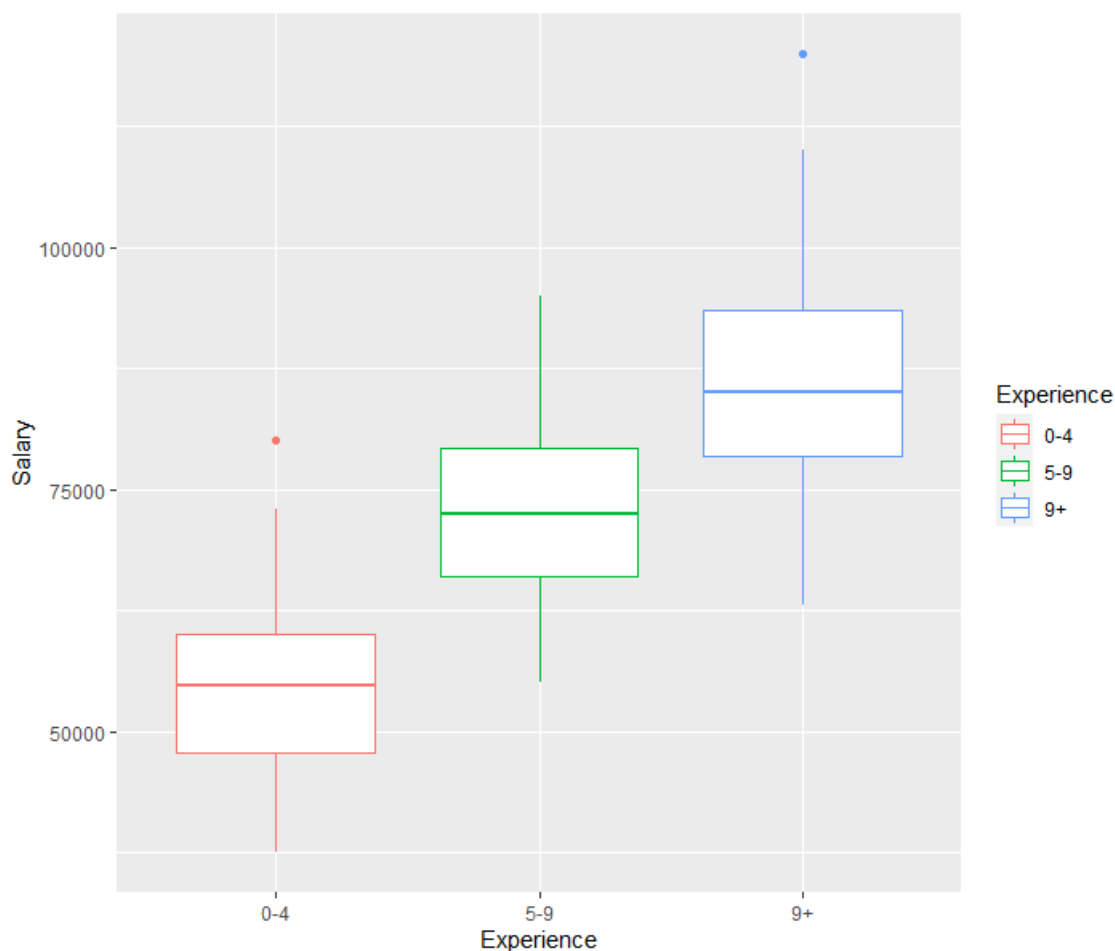
W tabeli 2. zostały zaprezentowane podstawowe statystyki opisowe dla zmiennych z podziałem względem lat doświadczenia. Zaobserwować można różnice w odchyleniu standardowym co może sugerować problem ze spełnieniem założenia o jednorodności wariancji.

Tabela 2. Statystyki opisowe dla danych grupowanych względem lat doświadczenia

item	group	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	0-4	1	40	54890.00	9883.65	54750	54378.12	10007.55	37500	80000	42500	0.37	-0.56	1562.74
2	5-9	1	40	72931.25	8644.83	72500	72520.31	9636.90	55000	95000	40000	0.38	-0.29	1366.87
3	9+	1	40	85576.50	12268.21	85000	84751.88	11119.50	63000	120000	57000	0.61	0.11	1939.77

Jako uzupełnienie powyższych danych na wykresie 1. zostały zaprezentowane wykresy pudełkowe względem lat doświadczenia w branży programistycznej. Na ich podstawie można przypuszczać, że wynagrodzenie developerów znacząco zależy od doświadczenia.

Wykres 1 Wykres pudełkowy rocznego wynagrodzenia w EUR względem lat doświadczenia



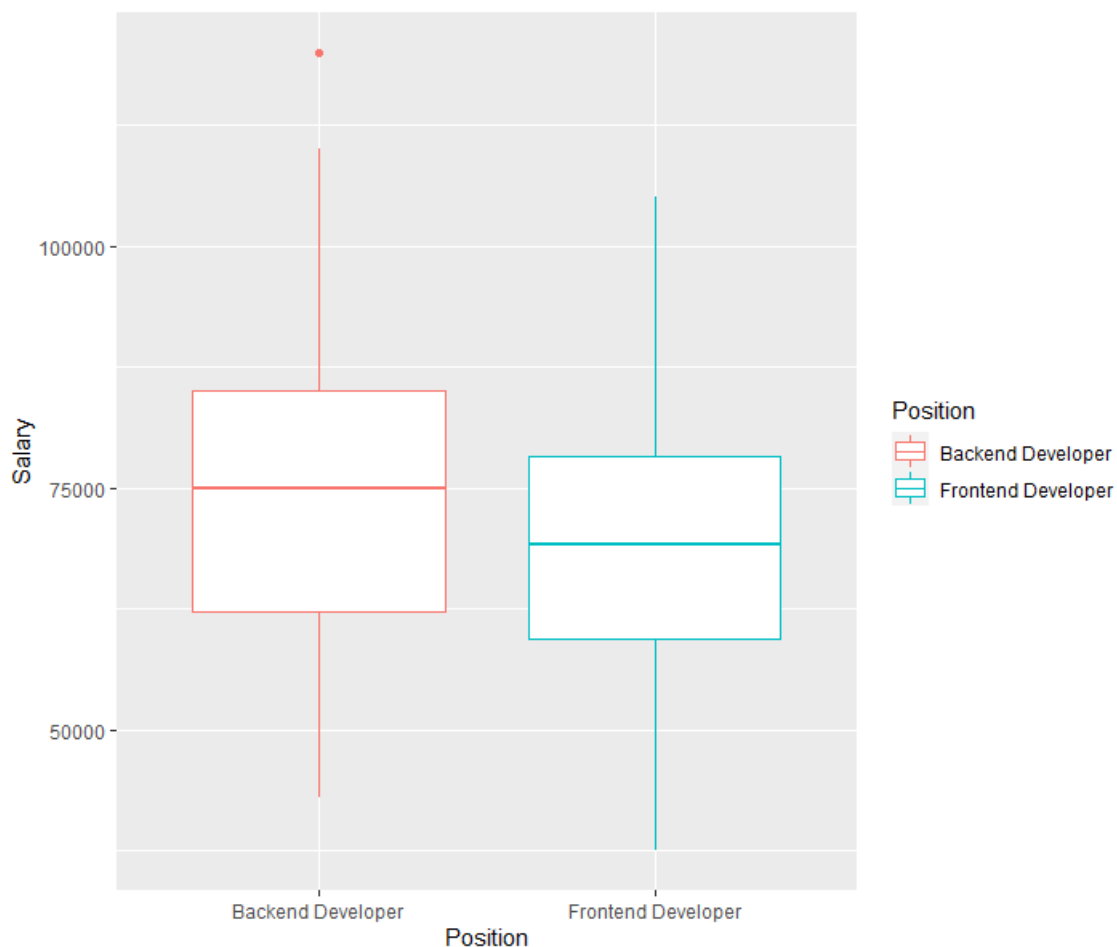
W tabeli 3. przedstawione zostały statystyki opisowe dla zmiennych z podziałem ze względu na zajmowane stanowisko. Z danych tych można wyczytać, że średnia oraz mediana są wyższe dla grupy backend developerów. Wariancja w obu grupach nie różni się znacząco co może sugerować, że nie będzie problemu z niejednorodnością wariancji.

Tabela 3. Statystyki opisowe dla danych grupowanych względem zajmowanego stanowiska

item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	Backend Developer	1	60	74133.33	16867.11	75000	73729.17	15270.78	43000	120000	77000	0.23	-0.26	2177.54
2	Frontend Developer	1	60	68131.83	15269.08	69125	68060.62	14084.70	37500	105000	67500	0.05	-0.41	1971.23

Na wykresie 2. zostały zaprezentowane wykresy pudełkowe dla zmiennej wynagrodzenie z podziałem na 2 grupy w zależności od pozycji. Zarobki dla grupy programistów backend wydają się być wyższe od grupy programistów frontend. W dalszej części zostanie zweryfikowane czy obejmowane stanowisko w znaczący sposób wpływa na otrzymywane wynagrodzenie.

Wykres 2. Wykres pudełkowy rocznego wynagrodzenia w EUR względem zajmowanego stanowiska



### 3.Sprawdzenie założeń ANOVA

#### Badanie normalności

Przed przystąpieniem do budowy modelu ANOVA zostały sprawdzone założenia o normalności rozkładów wyników zmiennej zależnej w każdej z analizowanych grup oraz jednorodności wariancji . Jako pierwsze zostało sprawdzone założenie o normalności za pomocą testów Shapiro-Wilka oraz Andersona-Darlinga. W przypadku obu testów hipoteza zerowa mówi, że próba pochodzi z populacji o rozkładzie normalnym. Wyniki zostały przedstawione w tabelach poniżej

Tabela 4. Wartości p testu Shapiro-Wilka z podziałem na grupy względem pozycji oraz doświadczenia

Position	Experience		
	0-4	5-9	9+
Backend Developer	0.420	0.826	0.783
Frontend Developer	0.567	0.995	0.368

Tabela 5. Wartość P dla testów na normalność rozkładu z podziałem według stanowiska

	Wartość P	
	Test Andersona-Darlinga	Test Shapiro-Wilka
<b>Frontend Developer</b>	0.5903	0.6520
<b>Backend Developer</b>	0.6044	0.5423

Tabela 6. Wartość P dla testów na normalność rozkładu z podziałem według lat doświadczenia

	Wartość P	
	Test Andersona-Darlinga	Test Shapiro-Wilka
<b>0-4</b>	0.4097	0.4533
<b>5-9</b>	0.5825	0.7471
<b>9+</b>	0.2313	0.2746

Wartości p testów normalności są większe od 0,05. Nie ma więc podstaw do odrzucenia hipotezy  $H_0$ . Stwierdzam więc, że rozkłady zmiennych w grupach są rozkładami normalnymi.

## Badanie jednorodności wariancji

W kolejnym kroku została zbadana jednorodność wariancji za pomocą testów: Bartletta, Levene'a oraz Browna-Forsythe'a. Hipoteza zerowa tych testów mówi, że wariancje we wszystkich badanych próbach są jednakowe.

Tabela 7. Wartości P testów jednorodności wariancji

	Wartość P		
	Test Bartletta	Test Levene'a	Test Browna-Forsythe'a
<b>Experience</b>	0.0868	0.2241	0.275
<b>Position</b>	0.4468	0.3426	0.3626

Wartości p testów na jednorodność wariancji są większe od 0,05. Możemy więc stwierdzić, że w naszych danych występuje homogeniczność wariancji.

## 4. Budowa modelu ANOVA

W poniższej tabeli został zaprezentowany model ANOVA. Wartości P dla zmiennych różnicujących: doświadczenie i pozycja są znacząco niższe od 0,05, więc odrzucamy hipotezy  $H_{A0}$  oraz  $H_{B0}$ . Oznacza to, że doświadczenie oraz pozycja są czynnikami wpływającymi na wysokość wynagrodzenia w branży IT. Natomiast w badaniu nie stwierdzono istotnego wpływu interakcji tych dwóch czynników, świadczy o tym P-value na poziomie 0.51545.

Tabela 8. Model ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Experience	2	1.903e+10	9.514e+09	95.299	< 2e-16	***
Position	1	1.081e+09	1.081e+09	10.824	0.00133	**
Experience:Position	2	1.331e+08	6.654e+07	0.667	0.51545	
Residuals	114	1.138e+10	9.983e+07			
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

## 5.Efekty eksperymentalne

W tabelach 9., 10. zostały zaprezentowane wyniki badania efektów eksperymentalnych

Tabela 9. Wartości cząstkowego eta kwadrat wraz z przedziałami ufności

Parameter	Eta2 (partial)	90% CI
Experience	0.63	[0.54, 0.69]
Position	0.09	[0.02, 0.18]
Experience:Position	0.01	[0.00, 0.05]

Tabela 10. Wartości Omega kwadrat wraz z przedziałami ufności

Parameter	Omega2 (partial)	90% CI
Experience	0.61	[0.52, 0.68]
Position	0.08	[0.02, 0.16]
Experience:Position	-5.59e-03	[0.00, 0.00]

Na podstawie wartości  $\omega^2$  można wnioskować, że doświadczenie ma silny wpływ na wynagrodzenie, natomiast pozycja ma średni wpływ.

## 6.Analiza post-hoc

W tabelach 11, 12, 13 zostały przedstawione wyniki testu Tukey'a. Na ich podstawie można wnioskować, że im większe jest doświadczenie programisty tym istotnie większe są jego zarobki. Ponadto programiści backendowi zarabiają istotnie więcej od developerów frontendu.

Tabela 11. Wyniki testu Tukey'a względem lat doświadczenia

\$Experience				
	diff	lwr	upr	p adj
5-9-0-4	18041.25	12735.748	23346.75	0e+00
9+-0-4	30686.50	25380.998	35992.00	0e+00
9+-5-9	12645.25	7339.748	17950.75	3e-07

Tabela 12. Wyniki testu Tukey'a względem pozycji

\$Position				
	diff	lwr	upr	p adj
Frontend Developer-Backend Developer	-6001.5	-9615.2	-2387.8	0.0013336

Tabela 13. Wyniki testu Tukey'a względem pozycji oraz doświadczenia

\$`Experience:Position`				
	diff	lwr	upr	p adj
5-9:Backend Developer-0-4:Backend Developer	20525.0	11366.07	29683.9299	0.0000000
9+:Backend Developer-0-4:Backend Developer	31325.0	22166.07	40483.9299	0.0000000
0-4:Frontend Developer-0-4:Backend Developer	-3920.0	-13078.93	5238.9299	0.8158737
5-9:Frontend Developer-0-4:Backend Developer	11637.5	2478.57	20796.4299	0.0046597
9+:Frontend Developer-0-4:Backend Developer	26128.0	16969.07	35286.9299	0.0000000
9+:Backend Developer-5-9:Backend Developer	10800.0	1641.07	19958.9299	0.0110410
0-4:Frontend Developer-5-9:Backend Developer	-24445.0	-33603.93	-15286.0701	0.0000000
5-9:Frontend Developer-5-9:Backend Developer	-8887.5	-18046.43	271.4299	0.0625955
9+:Frontend Developer-5-9:Backend Developer	5603.0	-3555.93	14761.9299	0.4869818
0-4:Frontend Developer-9+:Backend Developer	-35245.0	-44403.93	-26086.0701	0.0000000
5-9:Frontend Developer-9+:Backend Developer	-19687.5	-28846.43	-10528.5701	0.0000001
9+:Frontend Developer-9+:Backend Developer	-5197.0	-14355.93	3961.9299	0.5708496
5-9:Frontend Developer-0-4:Frontend Developer	15557.5	6398.57	24716.4299	0.0000420
9+:Frontend Developer-0-4:Frontend Developer	30048.0	20889.07	39206.9299	0.0000000
9+:Frontend Developer-5-9:Frontend Developer	14490.5	5331.57	23649.4299	0.0001667