

# Glassdoor Job Postings: Data Science

...

By Matthew Flaherty, Sarah Torrence, Logan King and Meena Muthusubramanian

# Motivating Question

What are the best options or recommendations to find a job in the data science field?

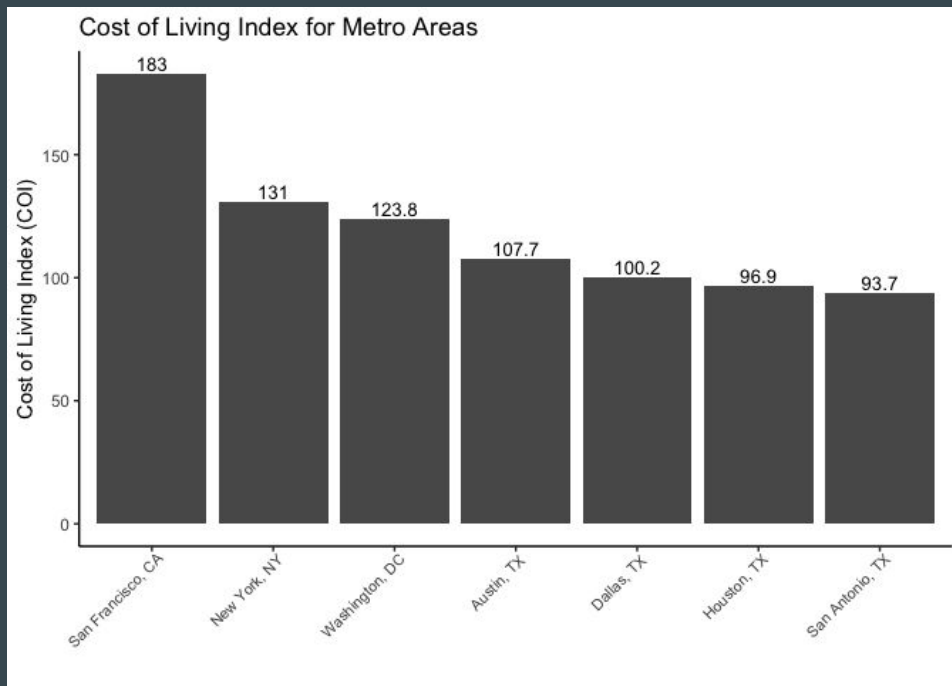
# Data

- Sources:
  - Glassdoor
  - Google Maps API
  - Advisorsmith
- Data Collection:
  - Search: “Data Scientist”
  - Locations: San Francisco, New York, Washington D.C., Texas
- Processing:
  - Combine data sets
  - Missing Values
  - Mislabeled Values
  - Feature Engineering
- Final Data Set:
  - 3,287 observations of 28 variables

city	state	job_title	industry	company	min_salary	max_salary	rating
Maspeth	NY	Diesel Mechanic	Construction, Repair & Ma...	Kingdom Associates	46298	55893	3.5
New York	NY	Principal, Sr. Consultant – Creative Te...	Telecommunications	TEECOM	64829	104769	4.1
San Francisco	CA	Senior Data Scientist – Underwriting ...	Business Services	Faire	151875	170407	4.3
San Francisco	CA	Data Scientist	Government	GovTech	78594	147225	3.6
San Francisco	CA	Data Scientist	Information Technology	Triplebyte	145000	225000	3.6
San Francisco	CA	Data Scientist	Information Technology	Notion Labs	105765	142959	5.0
San Francisco	CA	Staff Machine Learning Engineer	Information Technology	Tapjoy	137705	224163	3.5
San Francisco	CA	Senior Data Scientist	Information Technology	Autodesk	163578	182543	4.0

# Feature Engineering

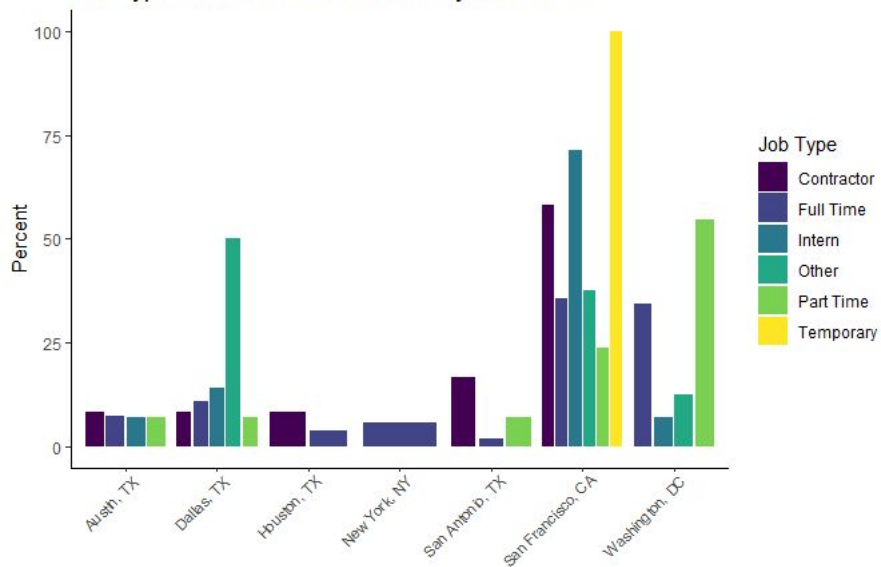
- Metro Area Feature
- Scaled Salary Feature:
  - $\text{Scaled\_salary} = \text{salary}/(\text{COI}/100)$
- Job Category Feature



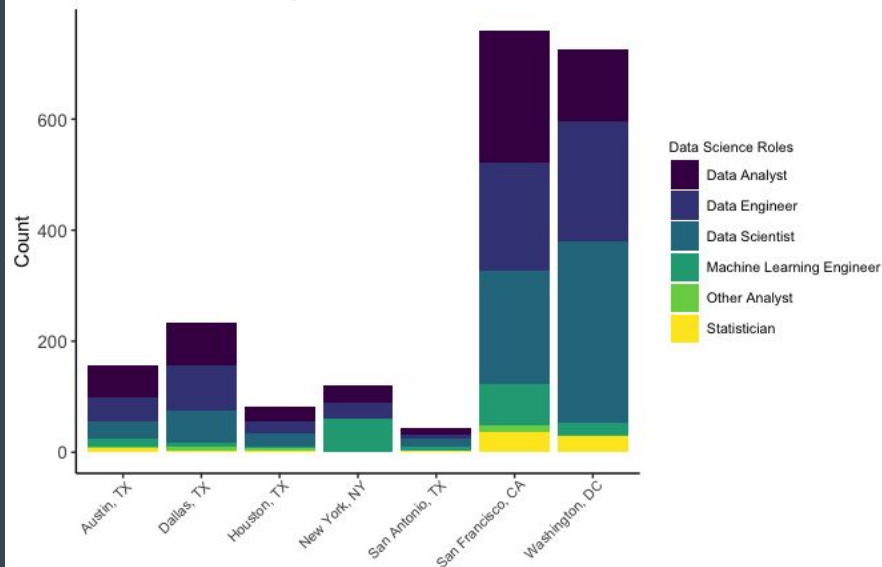
Job Category	Data Analyst	Data Engineer	Data Scientist	Machine Learning Engineer	Other Analyst	Statistician	Location Total
Austin, TX	57	44	31	13	3	8	156
Dallas, TX	76	83	56	9	6	3	233
Houston, TX	26	20	26	2	5	2	81
New York, NY	30	30	0	60	0	0	120
San Antonio, TX	12	7	15	8	0	2	44
San Francisco, CA	238	194	205	74	13	36	760
Washington, DC	130	214	327	24	1	29	725
Job Category Total	569	592	660	190	28	80	2119

# EDA Results - Data Science Roles

Job Types for Data Science Roles by Metro Area

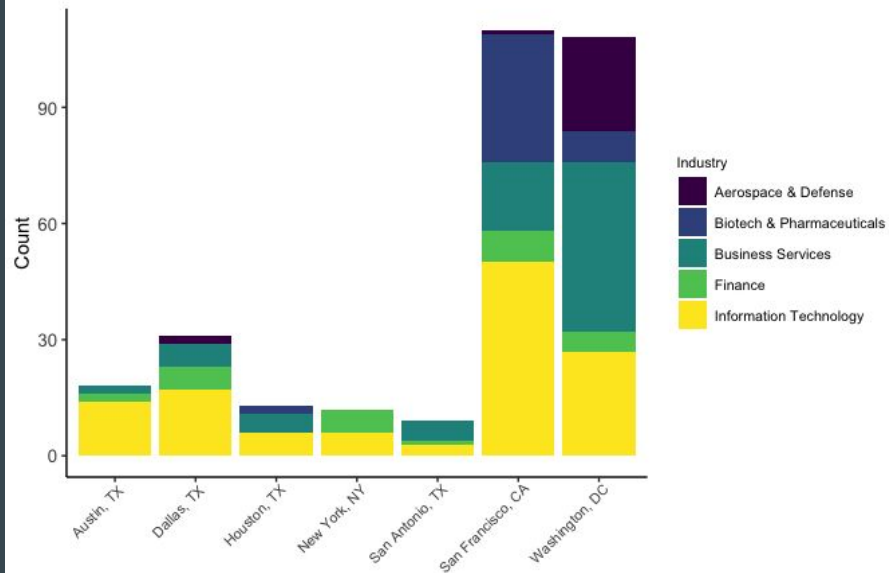


Data Science Jobs by Metro Area

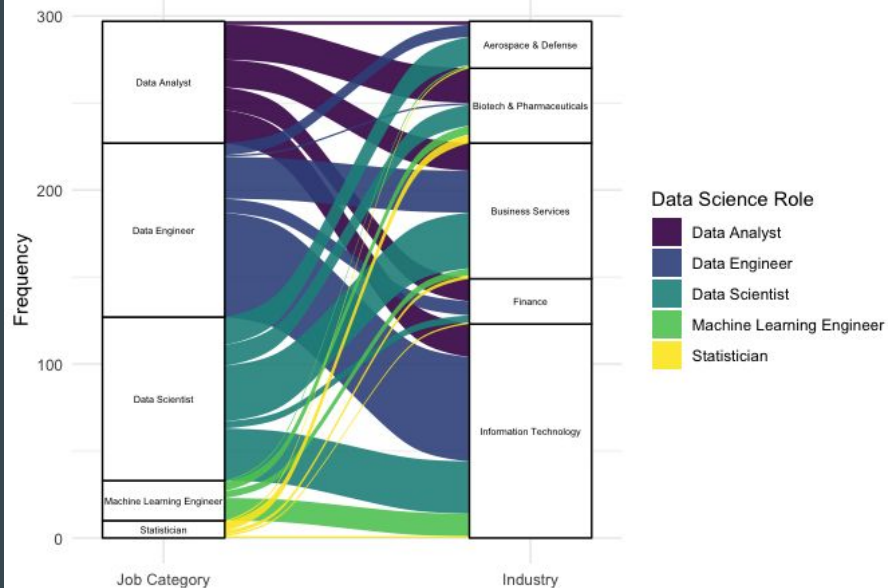


# EDA Results - Industry

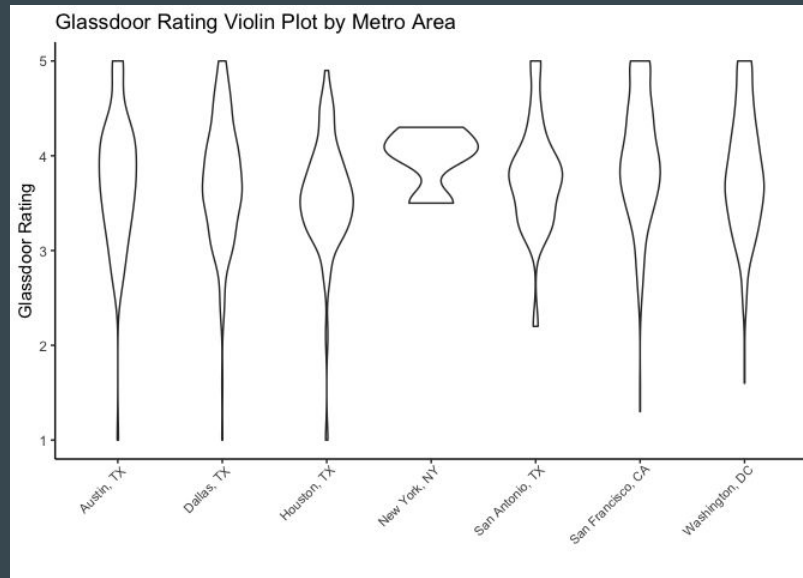
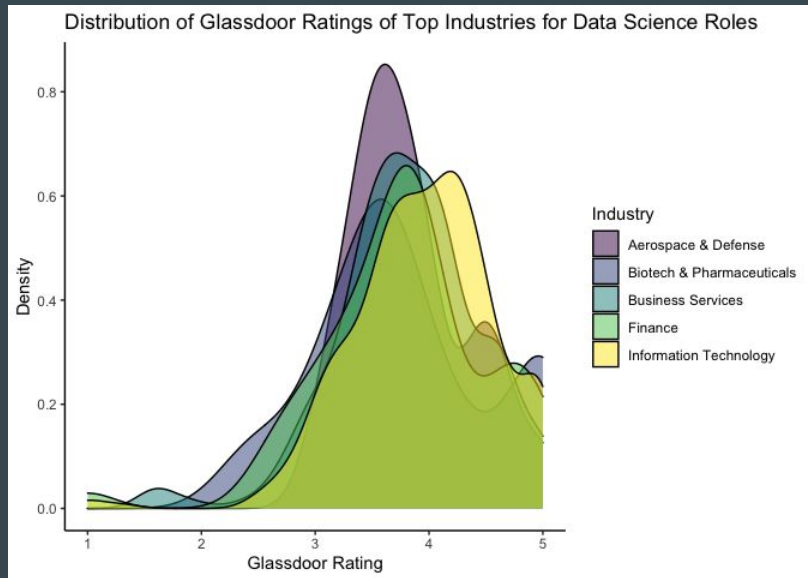
## Top Industries for Data Science Roles by Metro Area



## Data Science Job Roles in the Top Industries

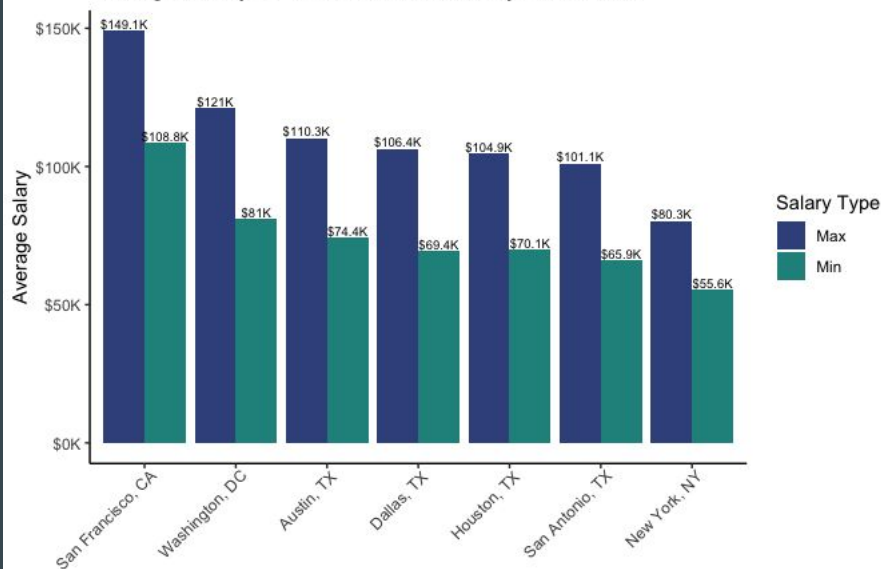


# EDA Results - Glassdoor Rating

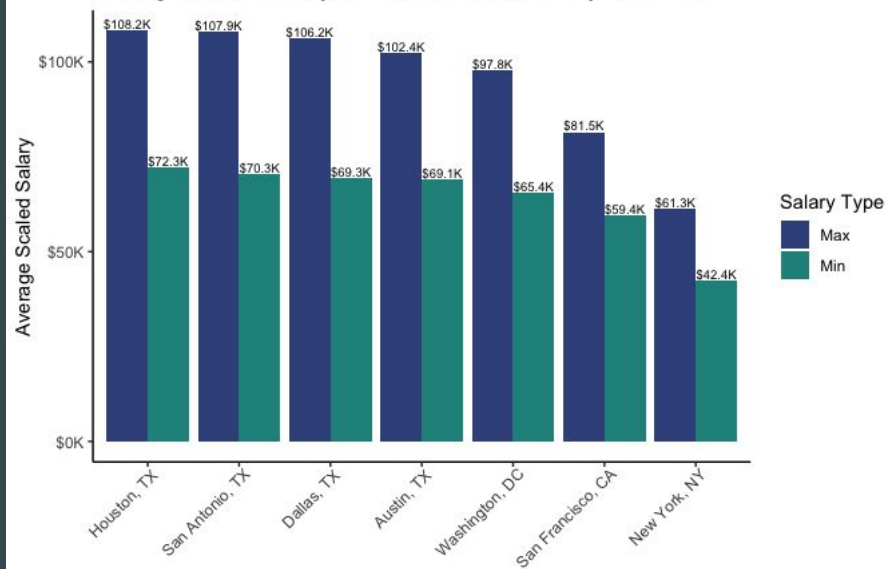


# EDA Results - Salary

Average Salary for Data Science Jobs by Metro Area

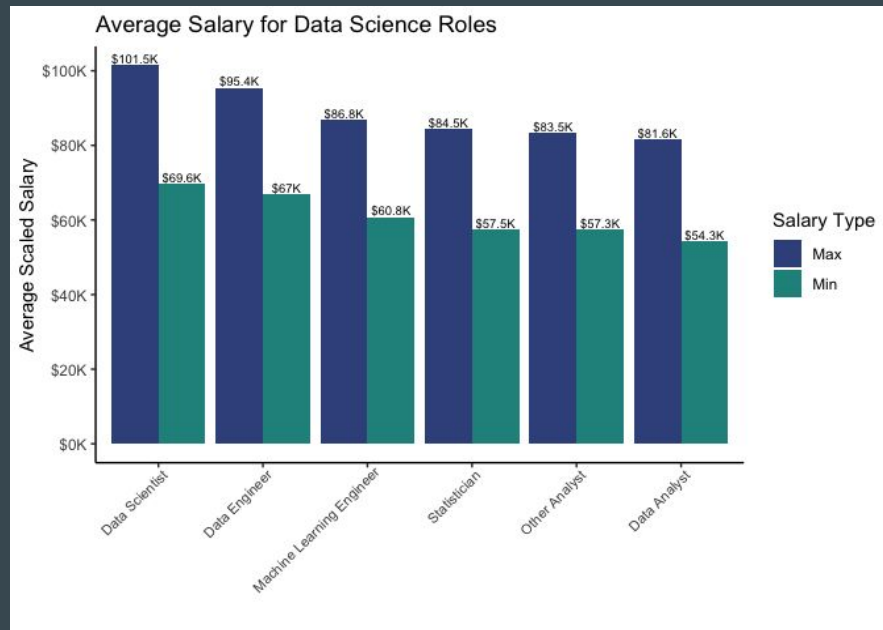
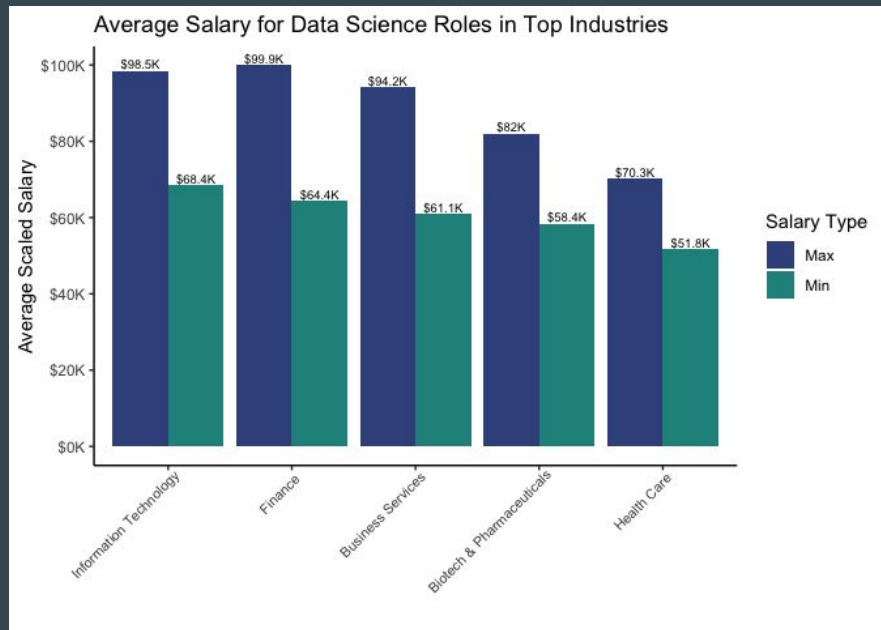


Average Scaled Salary for Data Science Jobs by Metro Area



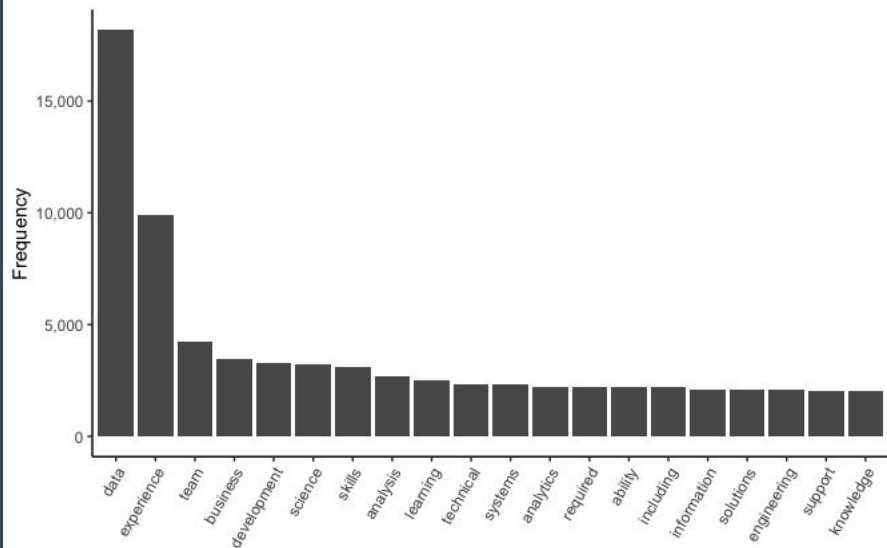


# EDA Results - Salary (Cont.)

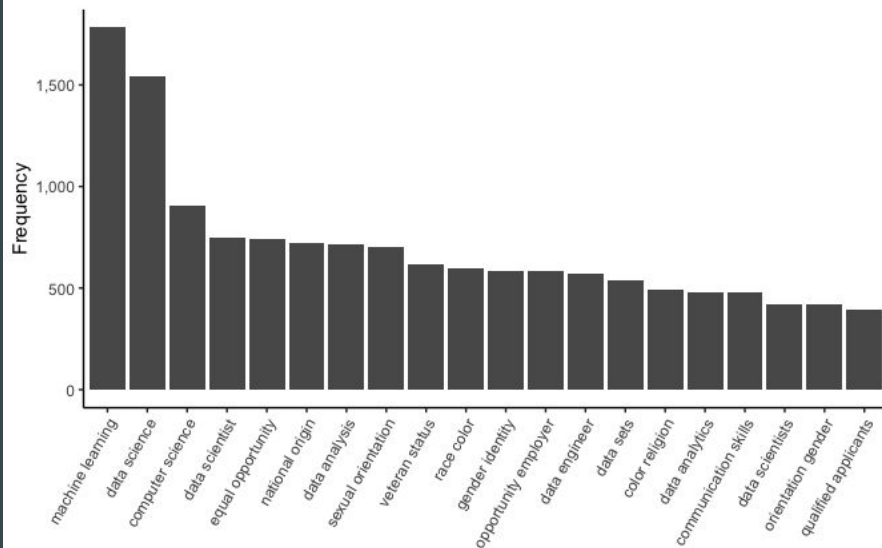


# Text Analysis

Frequency of Words  
for Data Science Roles



Frequency of Word Pairs  
for Data Science Roles



# Conclusion

This project is meant to help candidates looking for a career in the data science field. From choosing the right industry and searching for data science related jobs, this analysis is aimed to make the job searching process easier.

- Location - SF and DC
- Job categories - Data Analyst, Data Engineer, Data Scientist
- Industry - Information Technology

# Limitations, Improvements, and Next Steps

- Data collection issues
- Scope of included locations
- Nature of company ratings