

Data Science Job Salaries Analysis

Cindy Wang, Zitao Zhang, Aile Ma, Yi Luo

Motivation



- Analysis of salary for data science related jobs
- Having access to salary information allows individuals to benchmark their expectations against industry standards, ensuring that they are appropriately compensated for their expertise and experience.
- Our target group is people who are seeking for a new job, fresh undergraduates right out of college, and employers who want to understand the current economy
- Having the knowledge of salary ranges helps in negotiating fair compensation and making informed career decisions. Also help employers to fairly pay their workers

Data Overview

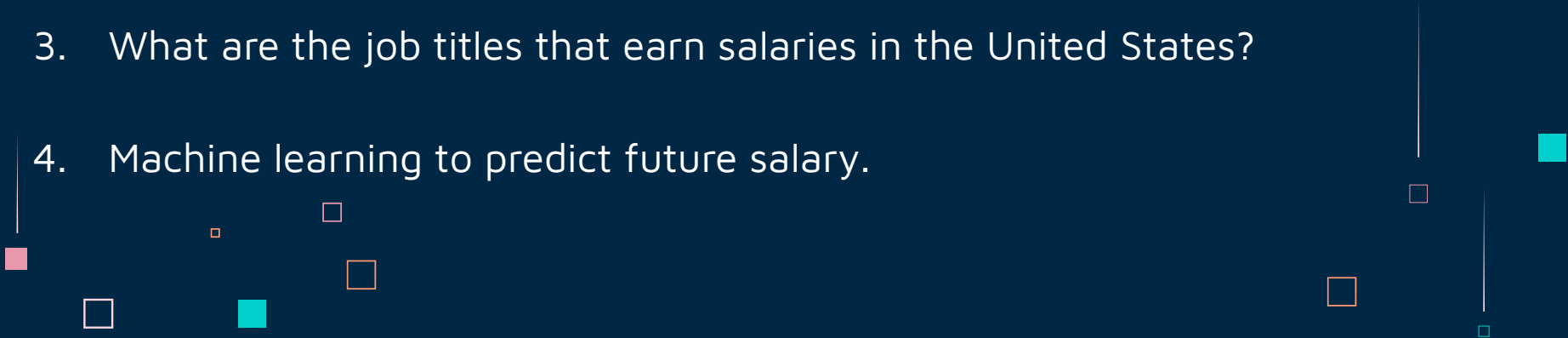


- Our first dataset is Data Jobs Salaries from Kaggle, and the data is collected from the website <https://ai-jobs.net/salaries/download/>, its data gathered from professionals and employers all over the world. In this dataset we have work year, job title, salary in usd, company location, experience levels and other variables.
- The second dataset is Employment Rate from OECD website, in this dataset, we have variables such as: country name and employment rate
- We used country name and year to merge these two datasets into one.
- There are no data quality issues, and they are both public data sets accessible by everyone

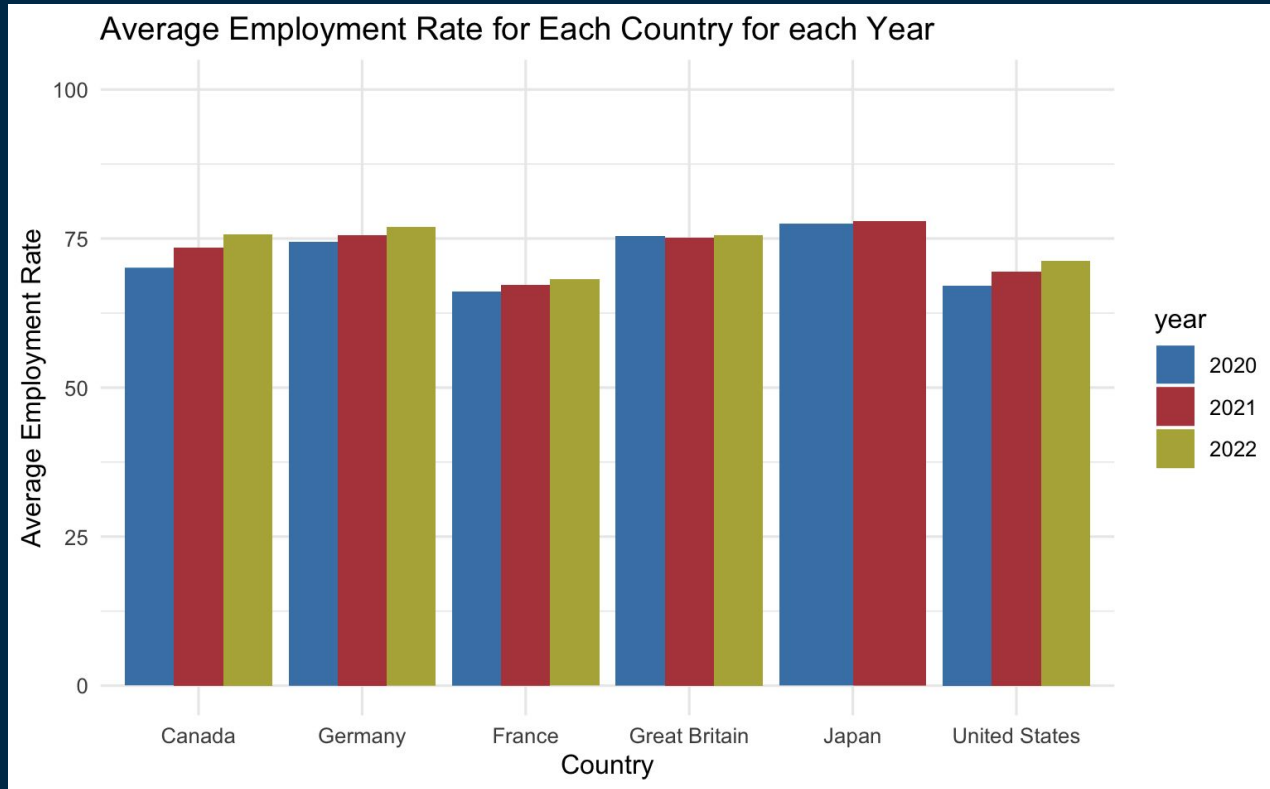
Questions



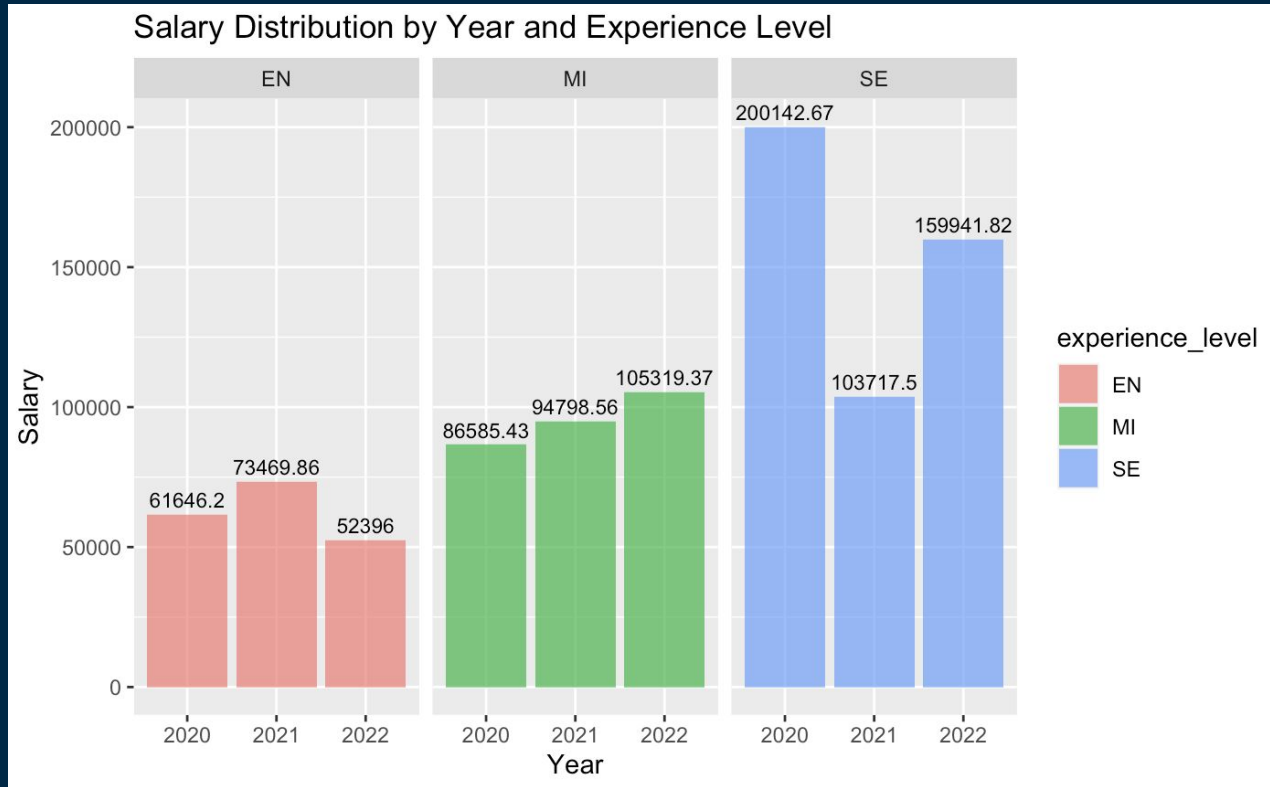
1. What is the average employment rate in different countries in different years?
2. What is the average salary for data science jobs with different experience levels?
3. What are the job titles that earn salaries in the United States?
4. Machine learning to predict future salary.



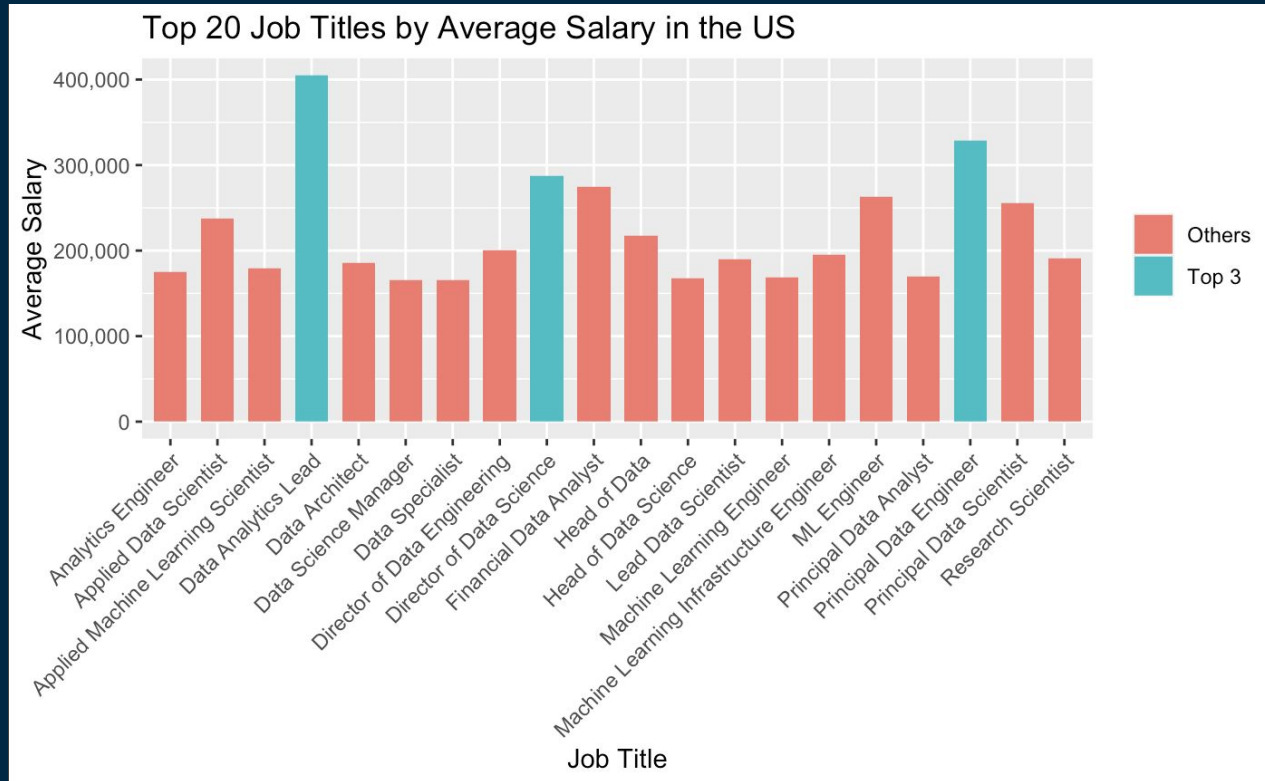
Average employment rate in different countries in different years



Average salary for data science jobs with different experience levels



Job titles and their salaries in the United States



Machine Learning

1. Select "US" for investigating target
2. Split training data and test data
3. Utilize random forest model to predict the average salary

Rank #1: Applied Data Scientist

Rank #20: Computer Vision Software Engineer

Root Mean Square Error
Mean Absolute Error
R-Squared
Mean Absolute Percentage Error

```
[1] "RMSE: 76964.1492345838"  
[1] "MAE: 48176.7444040039"  
[1] "R2: 0.170804446094393"  
[1] "MAPE: 0.39371009504221"
```

A tibble: 20 × 2

job_title <fctr>	average_predicted_salary <dbl>
Applied Data Scientist	226805.7
Financial Data Analyst	203950.2
Applied Machine Learning Scientist	195538.3
Principal Data Scientist	193749.0
Principal Data Engineer	179373.7
Analytics Engineer	173334.4
Machine Learning Engineer	167540.8
Head of Data Science	165598.0
Staff Data Scientist	156323.5
Data Science Manager	150047.3

1–10 of 20 rows

A tibble: 20 × 2

job_title <fctr>	average_predicted_salary <dbl>
Lead Data Engineer	145259.5
Machine Learning Scientist	145108.6
Data Analytics Manager	142447.2
Lead Data Analyst	138472.3
Data Scientist	137917.3
Data Engineer	136740.1
Data Analyst	131013.2
NLP Engineer	130008.7
AI Scientist	115569.5
Computer Vision Software Engineer	110653.6

11–20 of 20 rows

Privacy Concerns:

The salary data that we are using is collected anonymously, but there might still be concerns about individual privacy. Data about specific job titles, locations, and salaries could potentially lead to the identification of individuals, especially in smaller companies or specific roles.



Conclusion

In the United States, data science proves to be a high-paying career, with salaries steadily climbing annually across diverse experience levels. The field's continuous growth and demand contribute to the increasing compensation, making it an attractive and financially rewarding profession for those involved.





THANKS