Data Analyst Job Salary Prediction

Abby Liu

Brutal job market due to COVID-19 pandemic

1. The US unemployment rate reached the highest point in the past 10 years…

Questions From Two Perspectives:

- For job seekers: What limits or expectations can job applicants have about salary when looking for Data Analyst jobs?

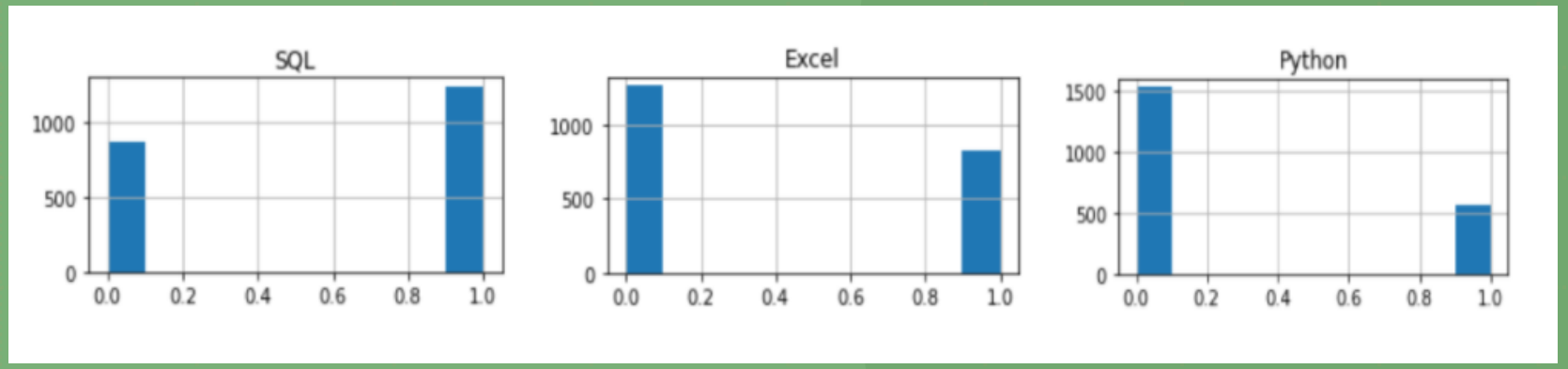
- For employers: How to set a competitive and reasonable salary to attract and retain talent?

Job Description	Rating	Company Name	Location	Size	Type of ownership	Industry	Sector	Revenue	Years Founded	Min_Salary	Max_Salary
Are you eager to roll up your sleeves and harn...	3.2	Vera Institute of Justice	NY	201 to 500 employees	Nonprofit Organization	Social Assistance	Non-Profit	100to500 million (USD)	59.0	37	66
Overview Provides analytical and technical su...	3.8	Visiting Nurse Service of New York	NY	10000+ employees	Nonprofit Organization	Health Care Services & Hospitals	Health Care	2to5 billion (USD)	127.0	37	66
We're looking for a Senior Data Analyst who ha...	3.4	Squarespace	NY	1001 to 5000 employees	Company - Private	Internet	Information Technology	<NA>	17.0	37	66
Requisition NumberRR-0001939 Remote:Yes We col...	4.1	Celerity	NY	201 to 500 employees	Subsidiary or Bu... Se...	IT ...	Information Technology	50to100 million (USD)	18.0	37	66
ABOUT FANDUEL GROUP FanDuel Group is a world-...	3.9	FanDuel	NY	501 to 1000 employees	Company - Private	Sports & Recreation	Arts, Entertainment & Recreation	100to500 million (USD)	11.0	37	66

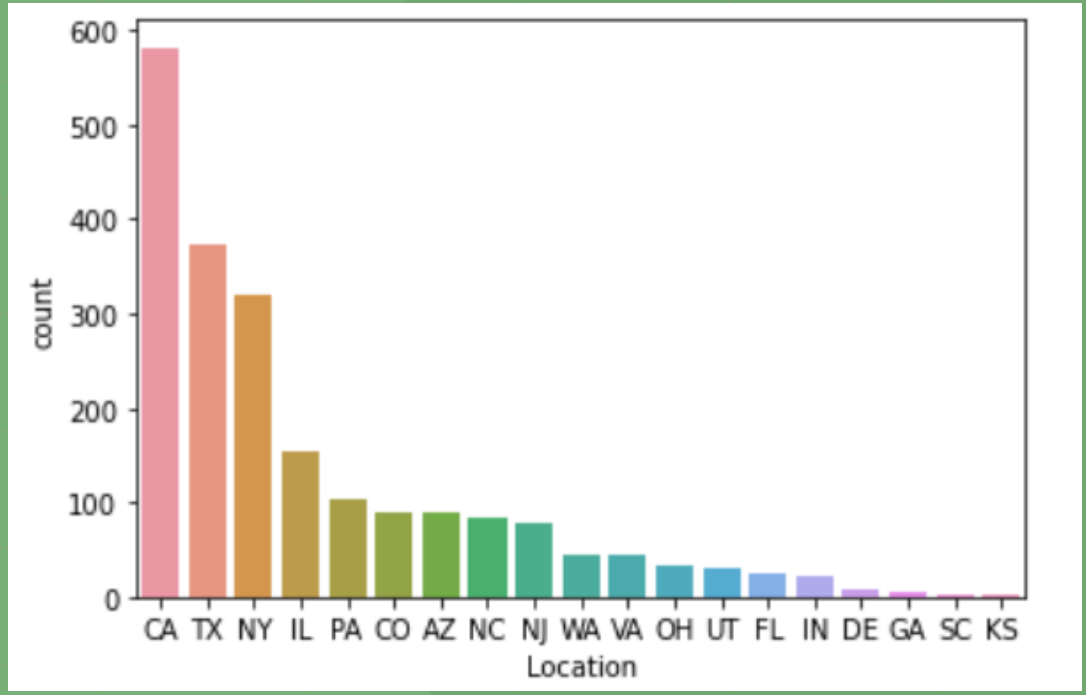
Data

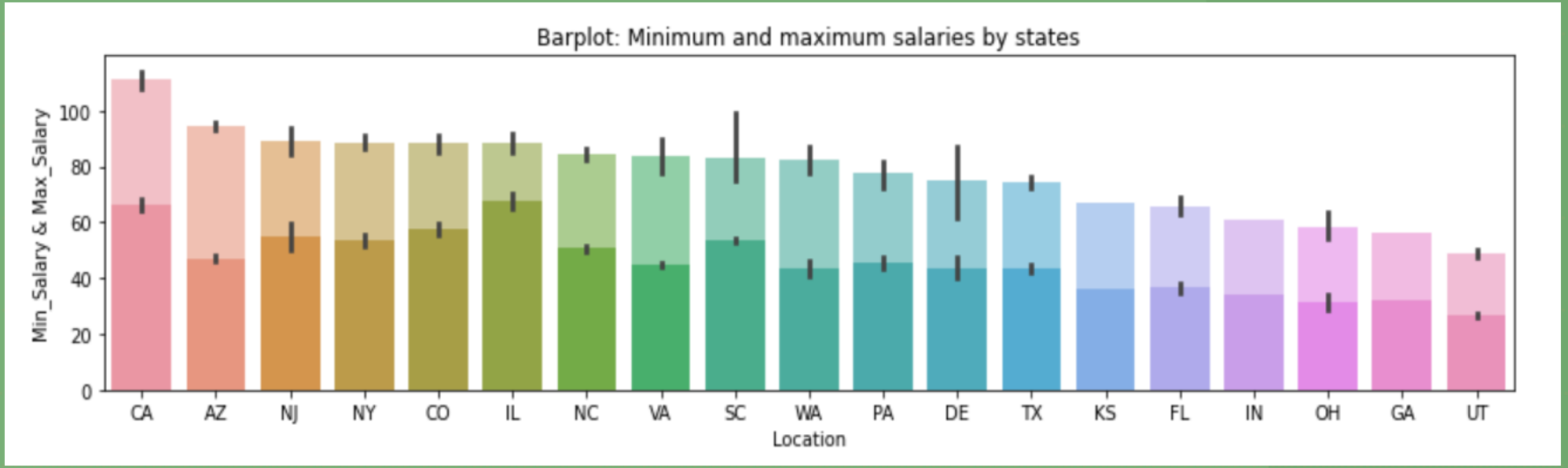
Data Source: https://www.kaggle.com/andrewmvd/data-analyst-jobs/metadata

EDA:
Skills Required

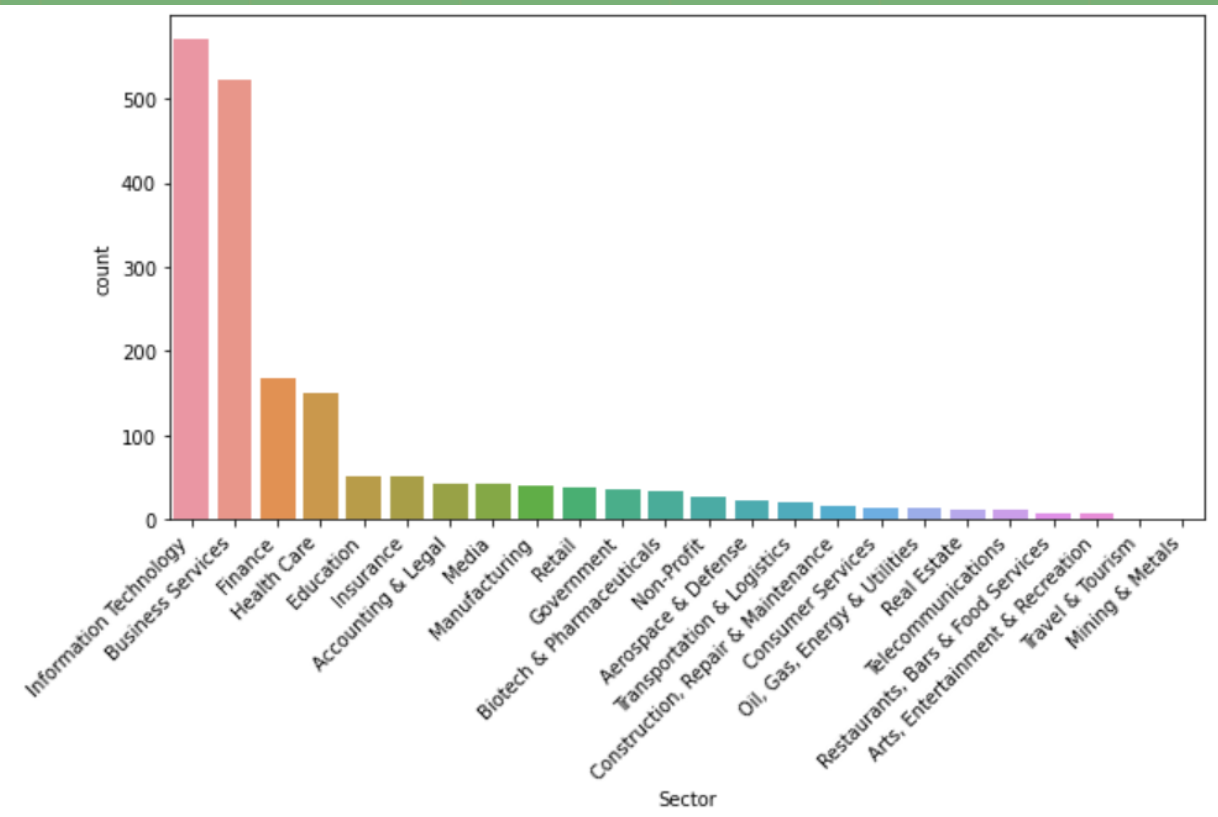


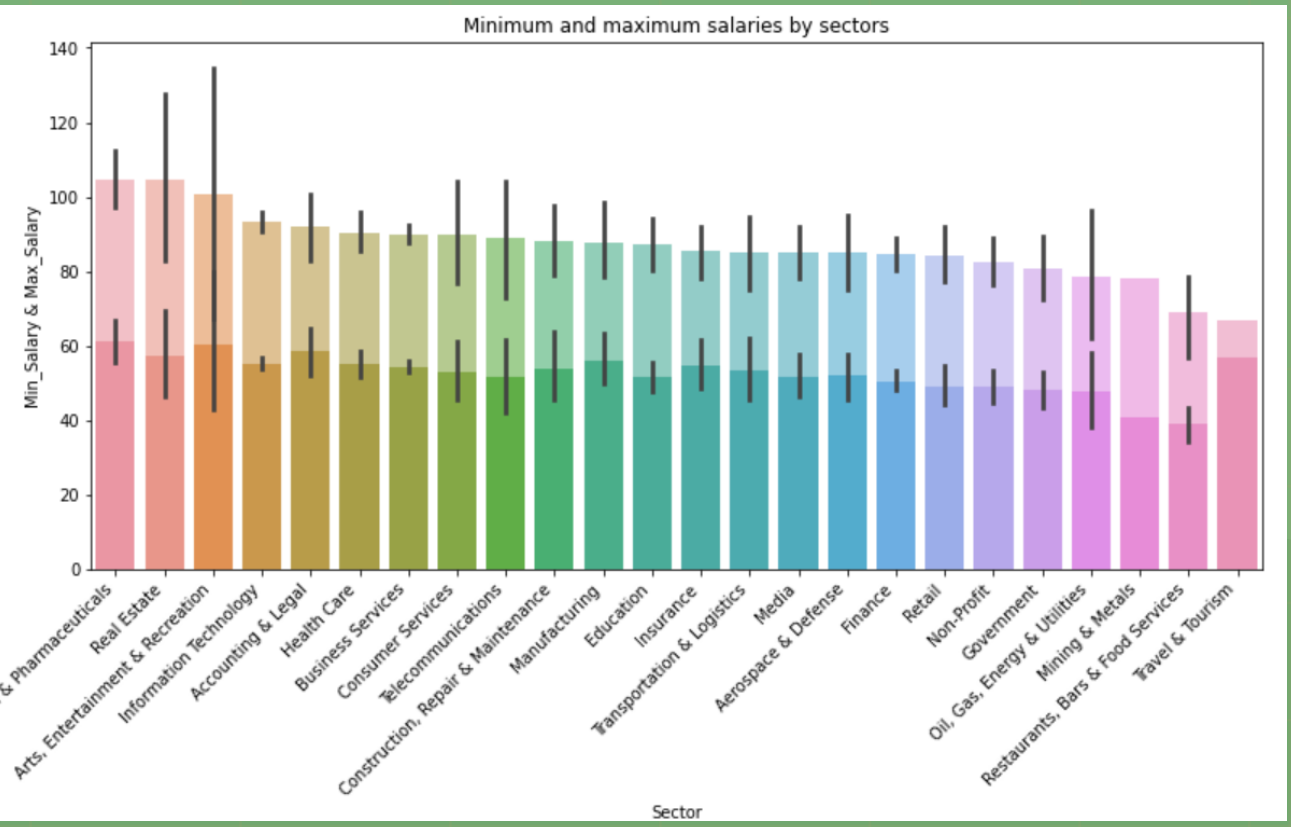
EDA: Location



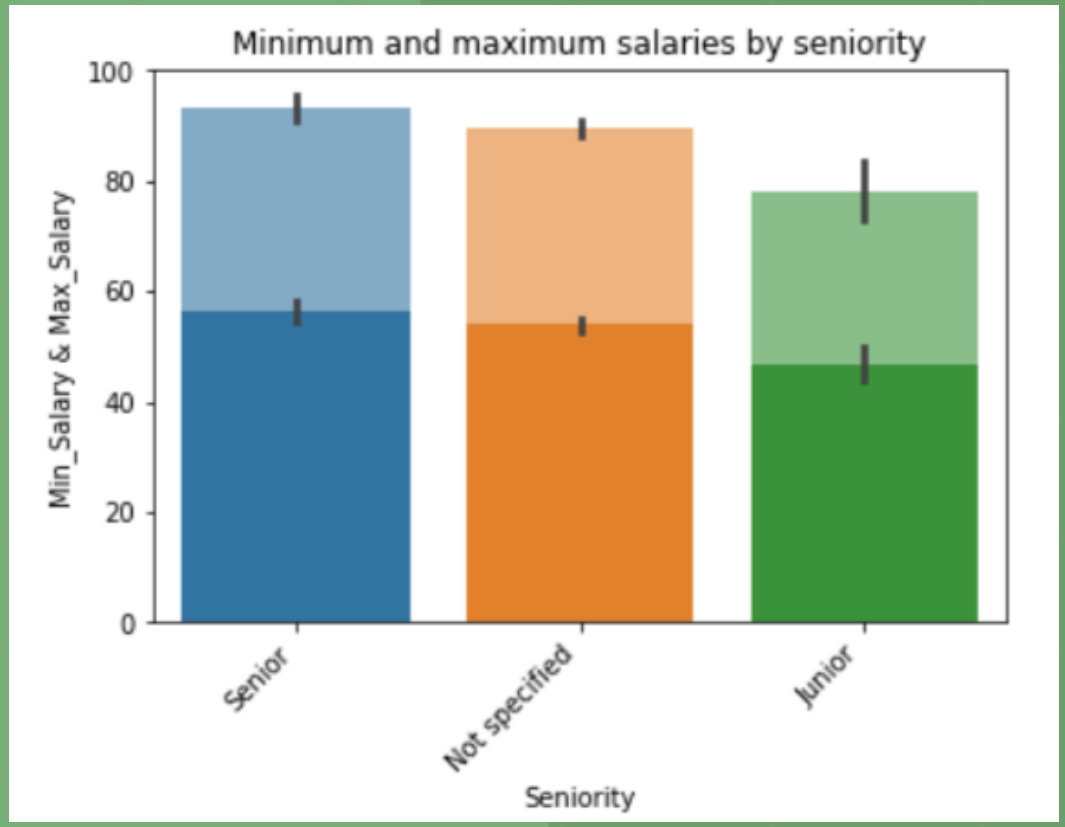


EDA: Sector


count
0
100
200
300
400
500
Information Technology
Business Services
Finance
Health Care
Education
Insurance
Accounting & Legal
Media
Manufacturing
Retail
Government
Biotech & Pharmaceuticals
Non-Profit
Aerospace & Defense
Transportation & Logistics
Construction, Repair & Maintenance
Consumer Services
Oil, Gas, Energy & Utilities
Real Estate
Telecommunications
Restaurants, Bars & Food Services
Arts, Entertainment & Recreation
Travel & Tourism
Mining & Metals
Sector

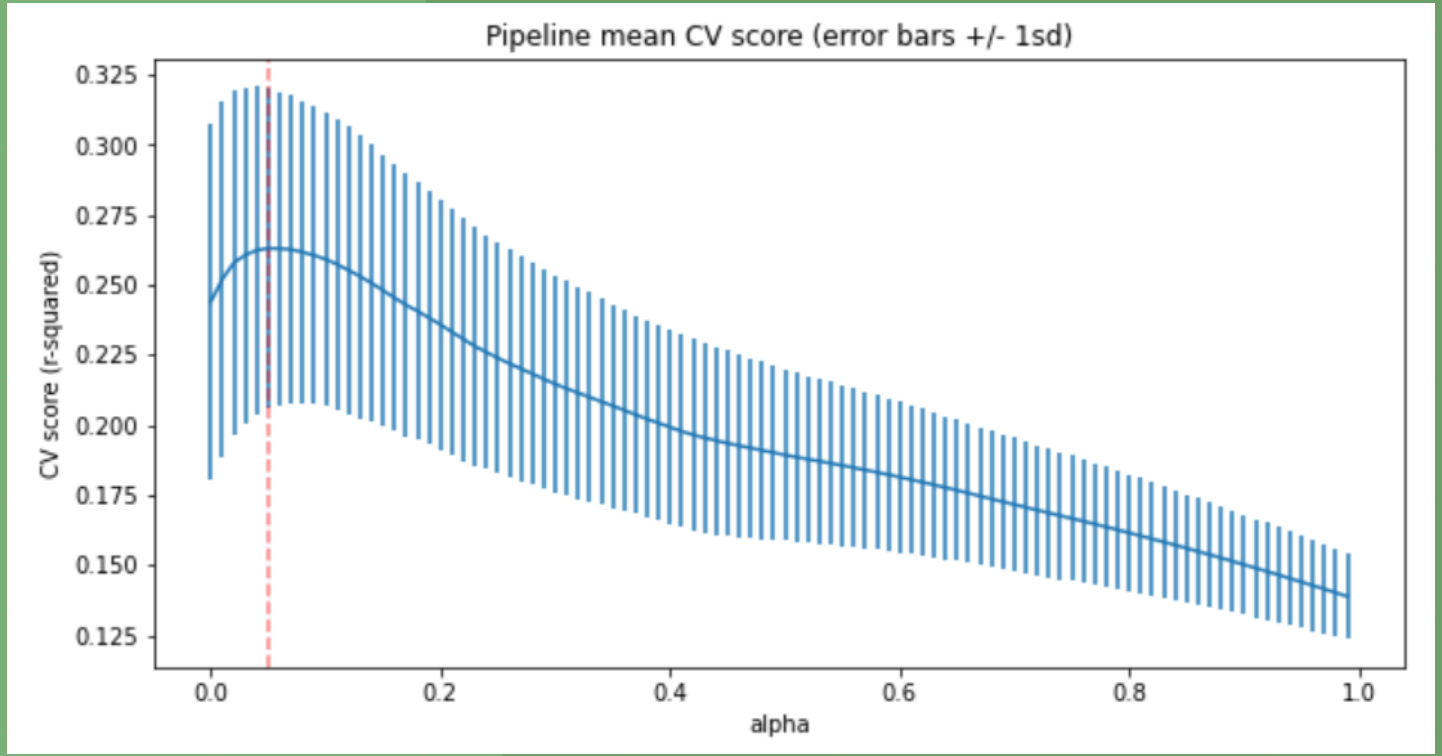

Minimum and maximum salaries by sectors
Min_Salary & Max_Salary
0
20
40
60
80
100
120
140
Biotech & Pharmaceuticals
Real Estate
Arts, Entertainment & Recreation
Information Technology
Accounting & Legal
Health Care
Business Services
Consumer Services
Telecommunications
Construction, Repair & Maintenance
Manufacturing
Education
Insurance
Transportation & Logistics
Media
Aerospace & Defense
Finance
Retail
Non-Profit
Government
Oil, Gas, Energy & Utilities
Mining & Metals
Restaurants, Bars & Food Services
Travel & Tourism
Sector

EDA: Seniority and Salary



Model Selection

- **Lasso Regression**
- **Random Forest Regressor**
- **Support Vector Regression (SVR)**



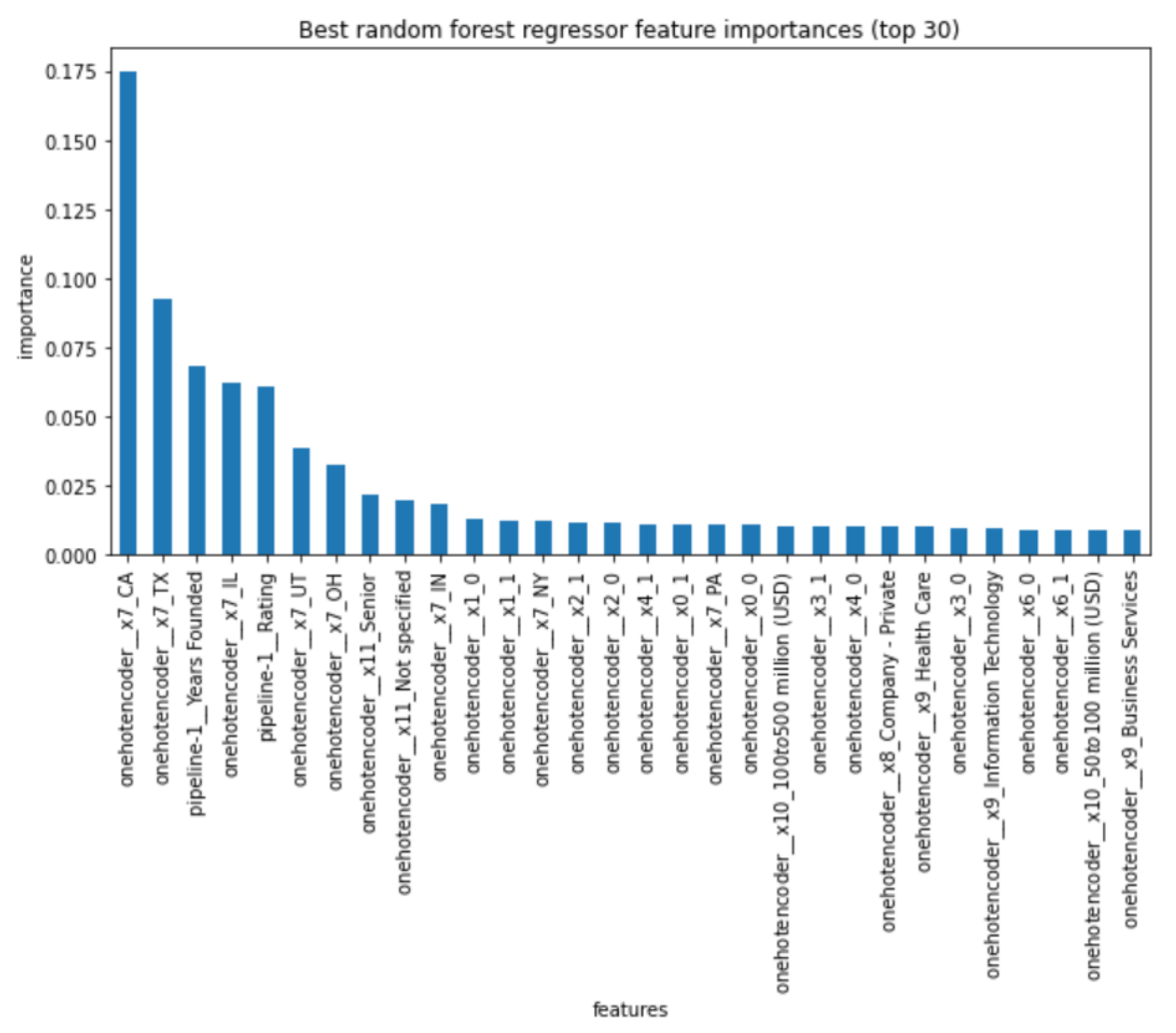
Hyperparameter Tuning

Grid Search Cross Validation

R-squared Increase by Model

- Lasso: 0.137 to 0.263
- Random Forest: 0.174 to 0.234
- SVR: 0.092 to 0.226

Feature Importance



Lasso Coefficients

	coefs	The absolute value of coefs
onehotencoder__x7_IL	20.885350	20.885350
onehotencoder__x7_CA	19.458577	19.458577
onehotencoder__x7_UT	-16.180112	16.180112
onehotencoder__x7_OH	-11.761645	11.761645
onehotencoder__x7_CO	10.928651	10.928651
onehotencoder__x7_NY	8.896904	8.896904
onehotencoder__x7_NJ	7.758899	7.758899
onehotencoder__x9_Media	-7.746343	7.746343
onehotencoder__x7_IN	-7.411508	7.411508
onehotencoder__x10_$1 to $5 million (USD)	3.494513	3.494513

Model	MAE	R-squared
Lasso(alpha =0.05)	11.72	0.263
RandomForestRegressor(max_depth = 10, max_features = 'sqrt', and n_estimators = 780)	12.17	0.234
SVR(C = 21 and epsilon = 5)	12.02	0.226

Model Performances (Minimum Salary Prediction)

Model	MAE	R-squared
Lasso(alpha =0.08)	18.26	0.254
RandomForestRegressor(max_depth = 10, max_features = 'auto', and n_estimators = 340)	18.15	0.290
SVR(C = 29 and epsilon = 10)	18.83	0.256

Model Performances (Maximum Salary Prediction)

Takeaway

Based on my findings, when determining the expected salary for a data analyst position, location, seniority, sector, rating of the company, and the number of years the company has been founded may be the most important features to be considered, especially location.

Recommendation

Recruiting firms or the department of **Human Resources** of businesses can use my models as a guide for salary benchmarking to confirm your company's compensation levels are in line with the current market, as well as to manage employees' salary expectations.

Thank You!