



DΦLab

# TETRIS PROGRAM

---

Data Jobs Salary Visualization in Developed Countries

Evan Kurnia Alim  
eka091294@gmail.com

#StackYourSkill





## Data Collection & Integration

Dataset Title: Data Science Job Salaries

Dataset source: <https://salaries.ai-jobs.net/download/salaries.csv> (as referenced on <https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>, perfect usability score of 10 out of 10 by Kaggle), downloaded on 26 July 2022

1. Download the dataset from <https://salaries.ai-jobs.net/download/salaries.csv>.
2. Load the data in Jupyter Notebook to display the head of the dataset as follows:

All salaries are annual in US Dollars (USD).

# TETRIS PROGRAM



In [3]:

```
import pandas as pd
import matplotlib.pyplot as plt

# read from .csv file
ds_salary_df = pd.read_csv('ds_salaries.csv')
ds_salary_df.head()
```

Out[3]:

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location
0	0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB
3	3	2020	MI	FT	Product Data Analyst	20000	USD	20000	HN	0	HN
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	US	50	US



## Data Cleaning (Step 1)

Unnamed column exists in the raw dataset, delete this column:

In [4]:

```
# delete unnamed column
ds_salary_df_1 = ds_salary_df.drop(columns=['Unnamed: 0'])
ds_salary_df_1.head()
```

Out[4]:

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
0	2020	MI	FT	Data Scientist	70000	EUR	79833	DE	0	DE	
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	JP	0	JP	
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	GB	50	GB	



## Data Cleaning (Step 2)

Check missing values. Fortunately, there are none in the dataset.

```
In [5]: # check missing values  
ds_salary_df_1.isnull().sum()
```

```
Out[5]: work_year          0  
experience_level         0  
employment_type         0  
job_title               0  
salary                 0  
salary_currency         0  
salary_in_usd          0  
employee_residence      0  
remote_ratio            0  
company_location        0  
company_size            0  
dtype: int64
```



## Hypotheses

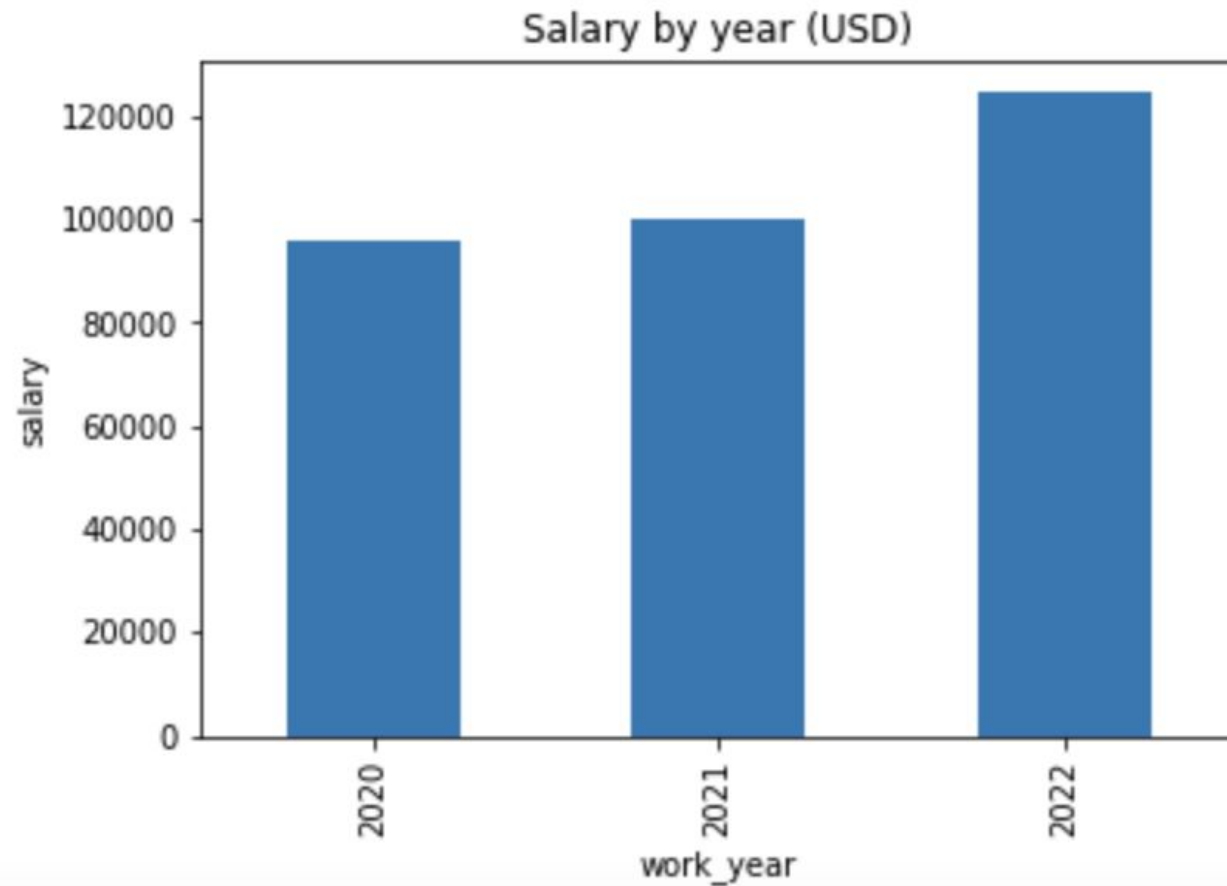
- The higher the job rank, the higher the salary.
- The larger the company, the higher the salary.
- The average salary in developed countries is much higher than the average salary in Indonesia (a developing country).
- Hybrid-style working has the lowest average salary.
- Salary distribution is right-skewed.



## Data Visualizations



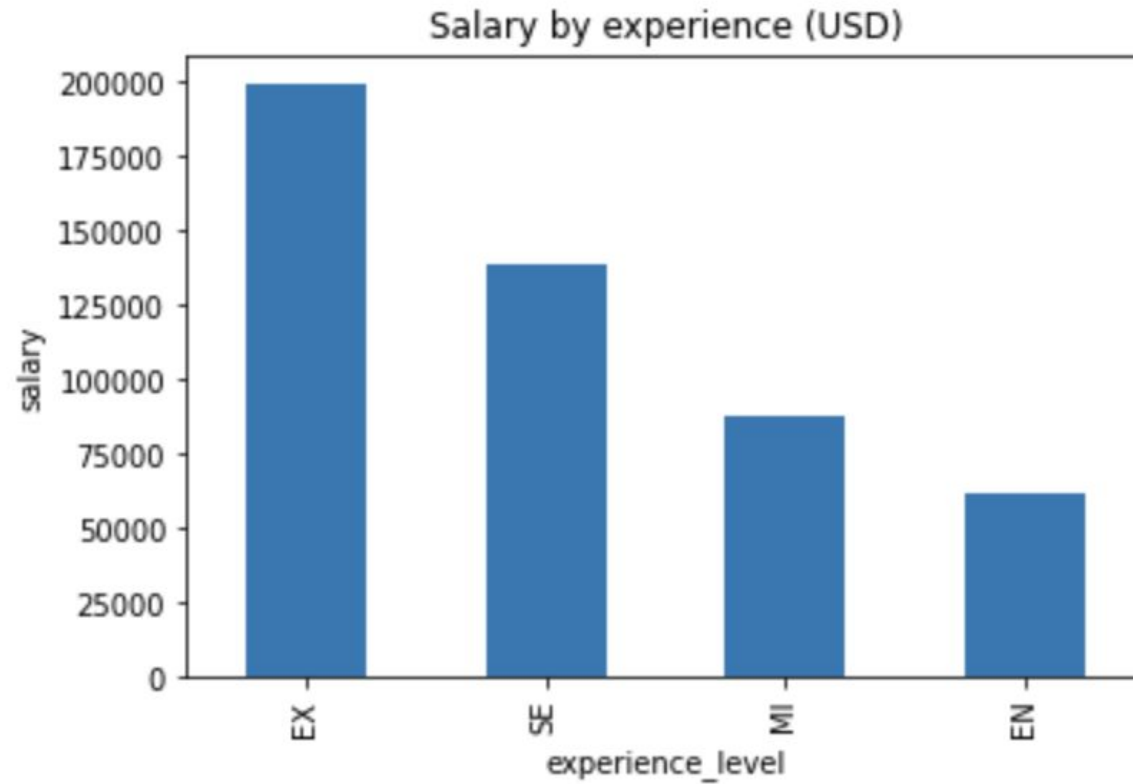
## Salary by Year





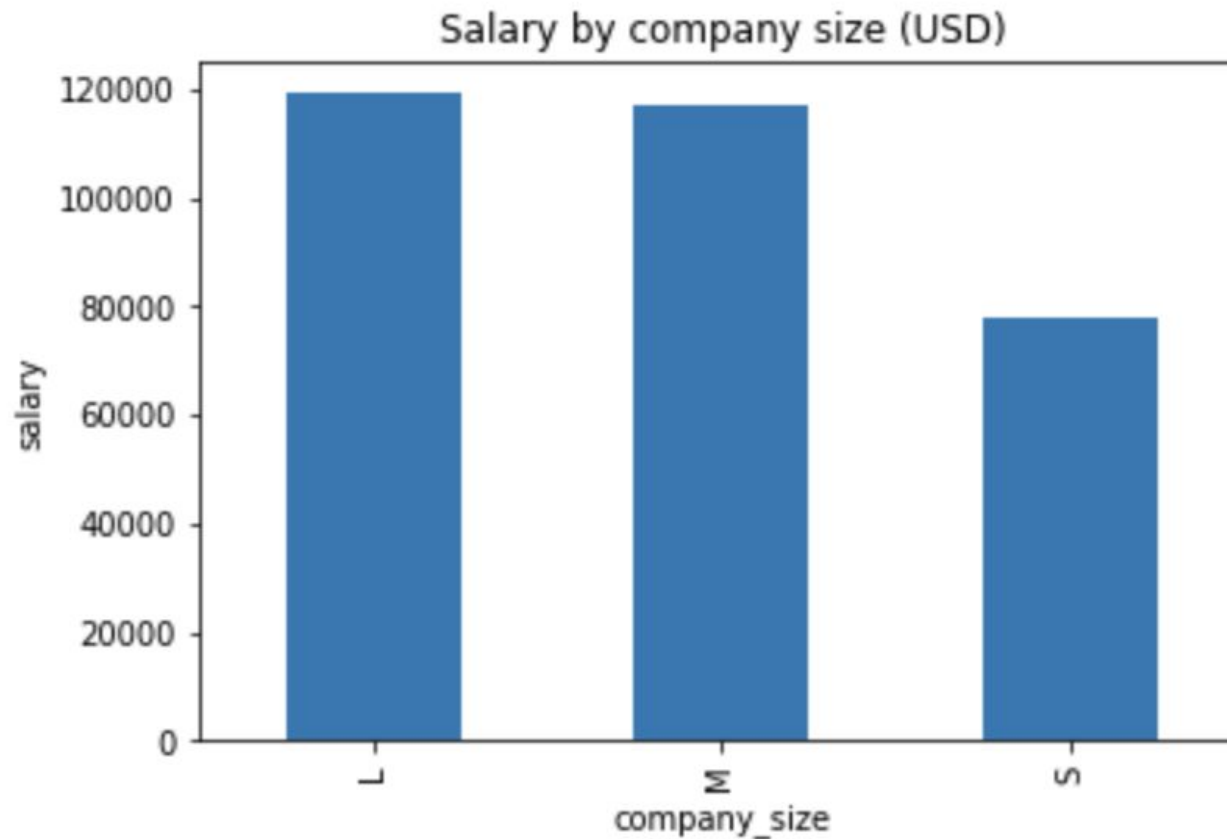


## Salary by Job Rank



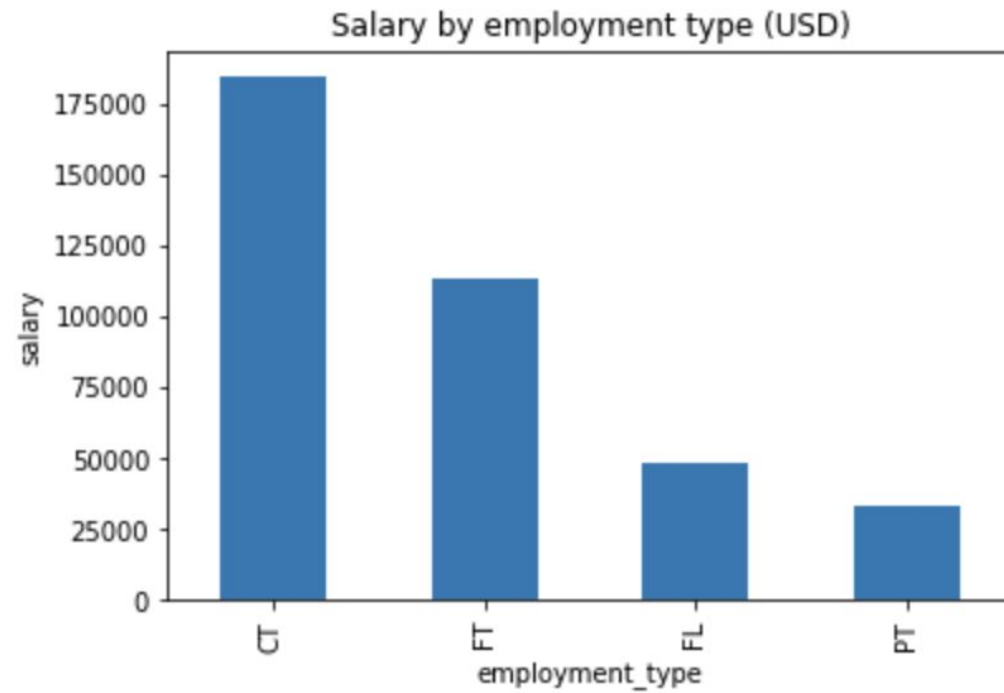


## Salary by Company Size



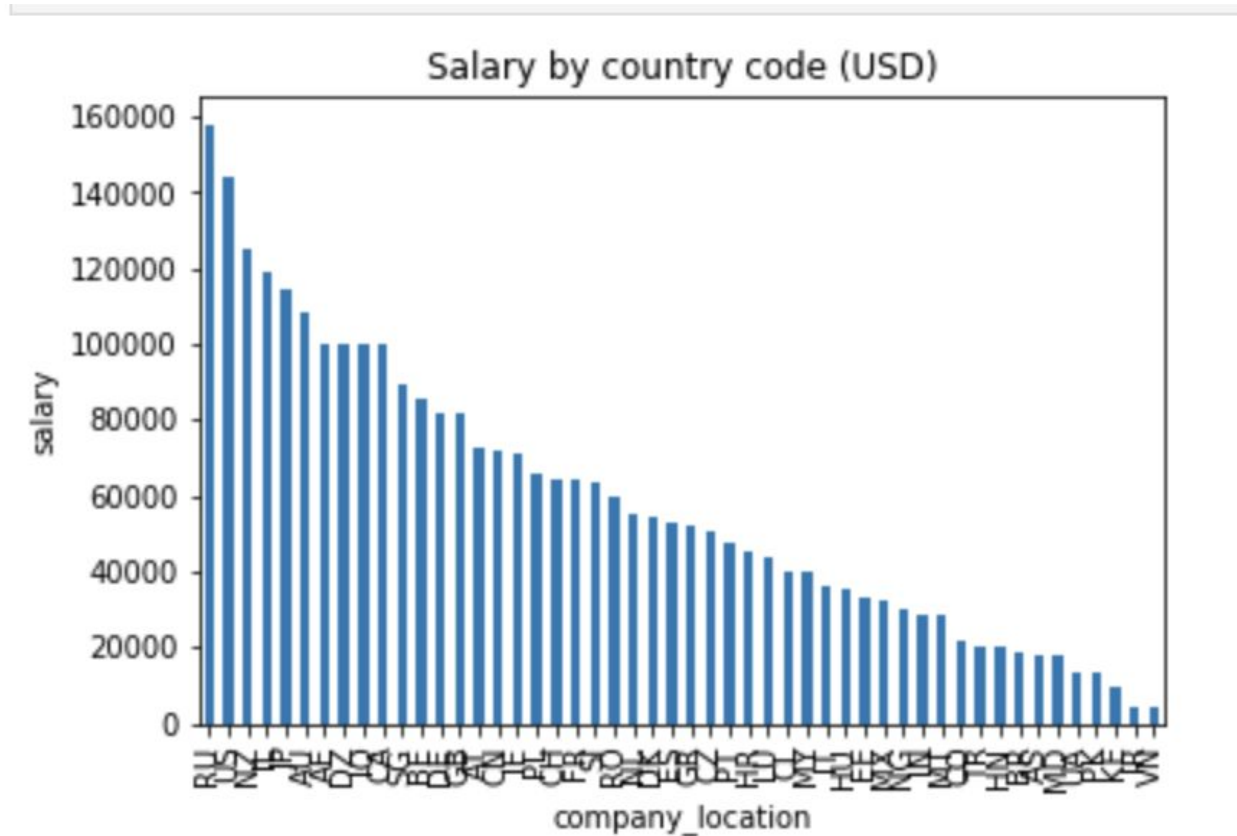


## Salary by Employment Type



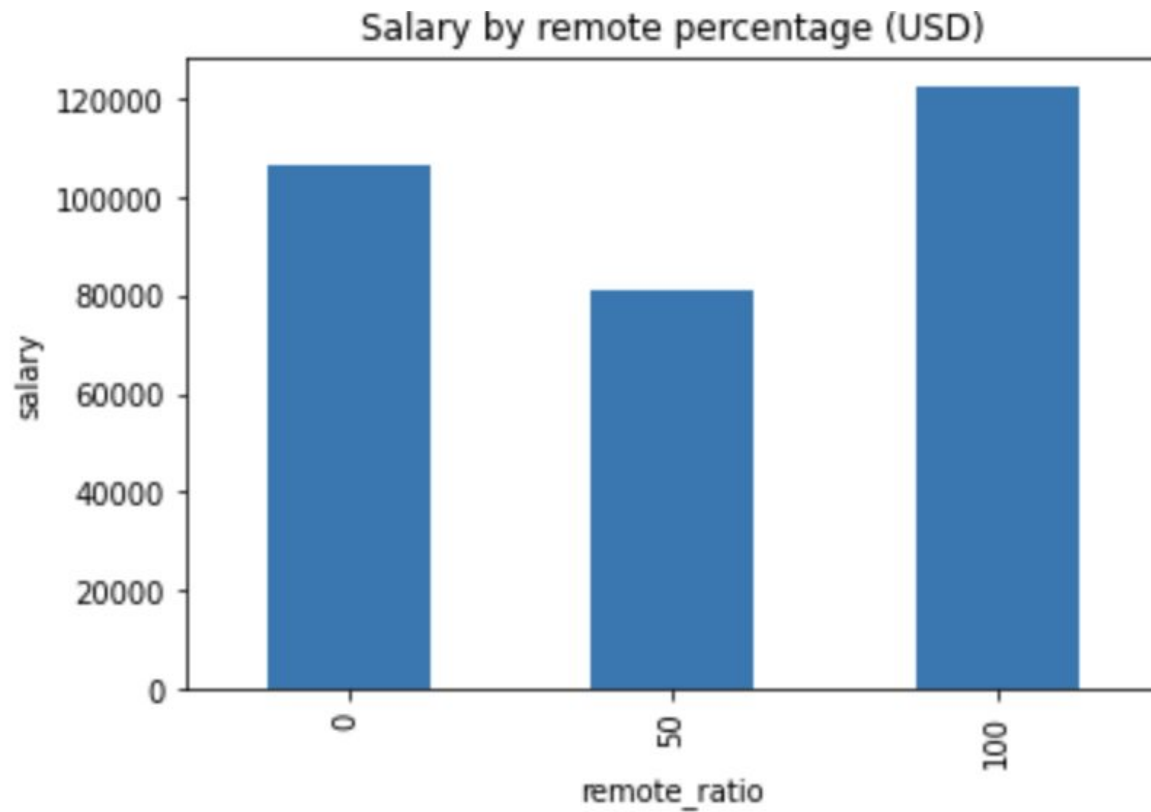


## Salary by Country





Salary by Remote Percentage (0=Full WFO, 50=Hybrid, 100=Full WFA)





## Experience level vs Remote Ratio

experience_level	remote_ratio	
EN	0	14
	50	25
	100	49
EX	0	3
	50	5
	100	18
MI	0	56
	50	42
	100	115
SE	0	54
	50	27
	100	199



## Salary Distribution





## Streamlit (data visualizations in 1 page)

See here: <https://evanka-tetris-capstone-salary-app-kmid40.streamlitapp.com/>





## Insight Analysis (Conclusion)

- Employees in developed countries typically earn more salary in average than in developing countries like Indonesia. This is justified that developed countries has much more developed economy and technology compared to developing countries, which results in much higher GDP (gross domestic product) per capita.
- Full-time (including contract) employees typically earn much higher than part-time (including flexible or project-based) employees.
- Hybrid employees typically earn less than full WFA and full WFO. Full WFA employees typically earn higher than even full WFO, this is justified by full WFA roles are typically higher in rank.
- Salary distribution is right-skewed, i.e. there are very high earning employees in extremely rare cases.

DΦLab

# **AYO #STACKYOURSKILL SEKARANG**

**dan Persiapkan Diri Menjadi Praktisi Data!**

---

