OVERVIEW

- Question
- Collection methods
- Models used
- Conclusions

QUESTION

- What factor(s) most directly impact data science salaries?

COLLECTION METHODS

- Web scraping Indeed.com
- Set up a request to the URL
- Extracted the information (job title, location, salary and company)
- Saved to a CSV file
- Performed analysis

COLLECTION METHODS

	company	job_title	location	salary	city	median	med_binary
0	Platinum Solutions	Data Scientist	Houston, TX	100000	Houston	LOW	0
1	Horn Solutions, Inc.	Sr Data Scientist (Hadoop) - Duluth, Georgia	Houston, TX	105000	Houston	HIGH	1
2	Genialis	Bioinformatics Software Developer	Houston, TX	65000	Houston	LOW	0
3	Baylor College of Medicine	Quality Assurance Analyst (Research)	Houston, TX	43794	Houston	LOW	0
4	Baylor College of Medicine	Flow Cytometry Specialist I	Houston, TX	46831	Houston	LOW	0

COLLECTION METHODS

	salary
count	286.000000
mean	108703.426573
std	47127.068596
min	24000.000000
25%	72904.750000
50%	100000.000000
75%	141875.000000
max	250000.000000

MODELS

- Logistic Regression based on location (city) for the entire sample of job listings with an annual salary
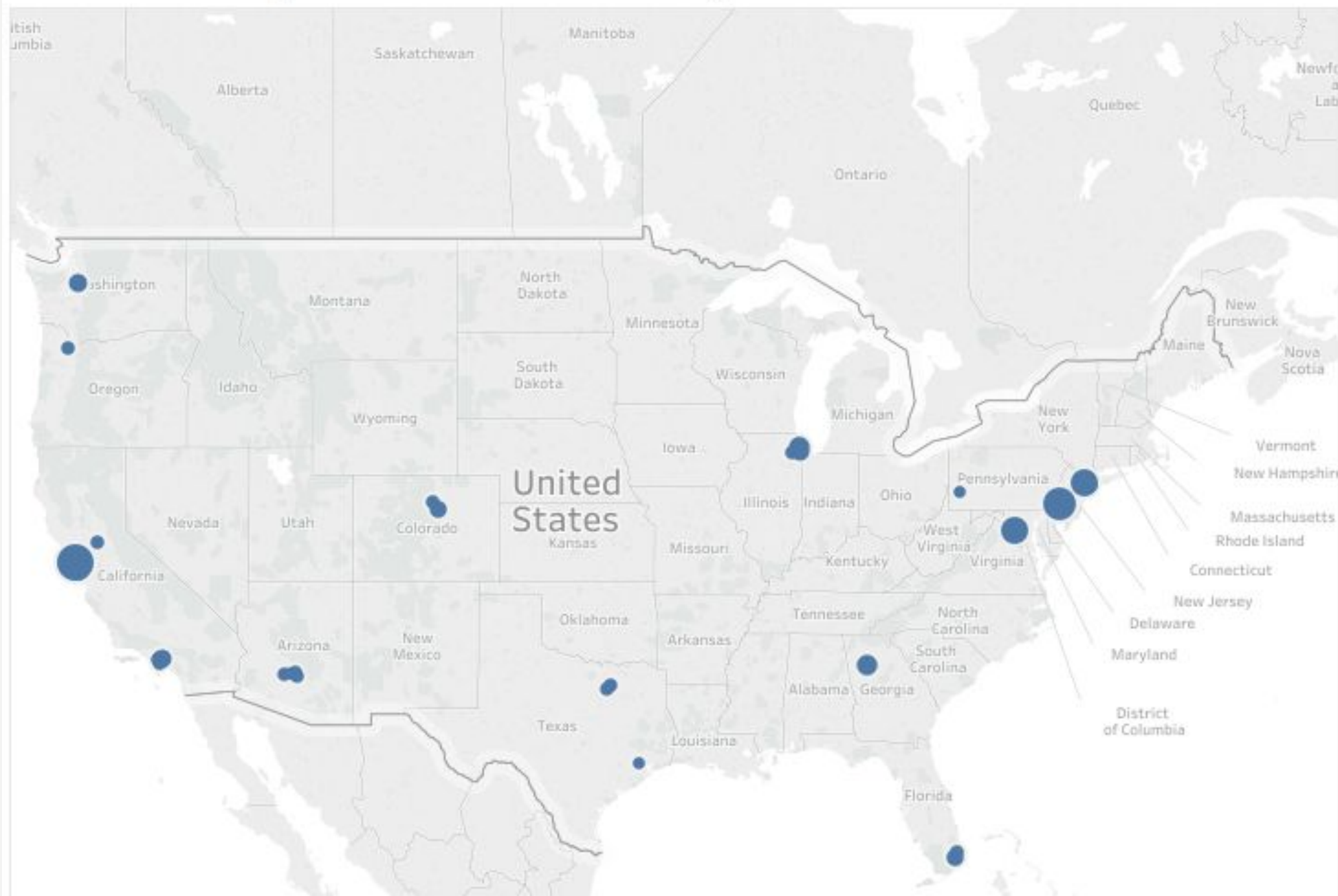
MODELS

- Senior Titles: 8
- Junior Titles: 0
- Scientist Titles: 108
- Data Titles: 114
- Manager Titles: 12
- Analyst Titles: 65

MODELS

- Accuracy scores for each model

Model Based on Location:	Score
Total sample	0.716783216783
Job titles with Data	0.763157894737
Job titles with Scientist	0.731481481481
Job titles with Analyst	0.676923076923

Odds to Earning Above the Median Salary

CONCLUSION

- The data shows us the odds of earning above the salary mean ($100,000.00) per city.

- The odds that someone would earn above the median salary is highest in San Francisco at 5.020650 and the odds that someone would earn below the median salary is lowest in Austin at 0.468744.