Regional Data Job Market Dashboard

—

By: Manisha Shetti, Cory Ptacek, Nish Trivedi, Beth Hakamy, Lisa Weinstein

Project Summary

This project looks at Kaggle.com 2018 data that was pulled on data scientist, data analyst and data engineering positions across the United States of America. The salaries, titles, companies, company headquarters and sector were all part of the dataset.

As a team, we analyzed, graphed and created a dashboard in order to see a picture of what types of jobs were available to those looking for data driven positions.

We used flask, leaflet.js, canvas.js, HTML and Bootstrap CSS to build the dashboard

Data - Kaggle.com -

Positive Attributes

- All files were in the same format and provided by the same source so merging the files was simple
- The data was well organized
- The data had enough consistent information for conclusions to be drawn
- The data listed company names and the salary ranges and you could see a difference in sector
- The data had millions of records

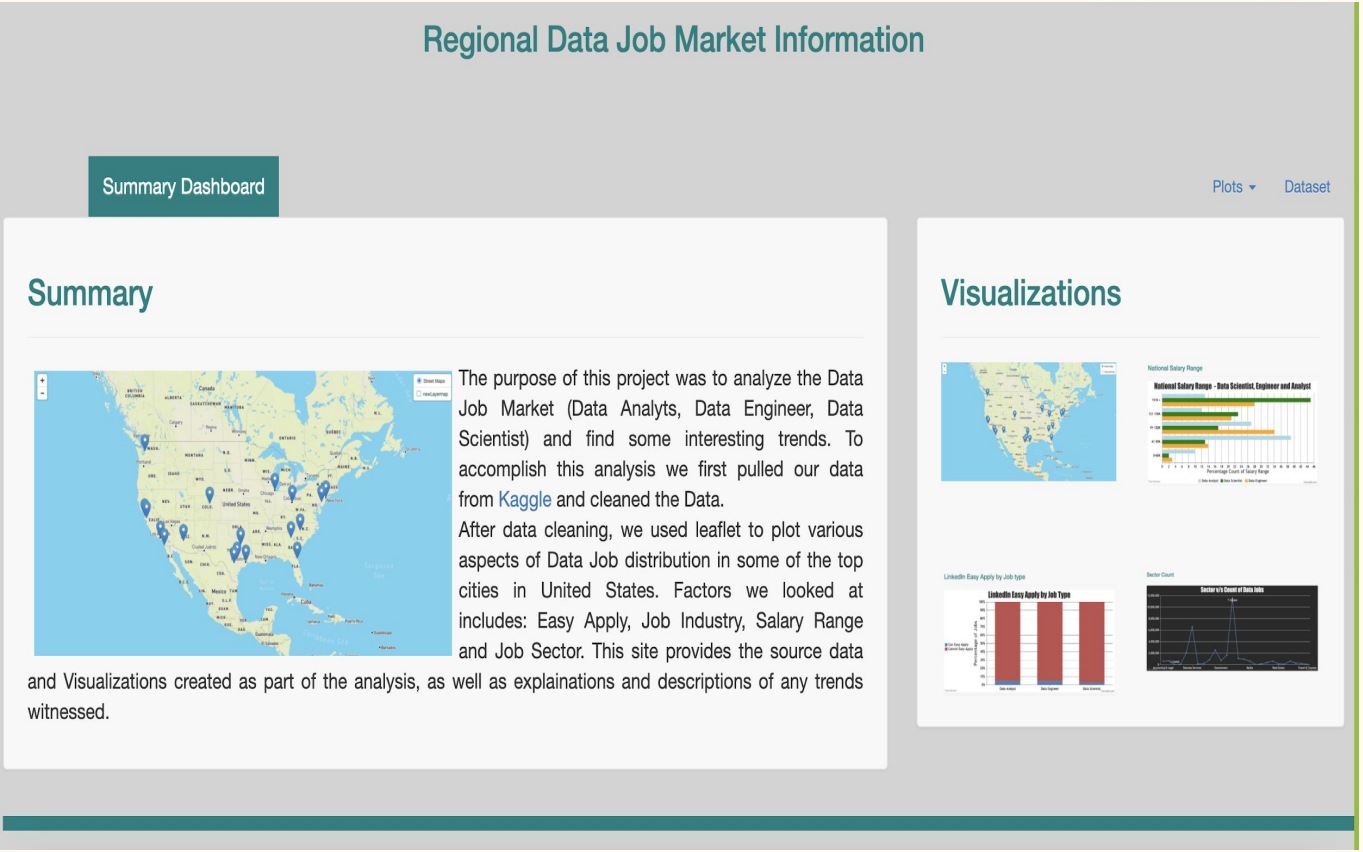
Negative Attributes

- The salary range was difficult to slice and dice based on the formatting (string with verbiage)
- The salary ranges were not consistent
- The job titles were not consistent
- The jobs did not have address to allow us to map a specific city to show a street view - we did explore a heat option but leaflet proved challenging
- The data had millions of records

Process

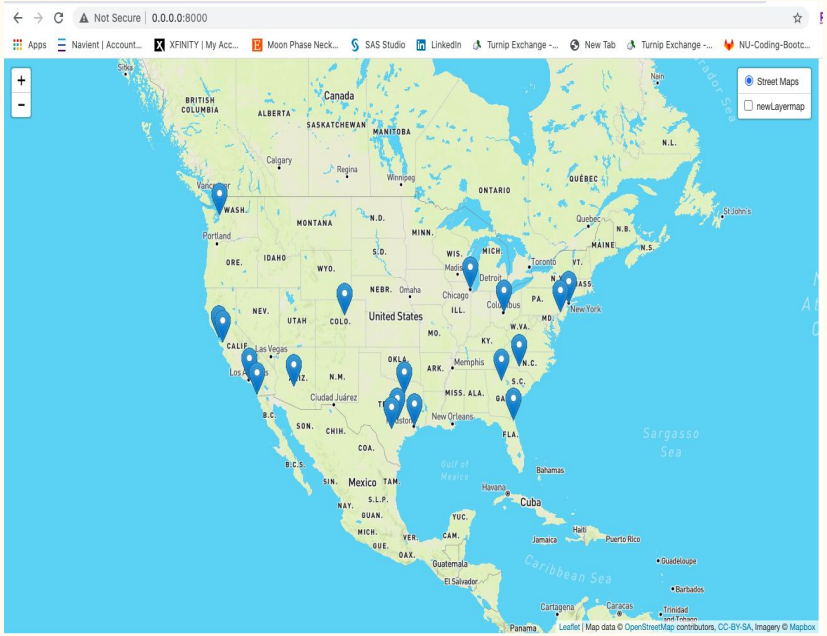
- Data cleaning was done using Pandas Dataframe.
- Merged all three Dataframe into one Combined Dataframe.
- Created Database Table by establishing connection with PostgresSQL.
- Created Python Flask app and connected this with Database using SQLalchemy.
- Created JavaScripts to pull data and perform analysis on various aspects.
- Created a dashboard page using HTML and Bootstrap CSS.
- Presented our visualization using Leaflet.
- Implemented a new JavaScript library called CanvasJS for additional visualizations.

Dashboard



- This dashboard page has multiple charts that updates from same data.
- This dashboard page includes user-driven interactions by click through and Plots drop down.
- This dashboard has 4 views/visualizations.

Leaflet Map Graph



Leaflet was used to visually show the Data Analyst, Data Engineer and Data Scientist opportunities were coming from.

Pros:

- Easy to visualize locations of job opportunities

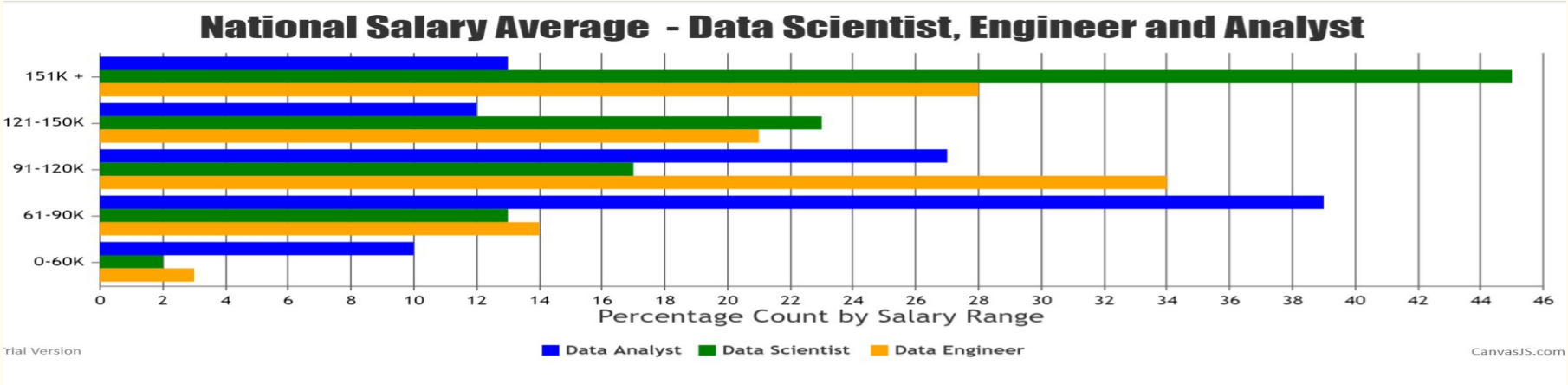
Cons:

- Lookup latitude and longitude for each location

Location with the most opportunities:

- Austin, TX, New York, New York, Los Angeles, CA, San Jose, CA, and Chicago, IL

Salary Information



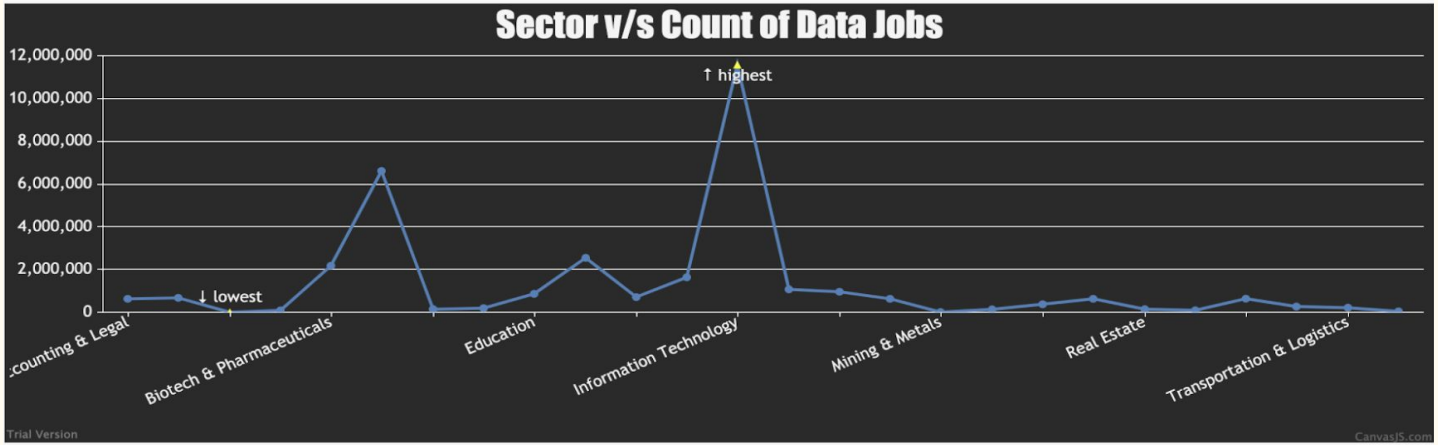
Technical Notes:

- Cleaned data to determine a salary range.
- Sliced it based on the maximum point into 5 ranges.
- Plotted bar charts for the 3 job titles in the data set.

Data Observations:

- “Data Scientists” get the highest pay.
- “Data Analysts” compensation comes in the 2 & Mid tiers (61-120). Signals it’s a step towards more specialized, or advanced future role.
- “Data Engineers” pay was distributed mostly in 2 tiers; mid & high = > specialization & experience!

Sector Information



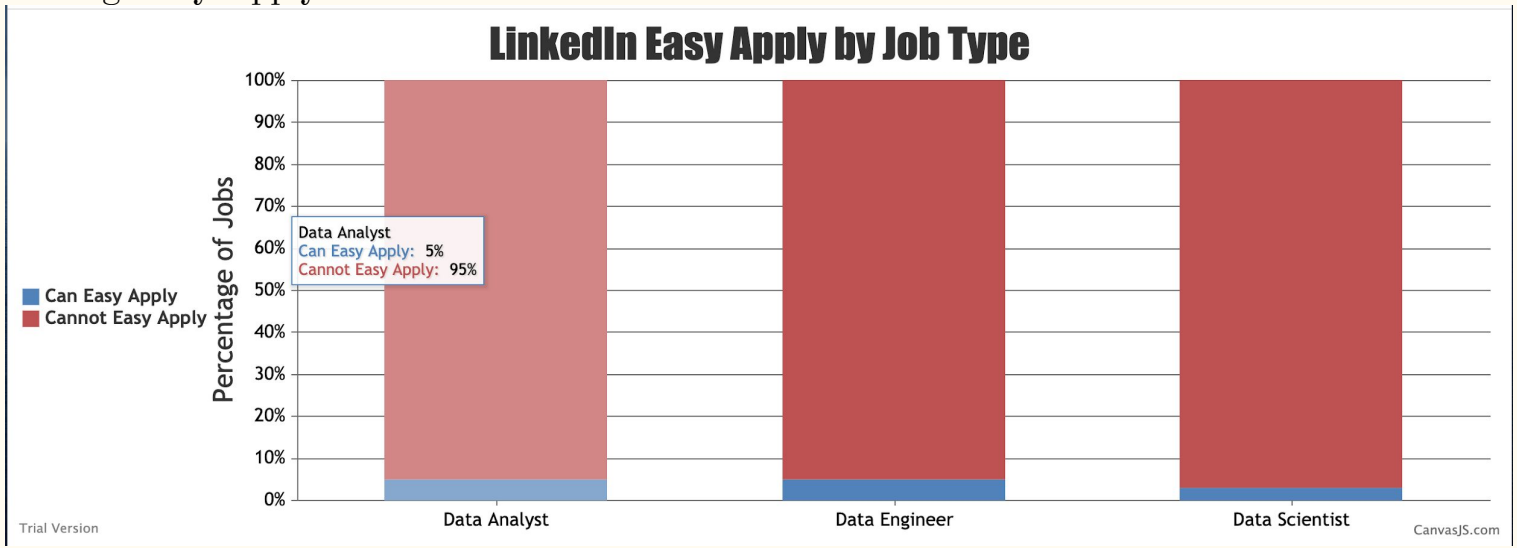
Row Labels	Sum of index
N/A	5320928
Accounting & Legal	628165
Aerospace & Defense	674876
Agriculture & Forestry	30626
Arts, Entertainment & Recreation	85203
Biotech & Pharmaceuticals	2177144
Business Services	6607606
Construction, Repair & Maintenance	144908
Consumer Services	190872
Education	861539
Finance	2545241
Government	710706
Health Care	1637650
Information Technology	11580247
Insurance	1071749
Manufacturing	958272
Media	629570
Mining & Metals	7549
Non-Profit	134593
Oil, Gas, Energy & Utilities	375065
Real Estate	143328
Restaurants, Bars & Food Services	90553
Retail	634691
Telecommunications	266417
Transportation & Logistics	213565
Travel & Tourism	41332
(blank)	
Grand Total	**37762395**

- **Top Fields** for data roles are **Technology, Business and Financial Services.**
- **Biomedical** (including pharma) field professionals is that this sector comes **4th.**
- **Limited Opportunities:** construction, agriculture, mining, restaurants, entertainment, and tourism.
- **Limitation**: non-categorized!

Easy Apply Information

- Data Analyst and Data Engineer had 5% of jobs in the dataset that use Easy Apply
- Data Scientist had 3%
- Still over 1,400,000 jobs able to be applied for using Easy Apply

Sum of index	Column Labels		
Row Labels	No Easy Apply	Easy App	Grand Total
Data Analyst	8217078	382280	8599358
Data Engineer	14814107	715293	15529400
Data Scientist	10485570	371014	10856584
Grand Total	**33516755**	**1468587**	**34985342**



Conclusion

- Using the aforementioned techniques, we have analyzed the data-oriented job market for job-seekers
- Popular areas include New York City, California, Texas, and Chicago (woo!)
- Salary
  - Most Data Analyst roles pay in the 61-90K range
  - Most Data Engineer roles pay in the 91-120K range
  - Most Data Scientist roles pay in the 150K+ range
- IT, business, and financial services has the most jobs, while other specialized fields offer limited (but unique) opportunities.
- Easy Apply analysis showed that while there are a significant amount of data jobs available through this platform, it is not the norm nor applicable to majority of these roles