

DATA SCIENTISTS SALARY PREDICTION IN USA

George Edgar

GOAL

- Gain clarity on pursuing Data Science program through Glassdoor data analysis.
- Predict salary change over long period.
- Uncover typical expectations of data science employees by companies.

TASK

- Data Collection
- Data Cleaning
- Exploratory Data Analysis
- Model Building
- Deploy Model

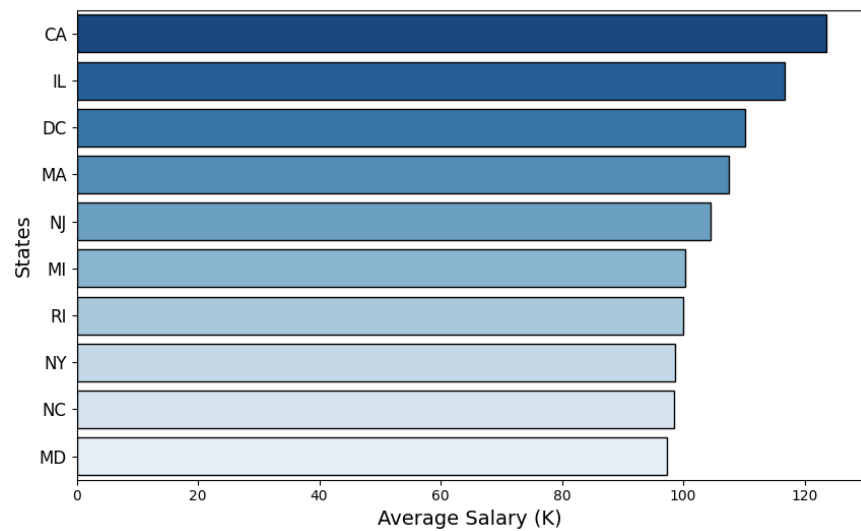
DATA COLLECTION & CLEANING

- Kaggle
- Salary parsing
- Identify state
- Job description parsing

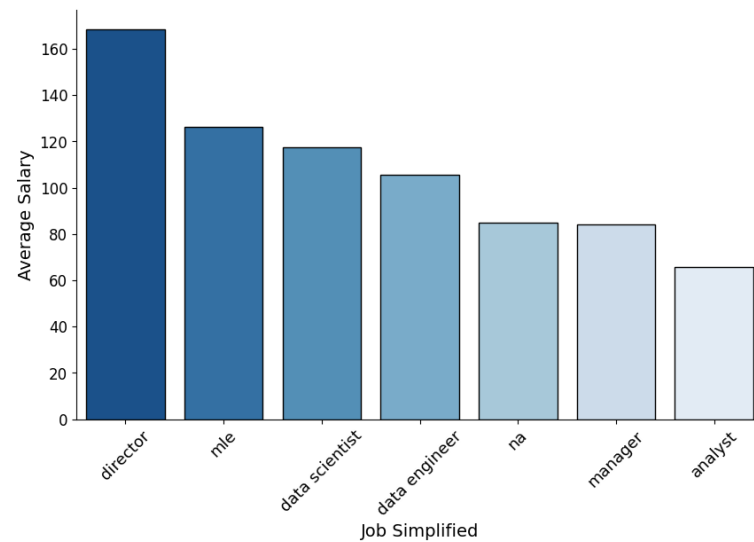


EXPLORATORY DATA ANALYSIS

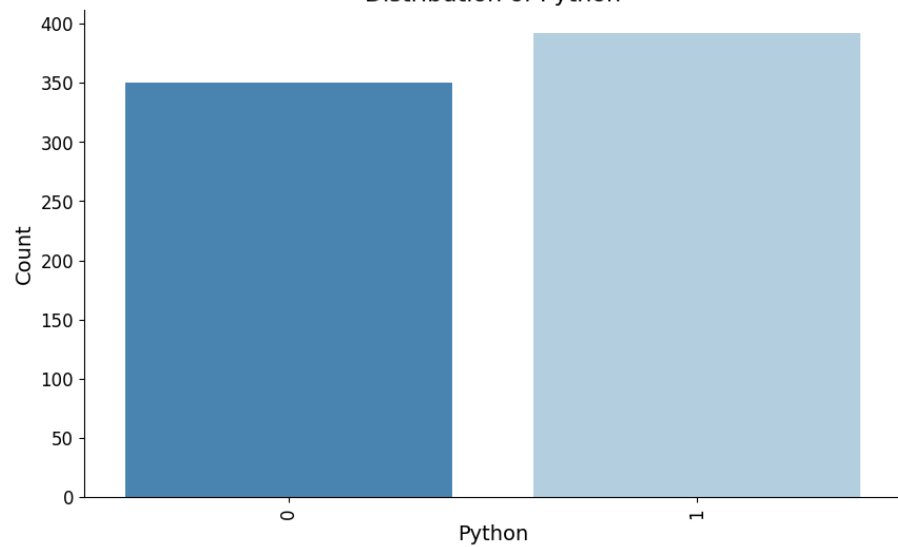
Average Salary for Different States



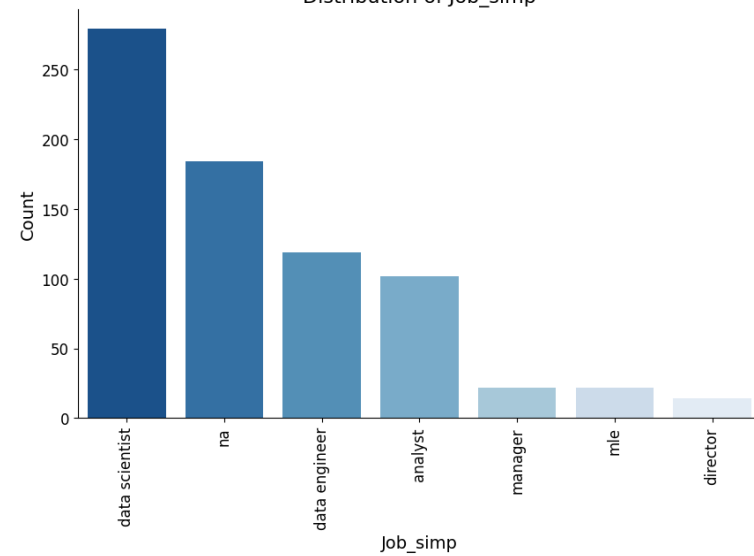
Average Salary by Job Simplified

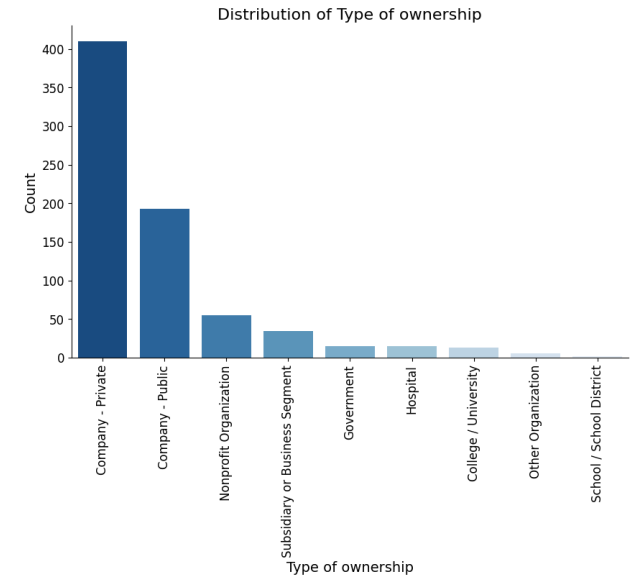
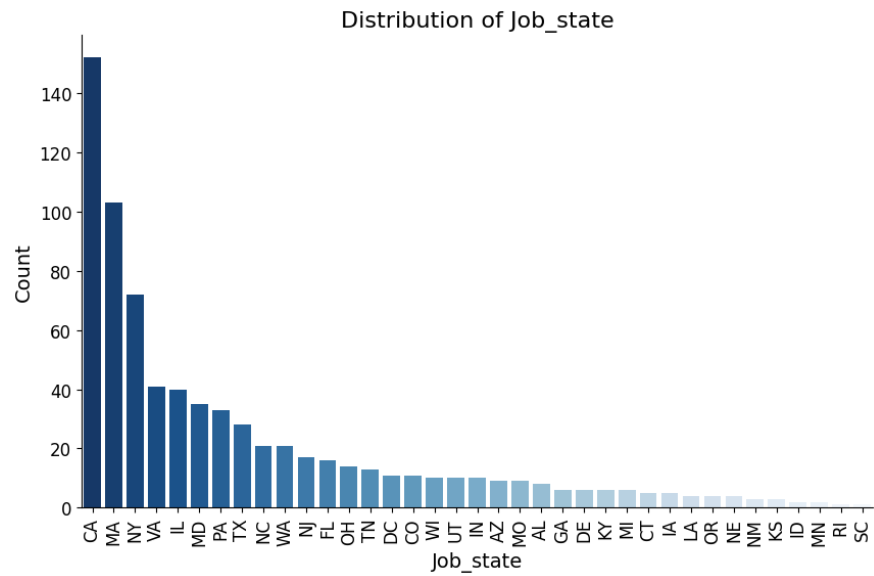
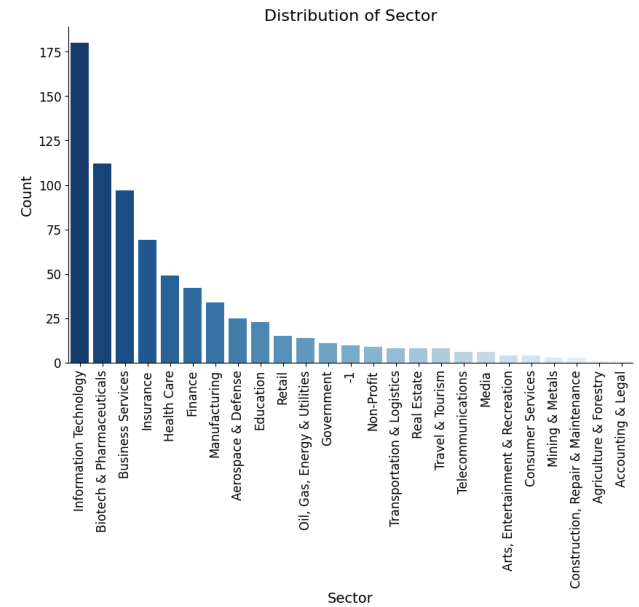
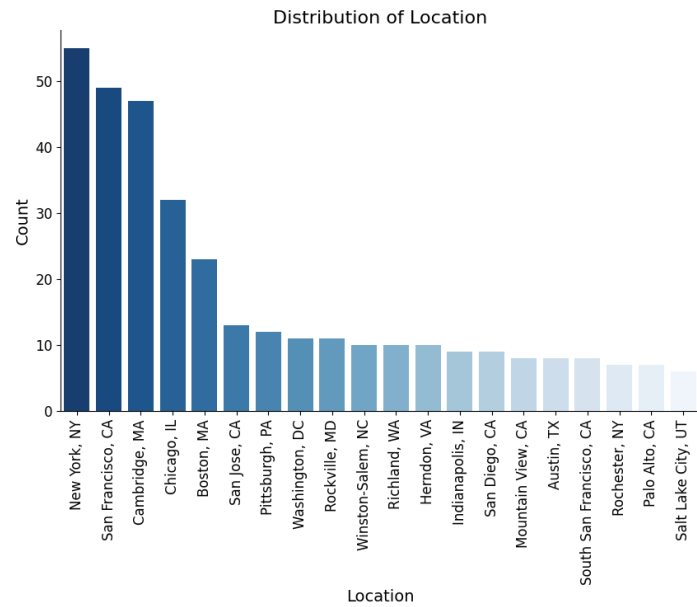


Distribution of Python



Distribution of Job_simp

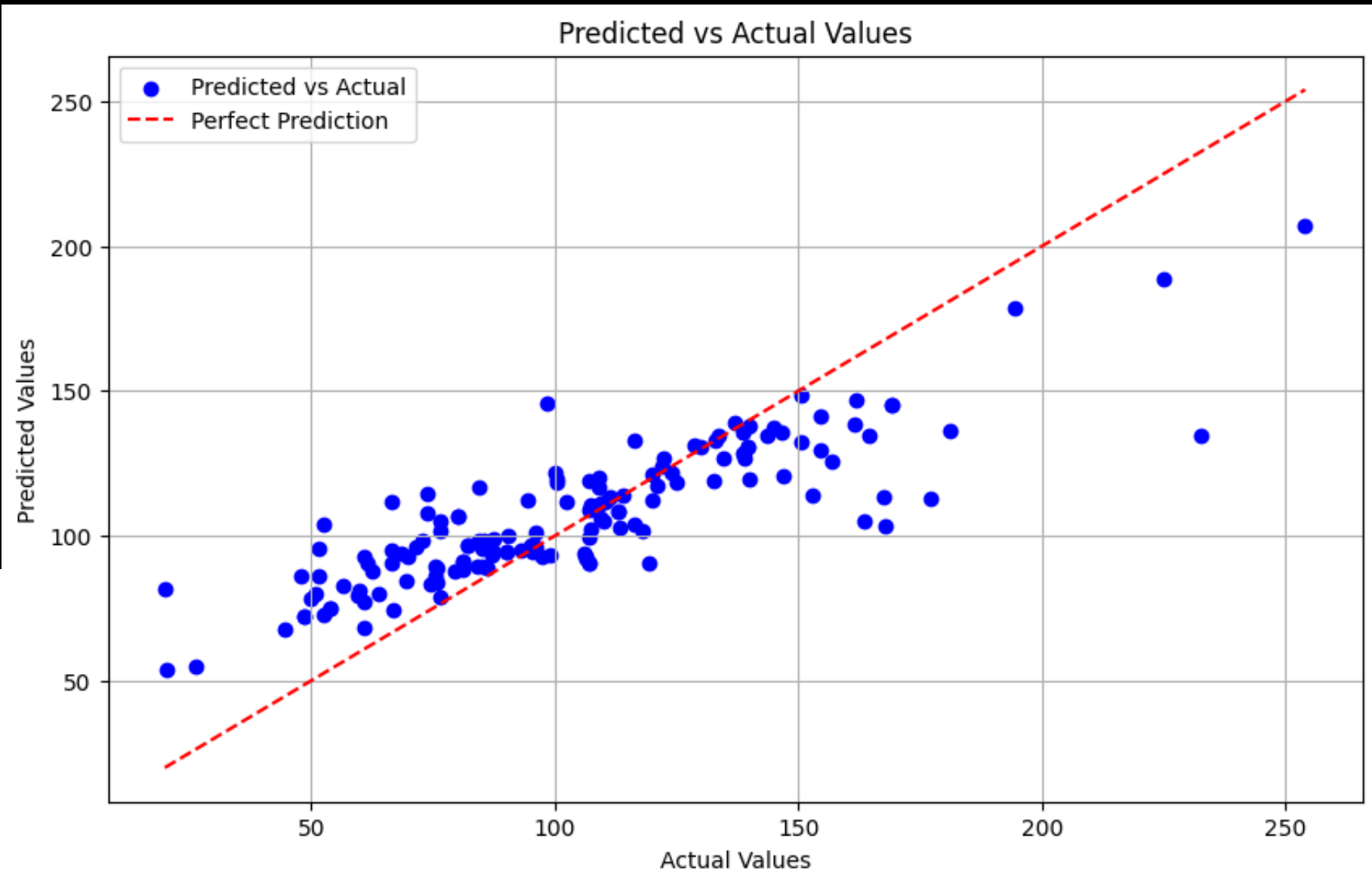




MODEL BUILDING

- Chose columns
- Dummy data
- Train Test split – 70% train, 30% test
- Multiple linear regression, Lasso regression, Random forest
- Tune model using GridSearchCV
- Test model

5-YEAR DATA SCIENCE SALARY PREDICTION



CONCLUSION

- Good accuracy of prediction
- No Glassdoor API
- Compare difference in salary over time
- Predict change in sector listings