

Salary Recommendation of Data Scientist in 2025

analyzed by

Lusitania Ragil Cahyaningsih

November 23, 2024

Business Context and Problem Background

1 **Topik Permasalahan:** **pentingnya peran** seorang Data Scientist*, sehingga perlu **"treatment"** dalam memberlakukan standarisasi gaji harapannya **tidak ada angka turnover** karena permasalahan gaji.

2 **Problem Background :** profesi Data Scientist semakin **berkembang pesat** dan menjadi salah satu pekerjaan yang paling dicari di berbagai industri. Menganalisis data gaji seorang data scientist bukan hanya penting bagi individu yang bekerja di bidang ini, tetapi juga bagi perusahaan dan lembaga yang ingin **menarik dan mempertahankan talenta terbaiknya**.



Objective and Key Questions

1

Objective: SMART framework

S = **standarisasi gaji** karyawan Data Scientist dan mempersiapkan perusahaan dalam persaingan pasar tenaga kerja

M = tidak ada karyawan Data Scientist yang resign karena permasalahan gaji

A = analisis dan mengaplikasikan standar gaji karyawan Data Scientist

R = melakukan standarisasi gaji karyawan Data Scientist maka perusahaan mampu mempertahankan maupun mendapatkan SDM terbaik

T = tahun 2025

Sasarannya adalah untuk **standarisasi** gaji karyawan, dalam hal ini Data Scientist dan mempersiapkan perusahaan dalam persaingan pasar tenaga kerja pada **awal tahun 2025** dengan cara **menganalisa data-data** gaji karyawan pada berbagai perusahaan diberbagai tingkat keahlian dan pengalaman serta menerapkannya.



Objective and Key Questions

2

Penjabaran masalah dengan metode **Key Questions**:

- a. Berapa **rata-rata gaji** Data Scientist, Data Engineer, dan Data Analyst?
- b. Bagaimana **trend gaji** Data Scientist, Data Engineer, dan Data Analyst dari tahun 2020 – 2023?
- c. Berapa **rekomendasi gaji** Data Scientist, Data Engineer, dan Data Analyst tahun 2025?
- d. Berapa **rata-rata gaji** Data Scientist, Data Engineer, dan Data Analyst tiap **experience** dan **expertise** level?
- e. Di negara mana **gaji tertinggi** Data Scientist, Data Engineer, dan Data Analyst?
- f. **Variabel apa saja yang mempengaruhi** tinggi - rendahnya gaji Data Scientist, Data Engineer, dan Data Analyst?



Dataset Description

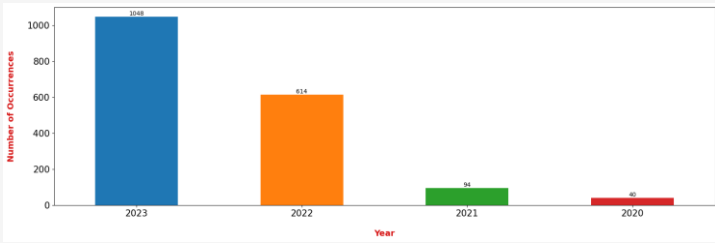
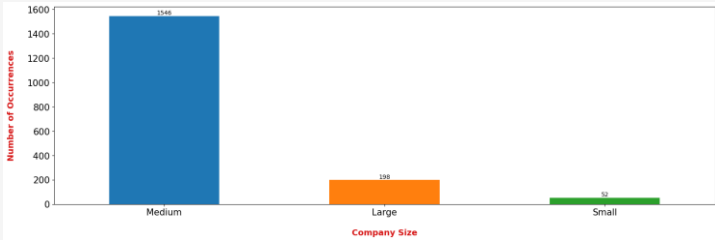
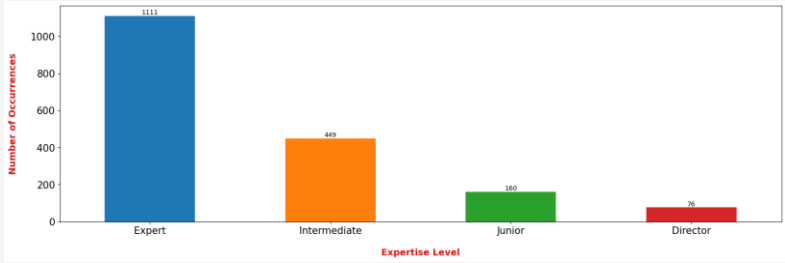
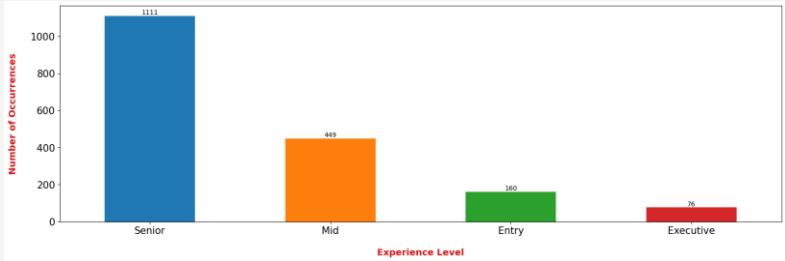
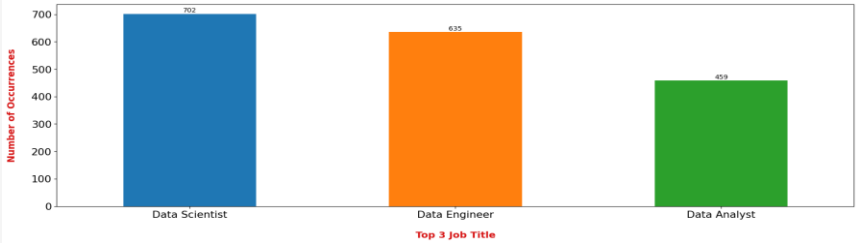
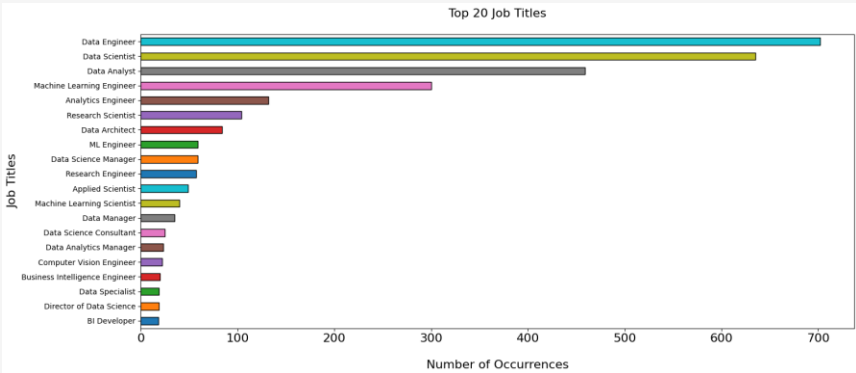
Data source: <https://www.kaggle.com/datasets/iamsouravbanerjee/data-science-salaries-2023/data>

	Job Title	Employment Type	Experience Level	Expertise Level	Company Location	Salary in USD	Employee Residence	Company Size	Year
0	Data Engineer	Full-Time	Senior	Expert	United States	210000	United States	Medium	2023
1	Data Engineer	Full-Time	Senior	Expert	United States	165000	United States	Medium	2023
2	Data Engineer	Full-Time	Senior	Expert	United States	185900	United States	Medium	2023
3	Data Engineer	Full-Time	Senior	Expert	United States	129300	United States	Medium	2023
4	Data Scientist	Full-Time	Senior	Expert	United States	140000	United States	Medium	2023
...
3288	Data Engineer	Full-Time	Mid	Intermediate	United States	130800	Spain	Medium	2020
3291	Data Scientist	Full-Time	Mid	Intermediate	Israel	119059	Singapore	Medium	2021
3293	Data Engineer	Full-Time	Mid	Intermediate	Malta	28369	Malta	Large	2021
3295	Data Scientist	Full-Time	Senior	Expert	United States	412000	United States	Large	2020
3297	Data Scientist	Full-Time	Entry	Junior	United States	105000	United States	Small	2020

1796 rows x 9 columns

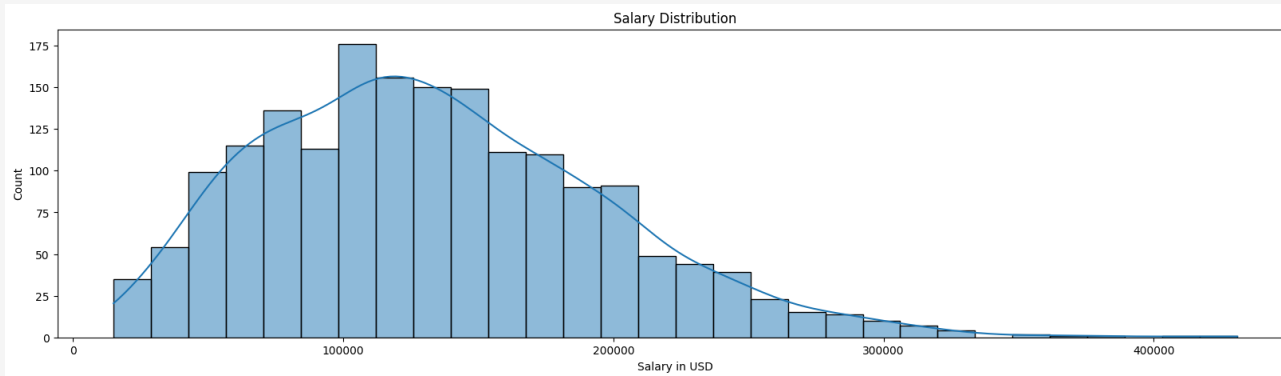
Key Variable:

- Job Title : Jabatan dalam lingkup Data Science
- Salary in USD : Gaji per tahun dalam US Dollar
- Experience Level : Tingkat pengalaman kerja
- Expertise Level : Tingkat keahlian kerja
- Company Size : Skala Perusahaan
- Year : Tahun kesesuaian data



Dataset Description

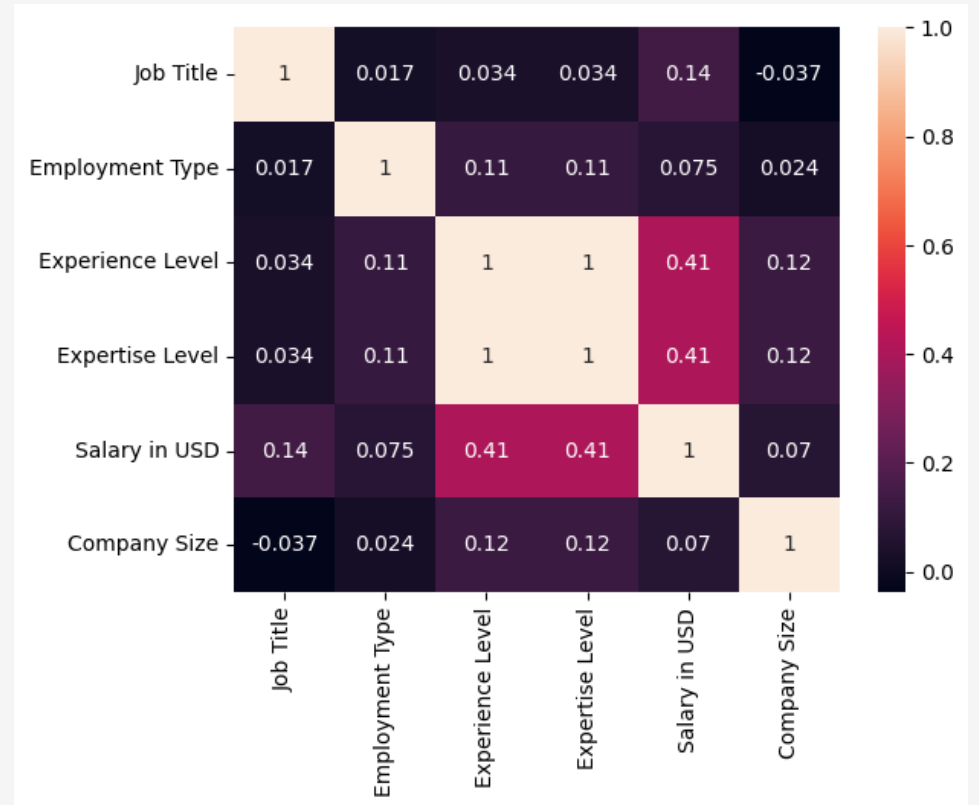
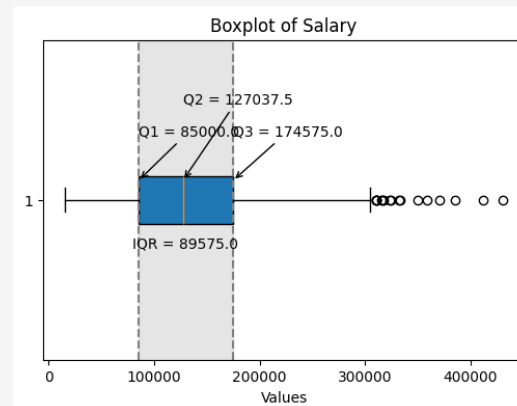
Data source: <https://www.kaggle.com/datasets/iamsouravbanerjee/data-science-salaries-2023/data>



- Mean of salary : \$ 133,667.13 /year
- Median of salary : \$ 127,037.50 /year
- Modus of salary : \$ 100,000.00 /year
- Skewness : 0.6266 -> **Distribution is moderately skewed (positive skewed)**
- Kurtosis : 0.4380 -> **Leptokurtic distribution**

Variasi : 4,011,776,939.22
Simpangan baku : 63,338.589

Variasi data gaji sangat beragam dan berdasarkan simpangan baku diketahui bahwa variasinya sendiri cukup menyebar.



Kecenderungan variabel-variabel pada data **tidak memiliki korelasi dengan gaji**. Namun jika diperingkat, terdapat 2 variabel yang hampir memiliki korelasi positif dengan gaji, yaitu **variabel experience level** dan **expertise level** sedikit **mempengaruhi gaji** namun tidak signifikan.

Methodology

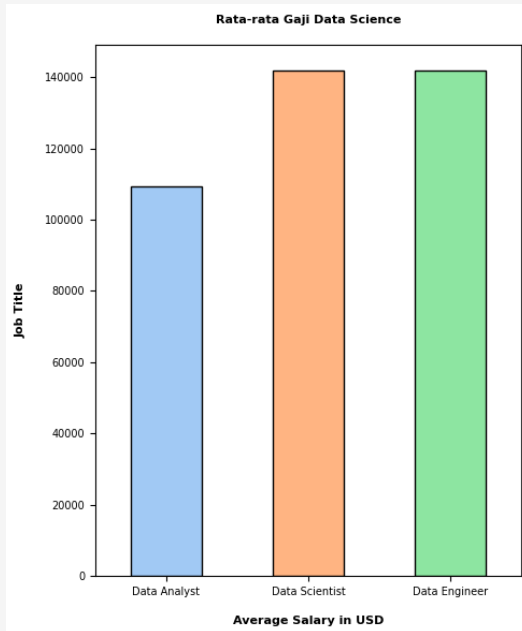
Pada analisa ini digunakan beberapa methodology untuk menjawab key questions, diantaranya:

- 1 Membandingkan data menggunakan plot **bar chart**
- 2 Membuat prediksi gaji tahun 2025 menggunakan metode **linear regression**
- 3 Membuat hipotesa dan menganalisa dengan metode **Chi-Squared**

Hint:

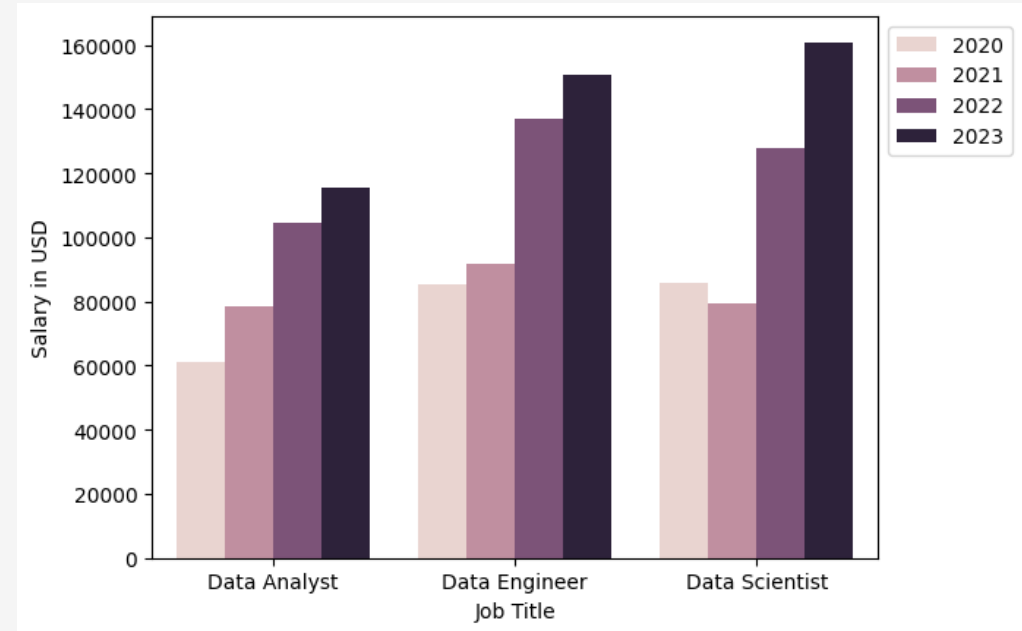
- (1) Metode **Linear Regression** adalah teknik analisis data yang digunakan untuk **memprediksi** nilai variabel yang tidak diketahui dengan menggunakan nilai data yang sudah diketahui. Metode ini memodelkan variabel yang tidak diketahui dengan variabel yang sudah diketahui sebagai **persamaan linier**.
- (2) Metode **Chi-Squared** adalah uji hipotesis dua sampel digunakan untuk **menguji hubungan antara dua variabel kategori**.

Analysis Result and Recommendations



1 Berapa **rata-rata gaji** Data Scientist, Data Engineer, dan Data Analyst?

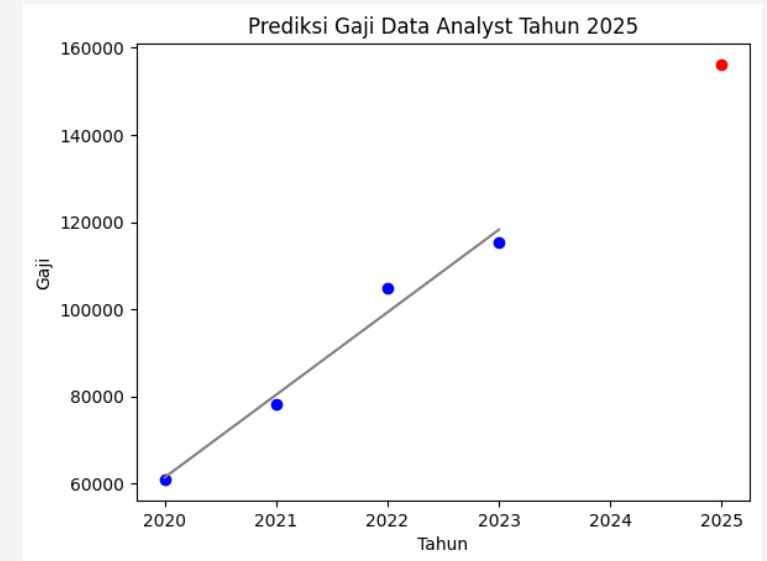
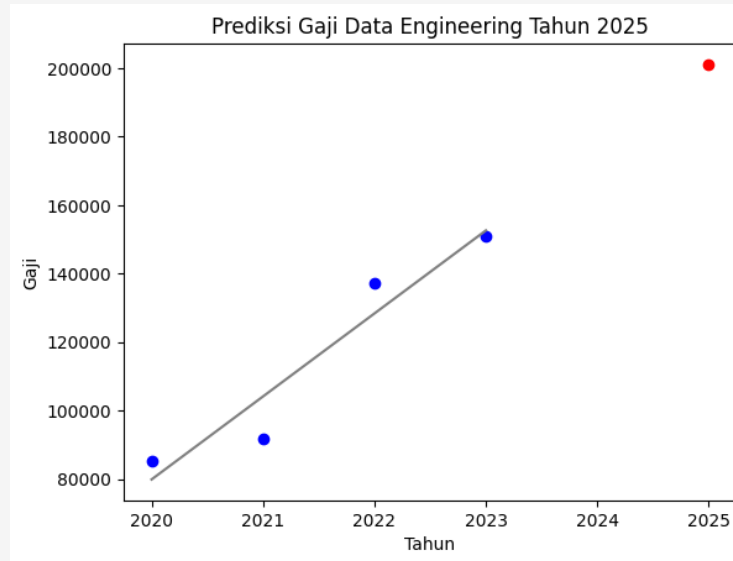
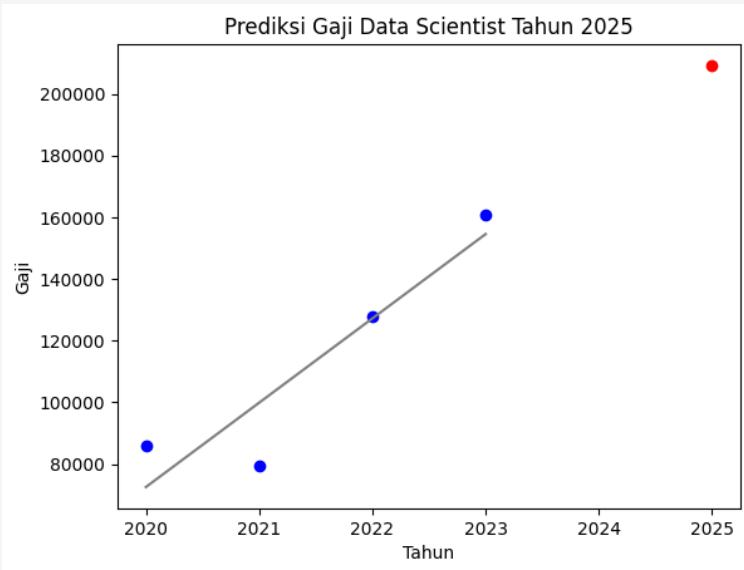
Data Scientist : \$ 141.886.16 /year
Data Engineer : \$ 142.033.87 /year
Data Analyst : \$ 109.500.38 /year



2 Bagaimana **trend gaji** Data Scientist, Data Engineer, dan Data Analyst dari tahun 2020 – 2023?

Ketiga job title tersebut cenderung mengalami **peningkatan dari tahun ke tahun**, walaupun pada tahun 2021 rata-rata gaji Data Scientist sempat mengalami penurunan, namun di tahun 2022 mengalami peningkatan kembali dengan nilai yang lebih tinggi dari tahun 2020.

Analysis Result and Recommendations



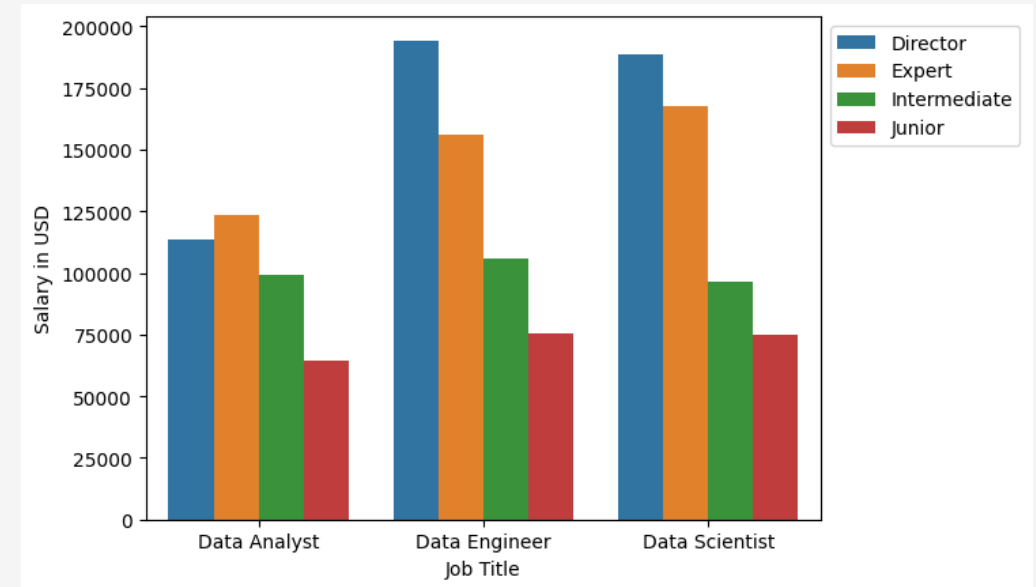
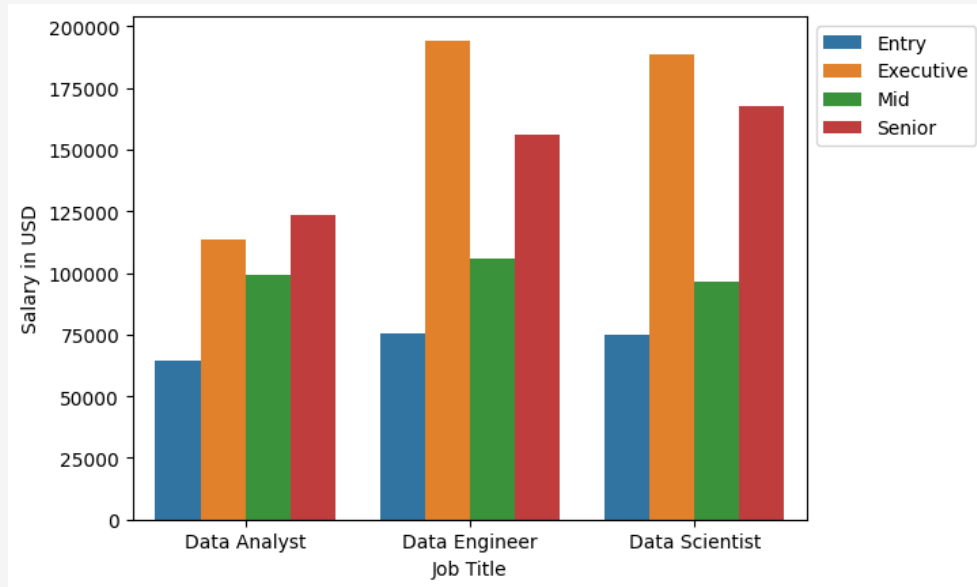
3

Berapa **rekomendasi gaji** Data Scientist, Data Engineer, dan Data Analyst tahun 2025?

Berdasarkan perhitungan prediksi dengan menggunakan metode *linear regression*, direkomendasikan gaji 2025 sebagai berikut:

- Data Scientist : \$ 209,204.33 /year
- Data Engineer : \$ 201,098.96 /year
- Data Analyst : \$ 156,177.84 /year

Analysis Result and Recommendations



4

Berapa **rata-rata gaji** Data Scientist, Data Engineer, dan Data Analyst tiap **experience** dan **expertise** level?

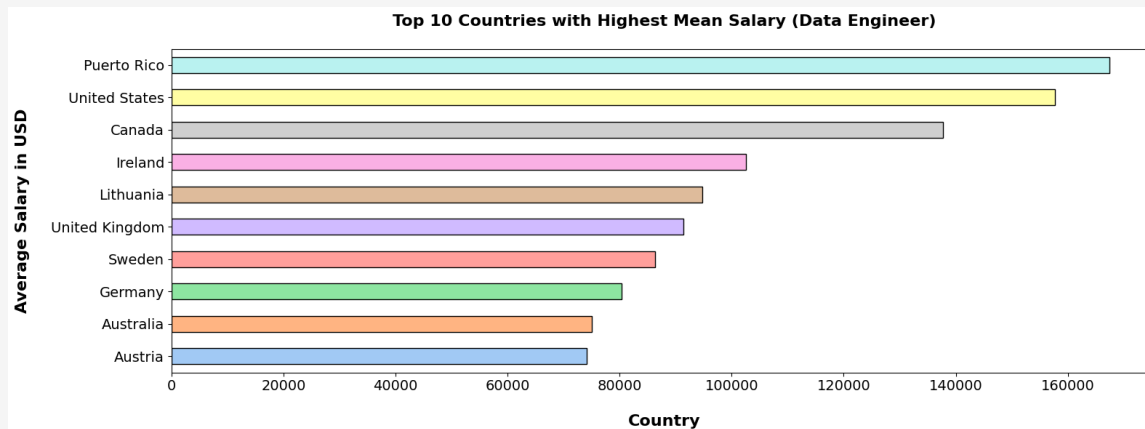
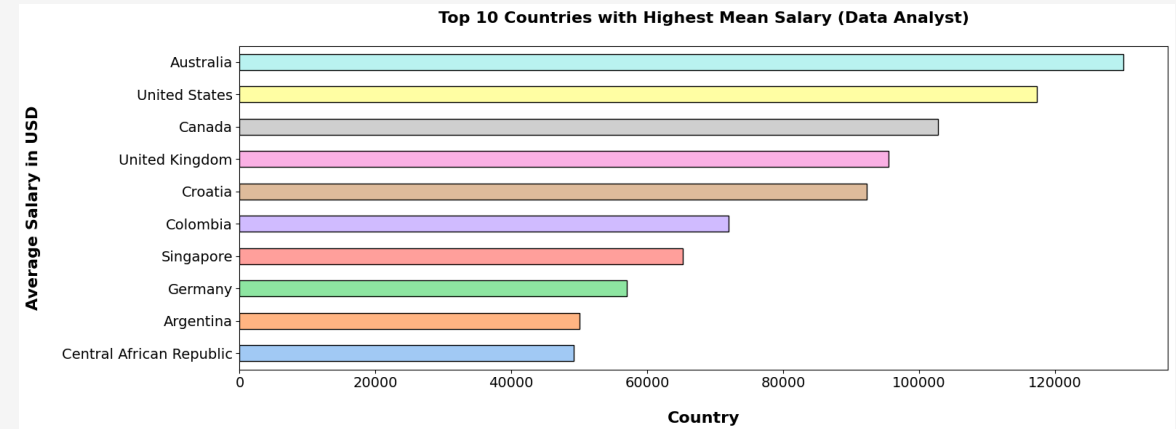
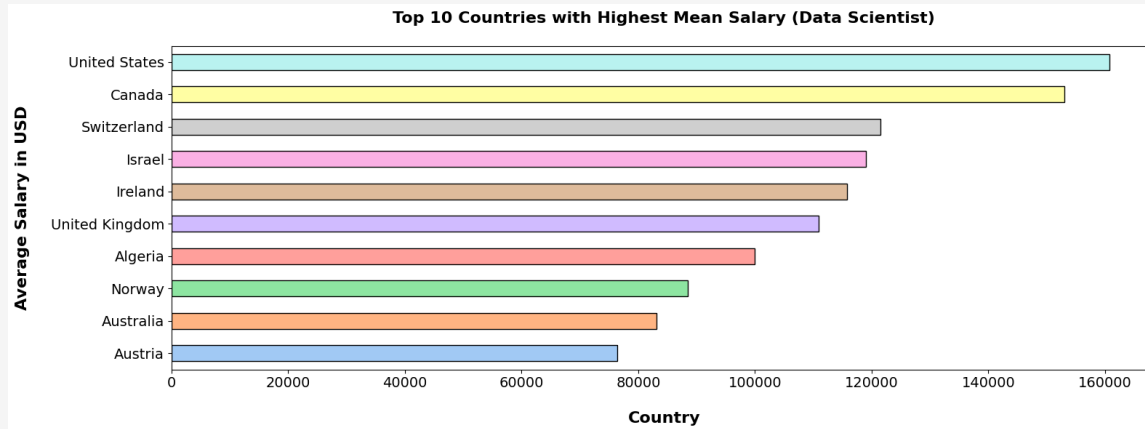
Experience Level

Berdasarkan table di atas, ketiga job title memiliki Gambaran keterikatan/hubungan antara experience level dan salary yang sama. Dimana jika diurutkan dari salary tertinggi ke terendah urutan experience levelnya adalah **Executive, Senior, Mid, dan yang terakhir adalah Entry.**

Expertise Level

Berbeda dengan experience level, pada expertise level ini Data Analyst memiliki model yang berbeda dari 2 job title lainnya. Dimana pada Data Engineer dan Data Scientist **Director menjadi expertise level yang memiliki gaji tertinggi, dilanjutkan dengan Expert, Intermediate, dan yang terakhir Junior.** Pada Data Analyst Director menjadi posisi kedua, dimana posisi pertama pada expertise level Expert.

Analysis Result and Recommendations



5

Di negara mana gaji tertinggi Data Scientist, Data Engineer, dan Data Analyst?

Berdasarkan diagram tersebut, gaji tertinggi Data Scientist berada di negara **United State**, sedangkan Data Engineer di negara **Puerto Rico**, dan Data Analyst di negara **Australia**.

Analysis Result and Recommendations

Metode Chi - Squared

Membuat Hipotesa:

H0 : variable tidak mempengaruhi salary

H1 : variable mempengaruhi salary

Digunakan confidence level 95%

```
cols = df4.drop(columns=['Company Location', 'Employee Residence','Year','Salary in USD']).columns

for col in cols:
    cross = pd.crosstab(df4[col],df4['Salary in USD'])
    pval = stats.chi2_contingency(cross).pvalue
    prompt = "H0 rejected" if pval<0.05 else "fail to reject H0"
    print(f"P-value of {col} and Salary: {pval} {prompt}")

✓ 0.0s

P-value of Job Title and Salary: 0.59817876488025 fail to reject H0
P-value of Employment Type and Salary: 1.0 fail to reject H0
P-value of Experience Level and Salary: 0.03189342325026161 H0 rejected
P-value of Expertise Level and Salary: 0.03189342325026161 H0 rejected
P-value of Company Size and Salary: 1.6808836636158922e-10 H0 rejected
```

6 Variabel apa saja yang mempengaruhi tinggi - rendahnya gaji Data Scientist, Data Engineer, dan Data Analyst?

Berdasarkan hasil perhitungan uji hipotesa dengan metode Chi – Squared, nilai p-value didapatkan bahwa variabel-variabel yang mempengaruhi gaji seorang Data Science (Baik scientist, engineer, maupun analyst) adalah **pengalaman bekerja**, **tingkat keahlian kerja**, serta **ukuran skala level perusahaan**.

Analysis Result and Recommendations

Conclusion / Recommendation:

Dari analisa di atas dapat disimpulkan bahwa:

1. Faktor **pengalaman kerja, tingkat keahlian, dan skala perusahaan mempengaruhi** nilai gaji para Data Science. Dengan jenis korelasi positif, artinya apabila nilai pengalaman, tingkat keahlian, dan skala perusahaan tinggi/besar maka nilai gaji juga akan lebih besar.
2. **United State dan Canada selalu masuk dalam jajaran Top 3** negara dengan tingkat gaji tertinggi untuk ketiga kategori Job Title yang dianalisa saat ini. Namun dalam hal ini, belum tentu negara tersebut merupakan negara favorit karena perlu analisa lebih dalam terkait *living cost* di negara tersebut.
3. Gaji tahun 2025 untuk Data Scientist direkomendasikan sebesar **209,204.33 USD**, Data Engineer **201,098.96 USD** dan Data Analyst **156,177.84 USD** per tahun. Artinya keseluruhan posisi Data Science dinilai baik dan bisa dijadikan rekomendasi job title saat ini karena nilai gaji yang meningkat dari tahun ke tahun, walaupun pada tahun 2021 sempat mengalami penurunan, besar kemungkinan adalah efek dari adanya wabah Covid pada saat itu.

Thank you!

Detail Analysis : <https://github.com/CODA-Assignment-Bay/p1-coda001-rmt-m2-lusitaniarc>

Dashboard Data Viz :