Sarah Torrence
Logan King
Meena Muthusubramanian
Matthew Flaherty December 10, 2020

**Glassdoor Job Postings: Data Science**

As data science master's students who will be joining the workforce soon, we are naturally interested in what the job market looks like in the data science field. We have all been through the job search process and have found that it is easy to become overwhelmed with the seemingly endless amounts of information and job postings available. Job sites aim to help streamline the job application process and provide users with information about various companies at which users may be interested in applying, so we began our search process here.

After looking through various data sources, we chose this data set because of its relevance to data science and how it has a "real-world" feel when it comes to searching for jobs in the market. The data comes from Glassdoor, one of the world's largest job and recruiting sites, in which one can find job listings, company reviews, ratings for the company, salary insights according to the role, interview questions and benefits. People often turn to Glassdoor to gain insight about companies and for more information about jobs or interviews. With our dataset, we attempt to answer the following motivating question: What are the best options or recommendations to find a job in the data science field?

In this report, we have explored the availability of job openings for data science related positions (Data Scientists, Data Engineer, Machine Learning Engineer, Data Analyst, Statistician and Other Analyst) with respect to certain metro areas, companies, and skills required. Further, we have analyzed the dataset based on job types and salary, and analyzed the number of positions available for data science specific jobs in each metro area. With this information, job seekers can search for openings based on specific metro areas, industries, companies, salaries and roles in which the most openings are available or which best suit their needs. As finding a job can be a grueling process, this report is meant to aid and inform data science professionals in their job search.

**Data:**

The main source of data we used for this analysis was scraped from the website Glassdoor. Glassdoor is a website in which current and former employees can anonymously review companies. This includes rating the company (on a scale from 0-5), leaving comments and documenting salaries for certain positions. Glassdoor anonymously posts reviews and reports salary ranges for the positions and companies inputted by users. Glassdoor also allows companies to add job openings to their website and combines the reviews and rating information with these job openings to allow the user to search for positions, review job descriptions, salary information and ratings for the company all on one page. The data for this report was scraped on June 5, 2020 from a search for job openings for the position "Data Scientist" in four separate scrapes for the locations New York, San Francisco, Washington, DC and Texas. These locations were chosen as the initial individual who collected the data felt these were good areas in which one could find data science related jobs.

The unit of analysis, each observation within the data, is for a job opening that popped up in the search for "Data Scientist" from one of the above locations. The data was originally

structured in four separate files, one for each location and includes information on city, state, company, Glassdoor rating, job description, job type, salary ranges, etc. for data science job openings. The cities in the data set are all within a small radius of the searched city, meaning they are not all in the searched location, but within the metro area of the searched city. For example, in the search for Washington, DC, a job opening could arise from Arlington, VA which is adjacent to the city limits of Washington, DC. The one exception is the Texas data set in which data was scraped for the entire state of Texas rather than the metro area of one city. The company ratings, a numeric variable, are determined by recent employee feedback and are on a 5-point scale as follows:

- 0.00 - 1.50 Employees are "Very Dissatisfied"
- 1.51 - 2.50 Employees are "Dissatisfied"
- 2.51 - 3.50 Employees say it's "OK"
- 3.51 - 4.00 Employees are "Satisfied"
- 4.01 - 5.00 Employees are "Very Satisfied"

Salary is recorded as a minimum and maximum salary for the particular job opening (in USD) based on salary values inputted by users for that same role at that company. The job type is the type of position listed such as full time, part time, intern or contractor. The job title and job description are the title and description for that particular position. The industry is the industry in which the company is categorized within.

We combined the four data sets, as they had all the same variables, to create one data source for analysis. We started with 3,324 observations of 13 variables, but the dimensions later changed as we cleaned and processed the data. Most of the categorical variables had blank cells to indicate missing values and the salary variables had a value of -1 to indicate salary information was missing for that observation. In these instances we coded in all missing values as NA in order to easily identify missing values and easily omit them in further analysis when applicable. There were also a few inconsistent state and city names for example “Texas” and “TX” where both in the data set. In this instance, to be able to analyze and visualize the data by groups, we need all job openings from Texas to be coded in as “TX”. We fixed these values as well.

To be able to perform meaningful analysis between locations and jobs roles within the field of data science, there were several additional features we had to create within the data set. First of all with some of our data for metro areas that spanned across different states and some of our data for an entire state, we needed to create a way to have meaningful location-based groups. In addition, salary information can vary greatly across the country based on the cost of living in different areas so we also wanted to find a better way to make meaningful and accurate salary comparisons. Lastly, as data science is a large field, we wanted to create a way to categorize job openings based on the type of role or expertise required. We used additional data and variables within our data set to wrangle the data and create features in each instance. Our final data set was 3,287 observations of 28 variables.

**Methodology/Feature Engineering:**

**Metro Area Feature:**

When collecting our initial data, there were four CSV files which contained job postings, each CSV representing jobs in different regions of the country (New York, San Francisco, Texas,

and Washington, DC). Upon our initial analysis of the data, we found several issues which would create problems for our analysis that needed to be addressed.

While we expected to find multiple states within a given dataset (Washington, DC would have DC, Maryland, and Virginia; New York would have New York and New Jersey), we also found several states which did not fit in with our initial regions outlined in the CSV files: Kentucky, North Carolina, and Tennessee (each containing one job posting). Since these observations did not make sense to include in our final analysis, they were removed.

Next came the issue of small sample sizes for individual cities. Job postings have listed the city that they are located in, so it came as no surprise that there were many unique cities to go along with the vast amount of job postings. However, with over 100 unique cities, there arose sample size issues that would negatively impact our final analysis. Our solution was to group each individual city into a larger metro area in order to have more reliable results.

We began by manually grouping cities with metro areas based on data obtained from (https://advisorsmith.com/data/coli/). The only city that was not grouped into a metro area by our manual metro area mapping was Paris, TX (which contained only one observation), so it was removed from the dataset. Additionally, six observations in Texas did not contain city values, so they were also removed. The manual metro area mapping left us with 16 distinct metro areas.

In order to check the accuracy of our manual metro area mapping, we conducted analysis of the individual cities and metro areas based on their location (latitude and longitude coordinates). Using geocoded location data from Google's API, we were able to assign both latitude and longitude values to each city and their associated metro area. We also constructed the weighted latitude and longitude of each metro area (value was a weighted average of the metro area's individual cities, weights were based on the number of observations in each individual city). The difference between the weighted and normal latitude and longitude never exceeded 0.7 degrees, so we are confident that our manual mapping was accurate.

With the locations of individual cities and the regular and weighted locations of their associated metro areas constructed, we can calculate the distance between each city and its associated metro area in order to capture any outliers in the city and metro area groupings (cities that may be far away from the metro area that they are grouped with). Our results showed that all except one city (Freeport, TX with metro area Houston, TX) were within 50 miles of the metro area with which they were grouped (50 miles was chosen after visual examination of maps along with research of local metro area considerations). Given its distance from its associated metro area, not routinely being considered part of the Houston metro area, and small sample size (2 observations), Freeport was removed from the dataset. Additionally, the following metro areas had small sample sizes and were not able to be grouped with any other metro area based on distance and were subsequently removed from the dataset: Brownsville, College Station, Corpus Christi, El Paso, Huntsville, Killeen, Lubbock, Lufkin, and Marshall (all located in Texas).

Our final set of metro areas is as follows: New York, NY; San Francisco, CA; Washington, DC; Austin, TX; Dallas, TX; Houston, TX; San Antonio, TX.

**Scaled Salary Feature:**

Depending on where one lives, the cost of living is different based on many factors including housing, transportation and food costs. To compensate for differences in cost of living, salaries for the same job might be different depending on where the job is located. We knew that a direct comparison of salaries between different metro areas was not going to be accurate so we

decided to scale the salary data based on the cost of living for each metro area. To scale the cost of living, we used the cost of living index (COI).

We collected data for the cost of living from AdvisorSmith (https://advisorsmith.com/data/coli/). This data from June 5, 2020 includes the COI for 509 metropolitan areas within the United States. The COI is modeled on national average household budgets and has weights assigned to the follow 6 major categories of household expenses (weights as percentages):

- Food: 16.1%
- Housing: 23.2%
- Utilities: 10.1%
- Transportation: 18.6%
- Healthcare: 9.6%
- Consumer Discretionary Spending: 22.3%

A COI of 100 is the average cost of living for the United States. If a city's COI is above 100, it has an above average cost of living and if it is below 100 it has a below average cost of living. For example, the New York City metro area has a COI of 131 meaning it has a 31% higher cost of living than the national average.

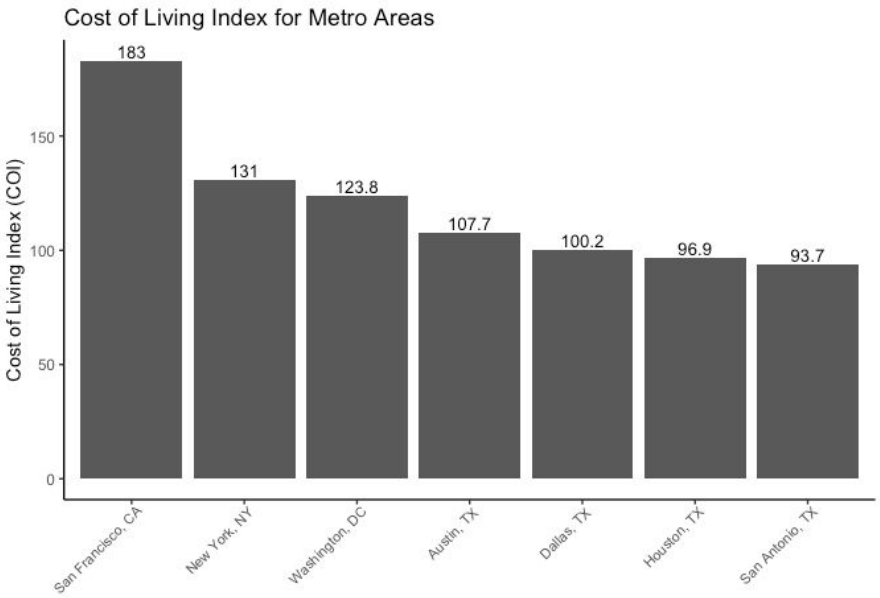


Figure 1

In Figure 1 we can see the COI for the seven metro areas in our data. The San Francisco Bay area has by far the highest cost of living followed by New York City and Washington, DC. All four metro areas in Texas are fairly close to the average cost of living in the United States.

We created two new variables for the minimum and maximum scaled salary, calculated in the following way:

- ***min_salary/(COI/100) = min_scaled_salary***
- ***max_salary/(COI/100) = max_scaled_salary***

This new scaled salary information allows us to compare metro areas against one another and utilize scaled salary ranges to be able to compare across regions.

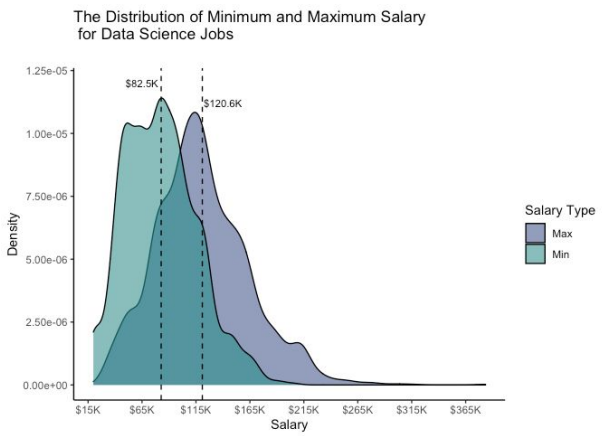


Figure 2a

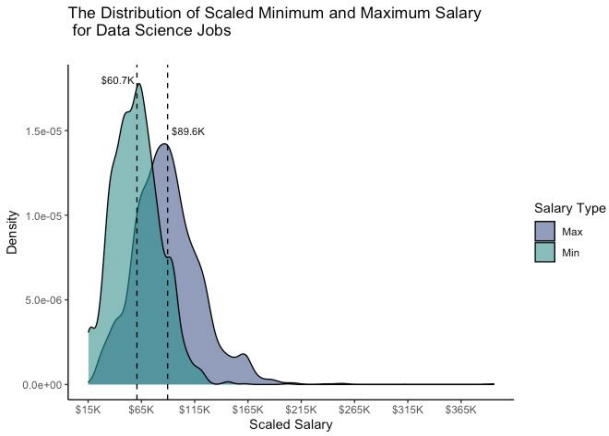


Figure 2b

We can see that from scaling the salary data (Figure 2b), the values have overall decreased and the spread of the distribution has decreased compared to unscaled salary ranges (Figure 2a). A large portion of the jobs listed in our data set are from cities with a higher cost of living than the average so it makes sense that the average maximum salary dropped from $120.6K to $89.6K and the average minimum salary dropped from $82.5K to $60.7K when scaling the data by COI. Now we can use these new scaled salary variables to compare salary ranges across the metro areas in our data set.

**Job Category Feature:**

When we began work with our data set, we saw that the unit of analysis was job openings, but what we really cared about was data science related job openings. Therefore, we began analyzing ways to see which jobs were data science related. Many of the job roles in our data set do not share the same title. This did not allow us to compare jobs based on job title so we decided to bucket the job titles into categories such as Data Analyst, Data Engineer, and Data Scientist. This allowed us to compare salaries for the categories and get locational information for the categories such as which city provides the most data science jobs. Thus, the purpose of the "job category" feature is to separate the jobs into data science and non-data science related jobs.

The first step in deriving the "job category" variable was to find the job titles that contain the data science categories of interest. The data science categories we began with were Data Engineer, Data Analyst, and Data Scientist. After getting the values for this variable, there were 1,814 NA values in our "job category" variable. We decided that there were too many NA values and that looking at other data science related job titles could reduce the number of NAs and increase the number of jobs that we use for our analysis. Thus, we also derived values for the new variable from the job description variable. We used the same three categories and found the job descriptions that contained these categories.

We were looking for ways to maximize the number of data science jobs in our data set so we then decided to include machine learning and modeling jobs. We could group these into a "Machine Learning Engineer" category. Then we included a "Statistician" job category because this is another area where recent data science grads can look for jobs.

Once we found the jobs that were data science related, we could begin adding these values into the "job category" variable. The values we chose were Data Engineer, Data Analyst,

Data Scientist, Machine Learning Engineer, Statistician and Other Analyst. The final NA value in the "job category" variable was 894; however, none of these jobs were related in any way and none of the jobs were data science related so we decided that our bucketing thus far was sufficient.

**Results:**

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

In order to answer our motivating question, we figured that it would be best to approach the job search process from the perspective of an applicant. We analyzed the following features in the subsequent order: data science roles, industry, company, Glassdoor rating, salary and conducted analysis on useful data science skills one should possess in order to be competitive in the job market.

**Data Science Roles Analysis**

The first thing that we think data science applicants will do when looking for a job is determine what job type they want. For example, our data set covers full-time, part-time, contractor, intern, temporary, and other job types. This graph gives the breakdown of data science job types by location. We are able to see the percent of each job type for each location.

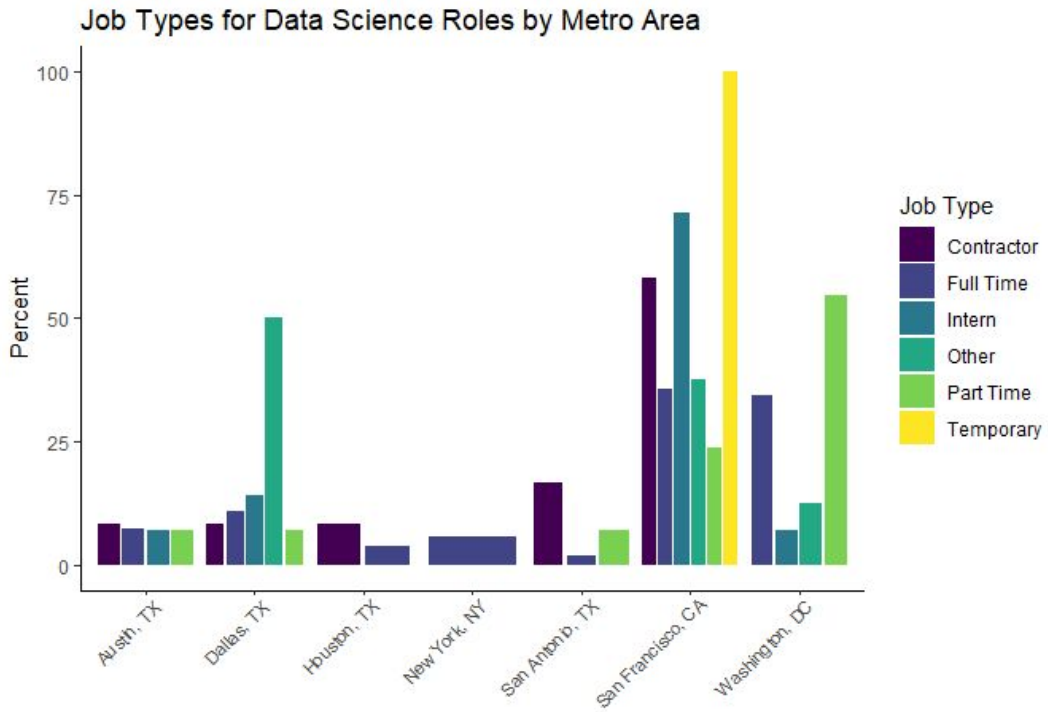


Figure 3

Figure 3 shows that San Francisco, CA is where a majority of the contractor, intern, and temporary job types are located. Thus, an applicant should consider this location if they are interested in these job types. San Francisco, CA and Washington, DC are where a majority of the

full-time positions are available so applicants looking for a full-time position will find more openings in these two locations than the other locations in our data set.

After the applicant has chosen their job type, they can begin to search for which data science job they prefer. Figure 4 below encompasses the job category and job location. It gives a count of the data science jobs for a location in our data set as well as how many of each data science job category are in each location.

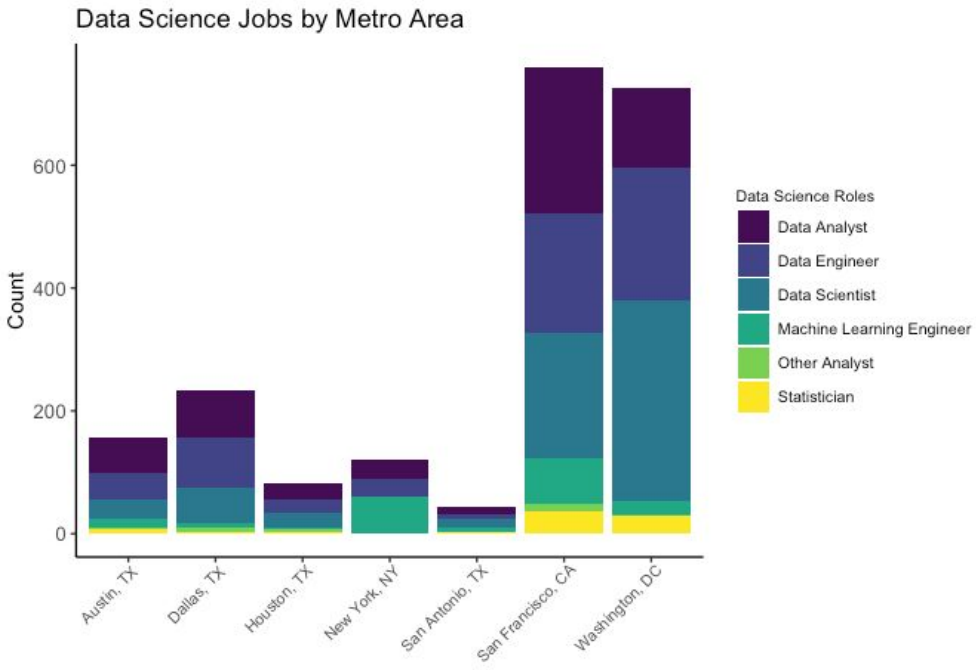


Figure 4

From Figure 4, we can determine that our data set has a majority of its jobs in San Francisco, CA and Washington, DC. Therefore, applicants ought to find the most data science jobs in these two locations and should consider applying to jobs in these locations as they offer more data science jobs than the other locations in our data set. Data Analyst, Data Engineer, and Data Scientist are also the top three data science jobs available in these two locations. Applicants should apply in these locations if they are interested in one of these three categories because of the abundance of positions as compared to other locations in our data set.

**Industry Analysis**

This column tells us about the various industries associated with data science related jobs. Candidates searching for a job may be more interested in finding a position in a specific industry. First, we analyzed the number of jobs available in the 5 industries with the most job openings according to the job category.

If a data science candidate is looking for a job, then based on this data, the job seeker can search for jobs in the 5 industries with the largest amount of data science related roles. From Figure 5, we see that San Francisco and Washington, DC have the maximum number of data science related jobs having open positions in all top 5 industries. Narrowing down to types of industries, Information Technology has the maximum number of jobs followed by Business

Services. Industries like Finance and Aerospace & Defense are limited to a few locations. With the information provided in this data set, if a person is looking for a job in Aerospace & Defense, then it's advised to look out for jobs in Washington, DC since it has the maximum number of jobs in this industry.

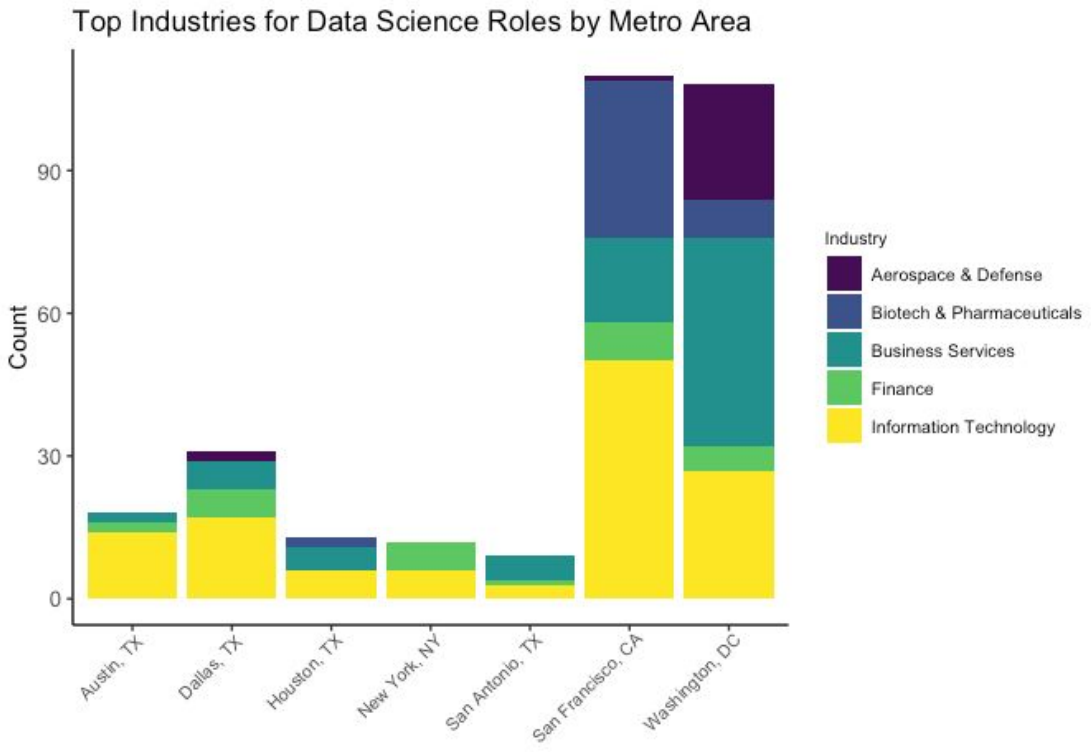


Figure 5

In Figure 6, we narrow down and look into different job categories in the top 5 industries. Here we see that if a candidate is looking for a data scientist position, it's recommended to look in Washington, DC since it has the most number of jobs. Irrespective of the type of industry, if a person is focused only towards the type of job, then this data will prove to be useful.

| Job Categories / Location | 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 |

Figure 6

Combining both the industry and the job category in Figure 7, we can see specifically which industry is looking to hire which data science roles. Here we can see Data Engineers are needed most in the Information Technology and Business Services industries. In addition, Statisticians are most commonly needed in the Biotech & Pharmaceuticals industry and Machine Learning Engineers are most likely to find a job in Information Technology. Data Analysts and Data Scientists are more flexible roles as they are found fairly evenly within most of the top 5 industries for data science job openings. This information can be useful to applicants with a specialization in a specific data science job category, by informing which industries most desire their services.

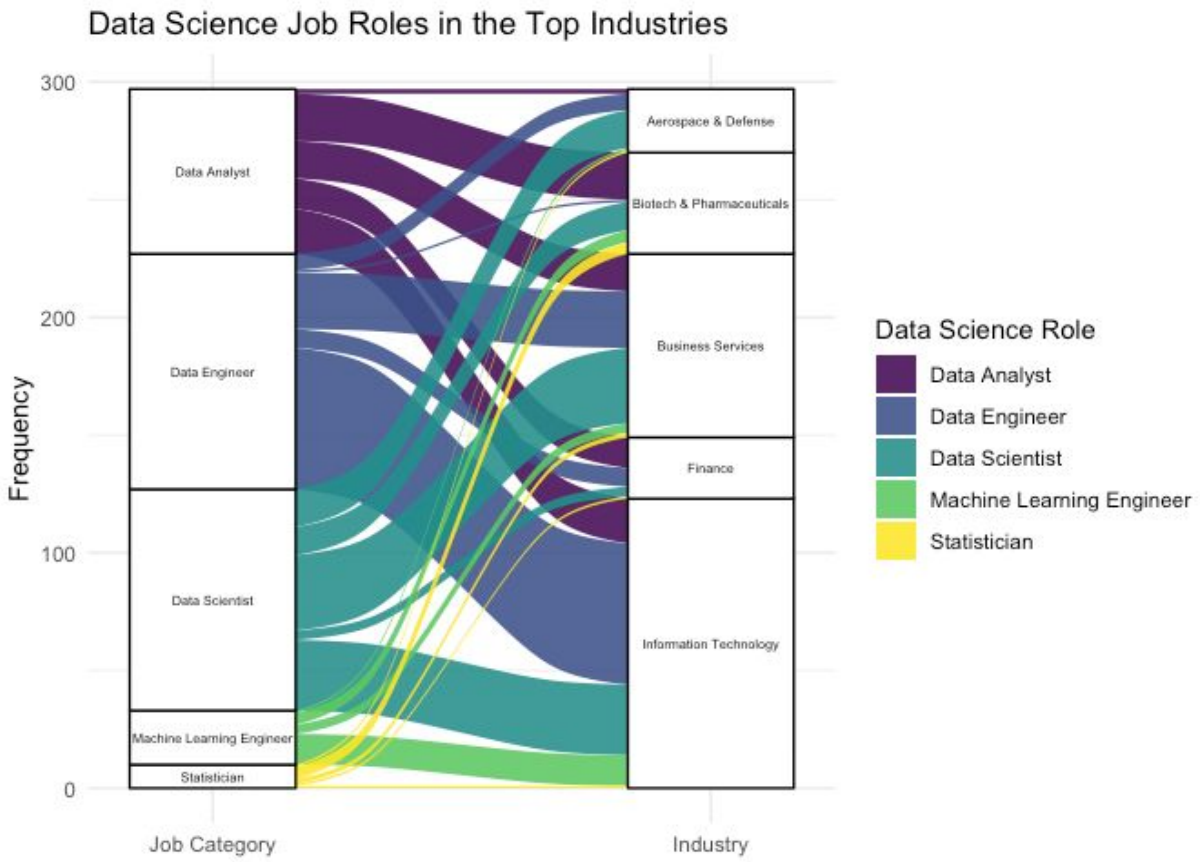


Figure 7

**Company Analysis**

After narrowing down the job roles, industries, and locations of companies for which to apply, job applicants often research individual companies offering jobs that match their search criteria. Below, we take a look at Figure 8 in order to gain a better sense of the makeup of some of the most popular data science destination companies. There are several companies with only one type of job posting: both Kingdom Associates and TEECOM are destinations for those interested in roles as Machine Learning Engineers, while Adepar and National Debt Relief only have openings for Data Engineers and Data Analysts, respectively. Genetech is unique from the rest of the popular data science companies in that it is employing across all categories of data science jobs. This plot can inform potential applicants of which company their specific area of data science expertise is most in demand.

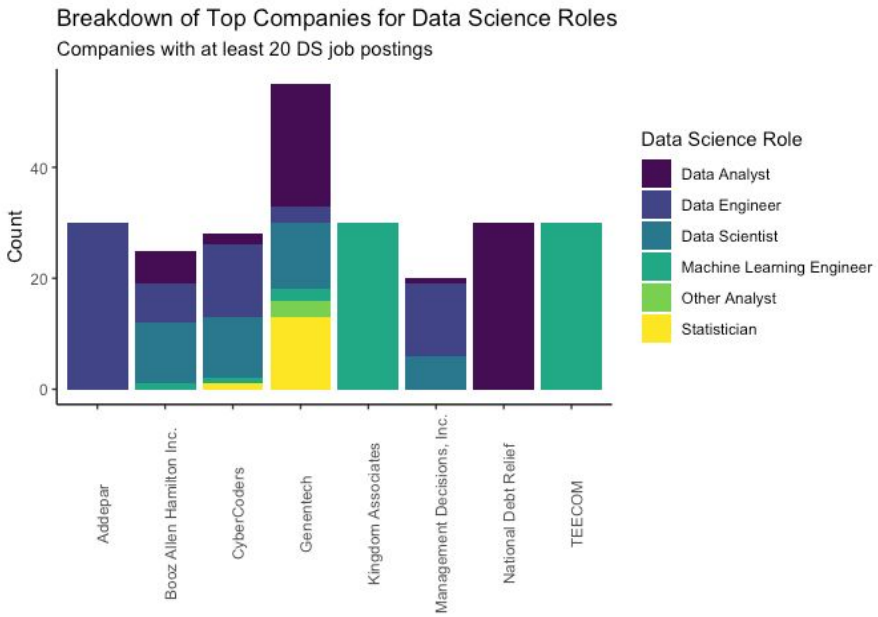


Figure 8

**Glassdoor Rating Analysis**

Additionally, job applicants may also be interested in the rating of the company that they are applying to, in order to gain a better idea of the work environment. First, we begin by observing the ratings of the most popular data science companies that were analyzed in the above company section. In Figure 9, we see that the ratings are all relatively consistent across these companies with the exception of Management Decisions, Inc., which holds a much lower rating. Glassdoor's company rating system could be the difference between where a job applicant chooses to work - all else equal. Both CyberCoders and Management Decisions, Inc. are in the business services industry and have the vast majority of their job postings in San Francisco. Furthermore, both hold a similar distribution of job categories. An applicant interested in both of these companies may be more likely to accept an offer from CyberCoders based on the significantly higher company rating.

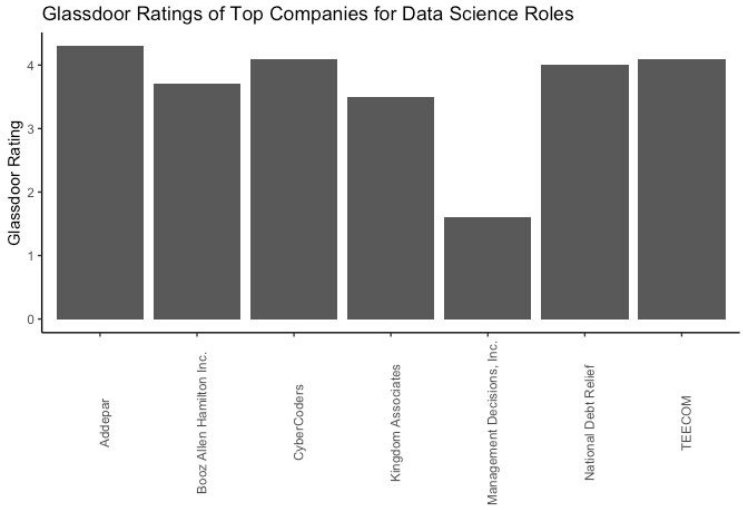


Figure 9

There may be applicants more interested in a particular industry as opposed to a particular company. In Figure 10, we breakdown the densities of company Glassdoor ratings for the top industries. The distributions all appear pretty similar with the exception of a high concentration just above 3.5 for Aerospace and Defense. Information Technology has the highest rating peak among the five largest industries, while Biotech and Pharmaceuticals has the highest number of companies with ratings of 5. We see that IT may be a slightly more appealing industry to job applicants from the company rating perspective. Applicants should take note of the fact that each of the most popular data science job industries have average company ratings where employees are 'satisfied' or 'very satisfied', which bodes well for their potential at-work experience if entering one of these industries.

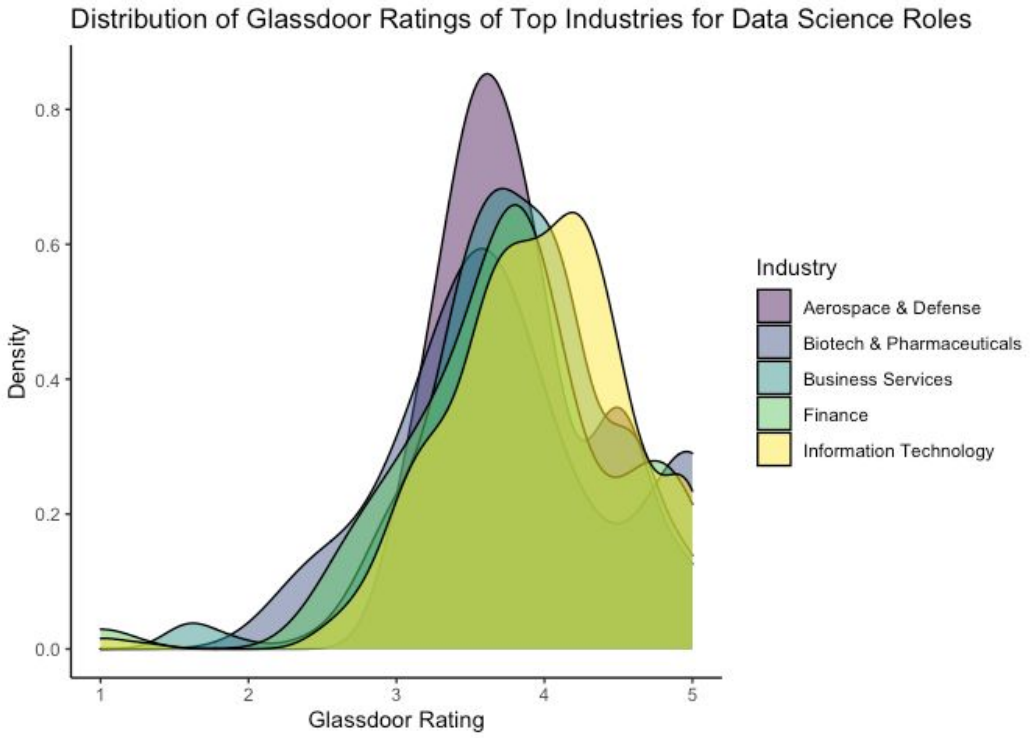


Figure 10

For applicants more sensitive to location as opposed to industry, we examine the Glassdoor ratings of companies located in each metro area (Figure 11). While the distributions of most areas appear similar, New York is clearly unique among the observed metro areas. NYC has all ratings concentrated between 3.5 and ~4.25. This may signal that applicants can be pretty certain about what life at the company is like in New York as opposed to other metro areas with greater variability in company ratings.

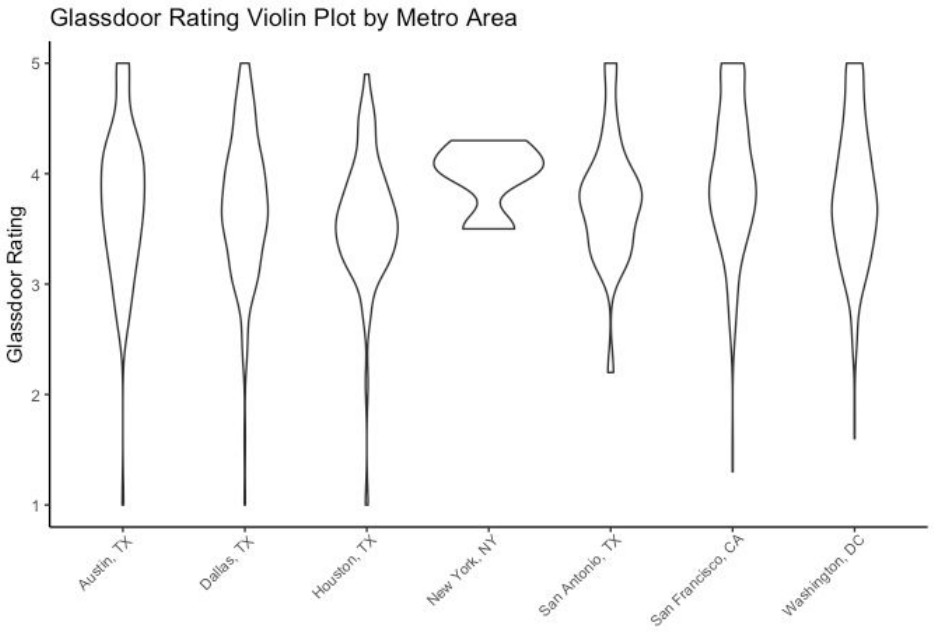


Figure 11

**Salary Analysis**

In addition to comparing job roles, industries and companies to work for, candidates often compare salaries. Although we would like to think the fulfilment of having a job that we enjoy is the most important factor in choosing a job, we need to be financially stable to support ourselves and our families. Thus, a competitive salary is a very important factor to consider. Our first question was about which metro location, regardless of the type of role, has a better overall salary range for data science related jobs. Figure 12a shows the average maximum and minimum salaries by metro location. We can see that San Francisco has the highest overall salary range for data science related jobs. However, this graph is not quite accurate in comparing metro locations due to differences in cost of living.

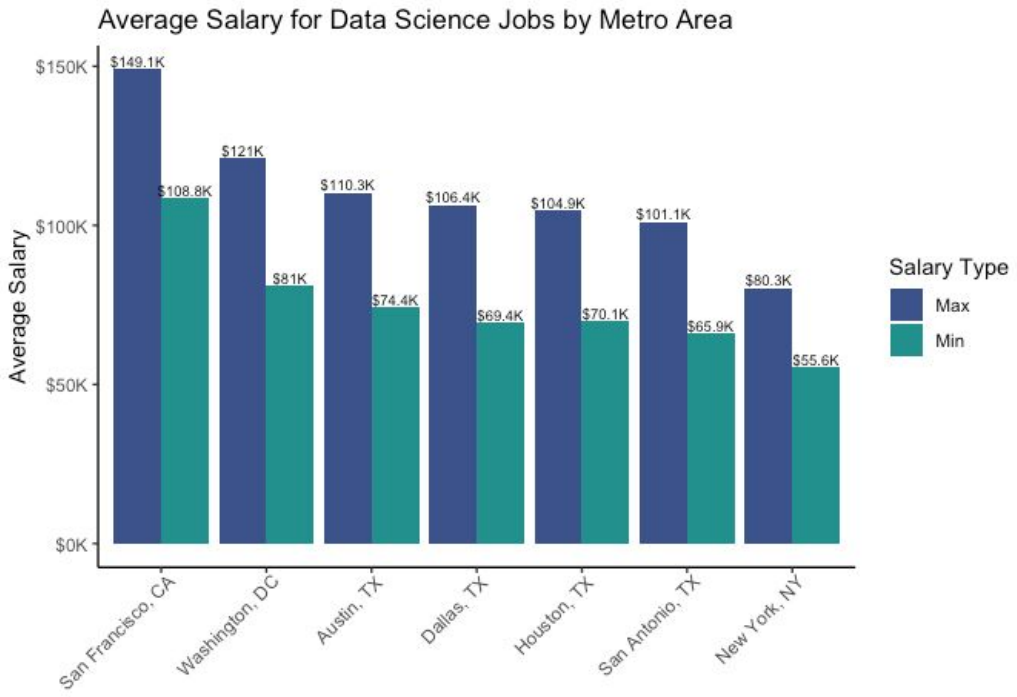


Figure 12a

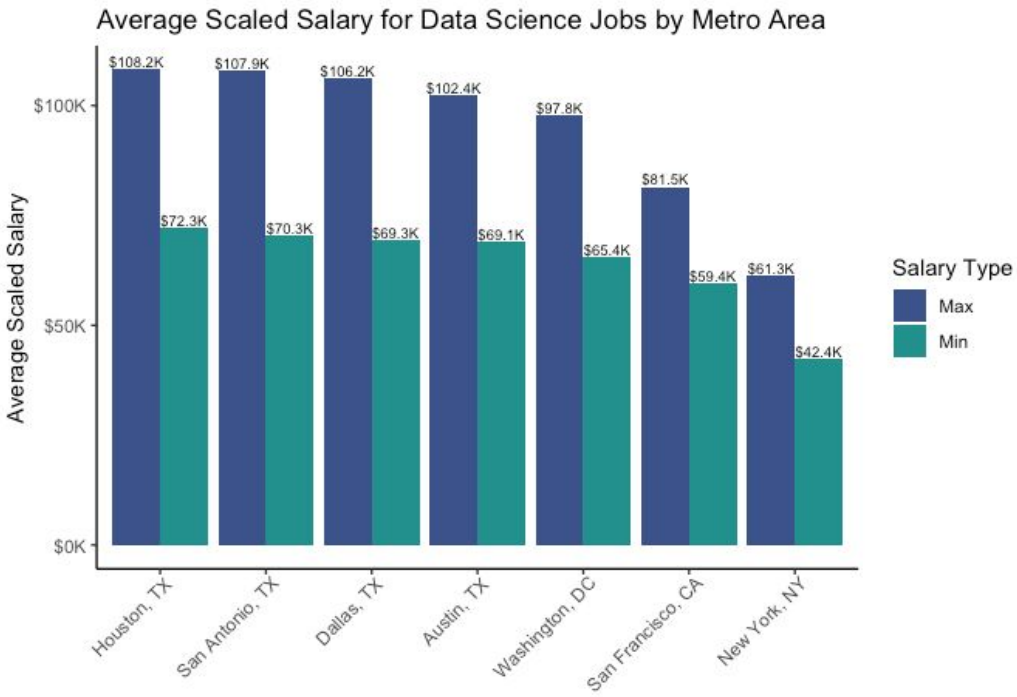


Figure 12b

We can see in Figure 12b that when we scale salary by cost of living, salaries in all the metro areas of Texas are higher. This means that although overall salaries are higher in San

Francisco and Washington, DC, a salary in large metro areas of Texas will go further in terms of daily expenses and salary compensation.

Different roles within data science require different skills and experience. Therefore, we wanted to understand which roles typically have higher salaries. As we can see below in Figure 13, the distribution of minimum and maximum salaries is similar between different data science roles. They are all approximately normally distributed with a slight right skew.

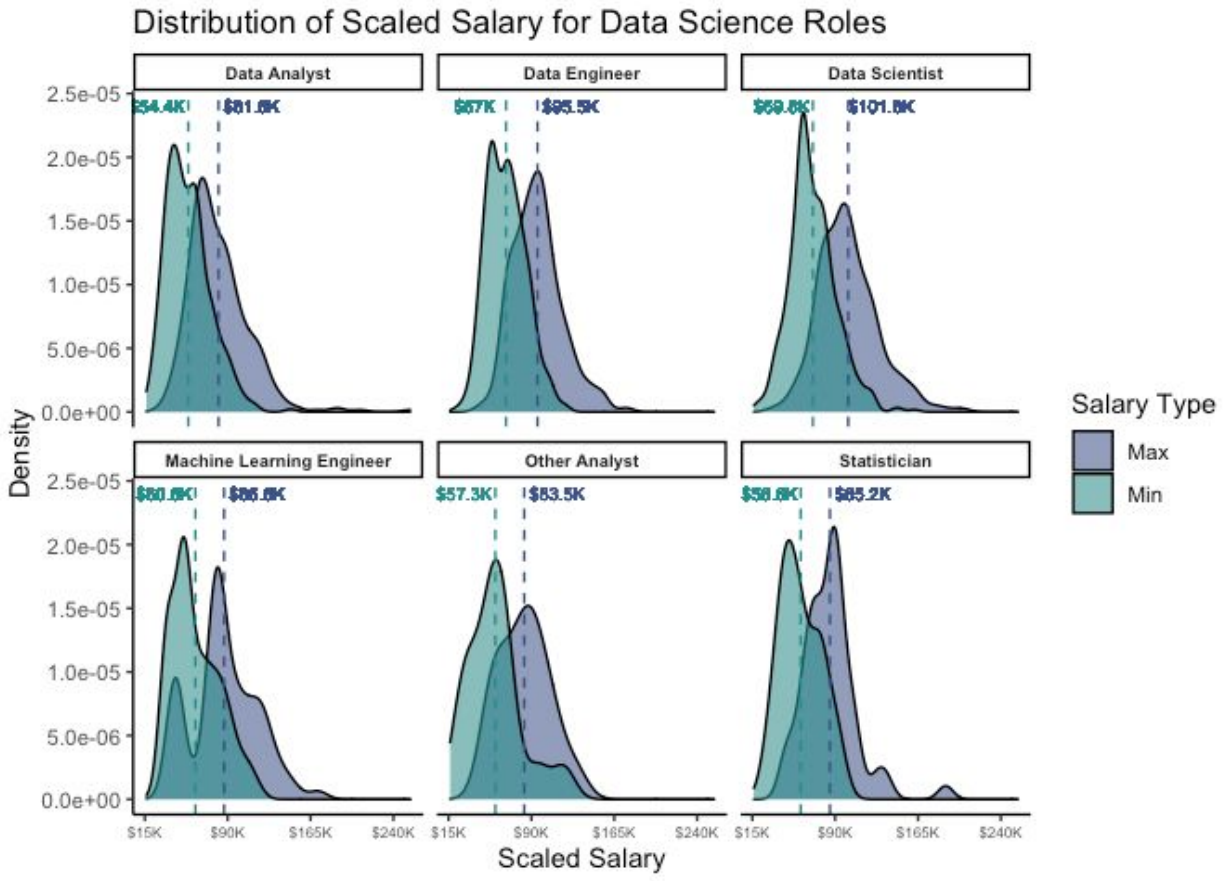


Figure 13

In looking at the average scaled salary for different data science roles (Figure 14), there is a clear trend. A Data Scientist makes the most, followed by the engineering roles, Data Engineer and Machine Learning Engineer, the Statistician and ending with the analysts, Data Analyst and Other Analyst. This means candidates might want to look at Data scientist and engineering positions over analyst positions if they have the applicable skills and experience.

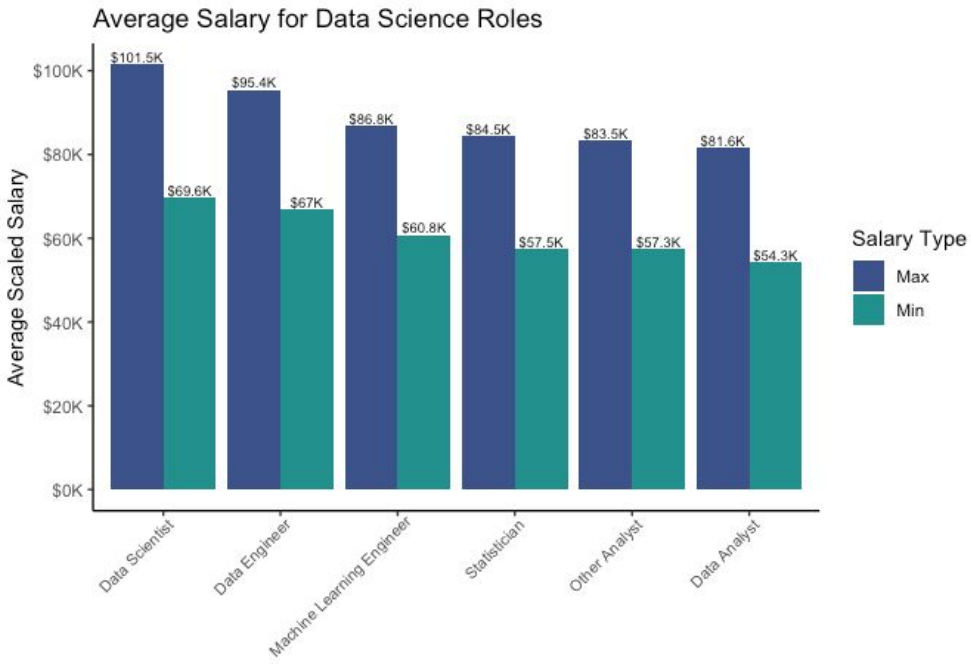


Figure 14

Even further, we were curious about which industries had the highest minimum and maximum average scaled salaries to see if there were better industries for individuals to look into depending on the salary range they wanted to pursue.

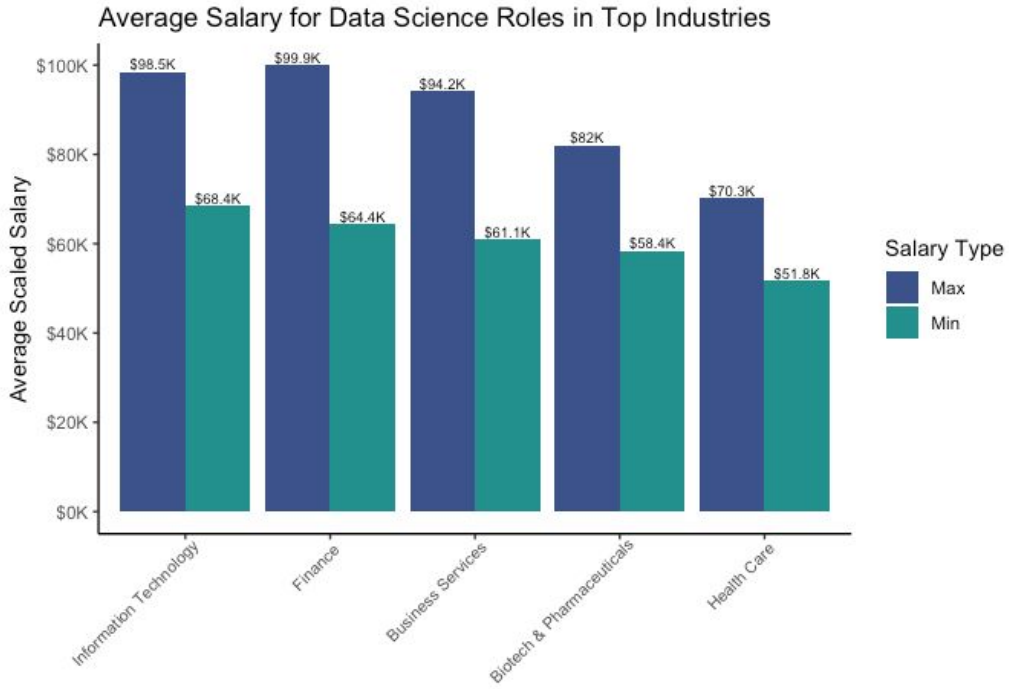


Figure 15

In looking at the top 5 industries within our data (Figure 15), Information Technology and Finance seem to have the highest average salary ranges followed by Business Services, Biotech & Pharmaceuticals and Health Care. Overall, when looking at average salaries for data science job openings, the best options for an optimal salary are to look in the four big metro areas of Texas, search for data scientist and data engineering roles and look at openings in the Information Technology and Finance sectors. Although certain areas, roles and industries have higher salaries than others, we can see that a candidate in the data science field can expect a pretty good salary no matter what their circumstances.

**Job Skills Analysis**

Once the applicant has figured out all of the above information on where to look and at what kinds of roles, industries, and companies to look for, he/she should find out important skills that would make his/her resume stand out. Our study consists of a text analysis that analyzes the job description of each job opening and lists the most commonly used words or phrases in these job descriptions. We hoped to find hard skills such as coding skills like R or Python that the applicant could ensure they have applicable experience using. Figure 16a shows commonly used words in job descriptions for data science roles while Figure 16b displays the commonly used two-word phrases in job descriptions.

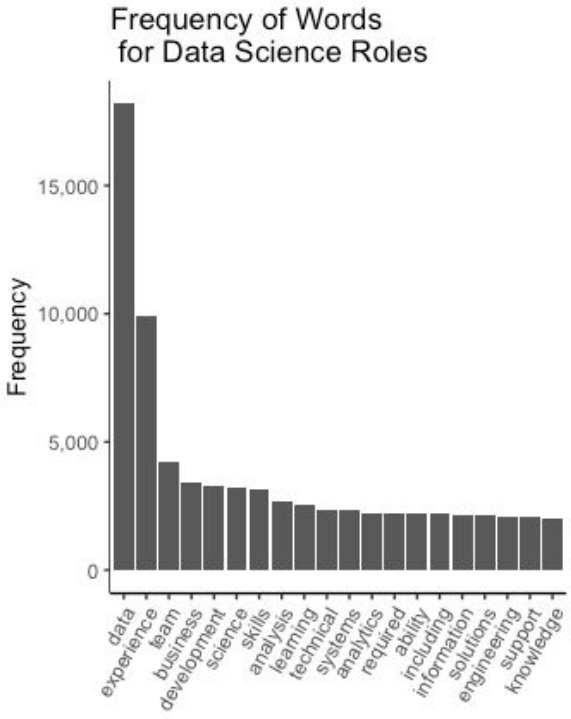


Figure 16a

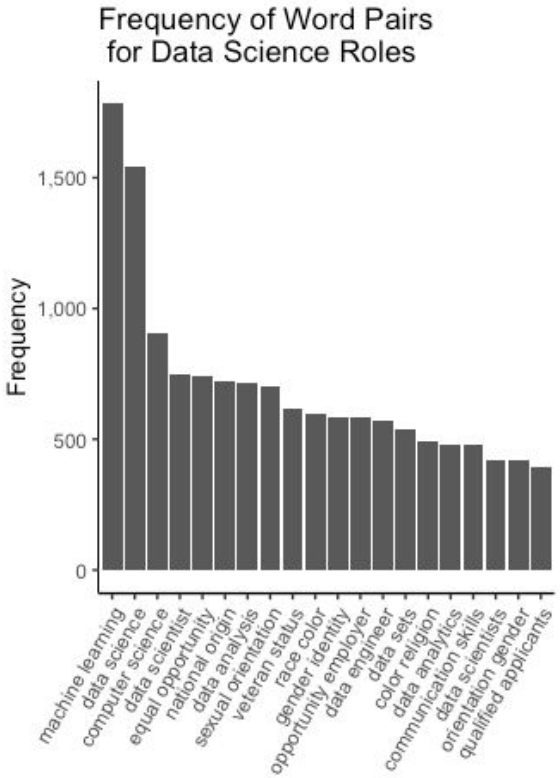


Figure 16b

Figure 16a does not display any hard skills necessary for the jobs; however, it provides soft skills such as ‘analysis’ and ‘team’. Thus, applicants ought to be able to work with others and be able to think analytically. Figure 16b also does not reveal any hard skills; however, this figure shows popular data science areas such as ‘machine learning’, ‘data scientist’, ‘data

analytics' and 'data engineer'. It also displays communication skills that the applicant ought to have. We know that to be successful in a data science field one must have some coding and data knowledge and experience, but these results highlight how important soft skills are as well in obtaining a job in the field of data science.

**Conclusions:**

Based on our analysis, we can make several recommendations to data science job applicants in order to find a job in the field. Data science appears to be booming in both the San Francisco Bay area and Washington, DC, providing options for applicants on both sides of the country. Additionally, these specific data science roles hold the most opportunity for applicants to land a job: Data Analyst, Data Engineer, and Data Scientist. Across industries, Information Technology provides the highest number of jobs and heavily recruits Data Engineers. Furthermore, IT workers are happiest among the top data science industries and are among the best compensated in terms of salary. If job satisfaction is important to an applicant, New York might be their best option as it holds the highest Glassdoor rating across our observed metro areas. While gross overall salaries are lower in Texas metro areas, when scaled by cost of living they lead our observed set of metro areas in salary compensation. No matter what aspects of a job may appeal to an applicant, they must possess the skills required to land the job offer and perform the job effectively. Soft skills such as communication and teamwork are key traits that companies look for in their data science applicants and can be a great compliment to one's technical skills.

**Limitations/Next Steps:**

When conducting our analysis, we came across several issues which may cause limitations in our analysis. The data we used was scraped from the Glassdoor website by a program written by a Kaggle user. Upon closer inspection of the data, we found that it was scraped with a combination of python packages, selenium and beautifulsoup. The selenium package allows for remote navigation of web pages. The methodology as to how the final Kaggle data was collected was made available on the user's GitHub page (https://github.com/Atharva-Phatak/Glassdoor-Jobs_Data-Analysis ).

In our analysis, we found that the counts for many groups of job postings (job categories or companies for example) often were multiples of 30. Upon closer inspection, we discovered that there were 30 observations displayed on a single page of job postings. This implies that the methods used to scrape the base dataset may not have collected data on all job postings, as even multiples of 30 for several companies does not seem highly likely. Furthermore, the seemingly low data science job numbers of New York as compared to San Francisco and Washington, DC may imply further issues in the data collection process. If given more time, we would look into scraping the job posting data ourselves from Glassdoor, in order to combat the issues that we found with the underlying data. This improvement would also strengthen the validity of our findings.

Additionally in looking at Figure 17, we can see there were inconsistencies in the counts of jobs posted for specific dates. While this may be due to a coincidence of companies posting jobs on specific days of the week or month, there is also the very real possibility that data was

only scraped on certain dates. If given more time, we would be able to scrape our own job posting data to ensure that there was a more uniform distribution of job postings to ensure the validity of our findings. This would also allow for time series analysis of job postings as an extension to our initial analysis, which could help inform applicants of the best time of year to apply for data science jobs.

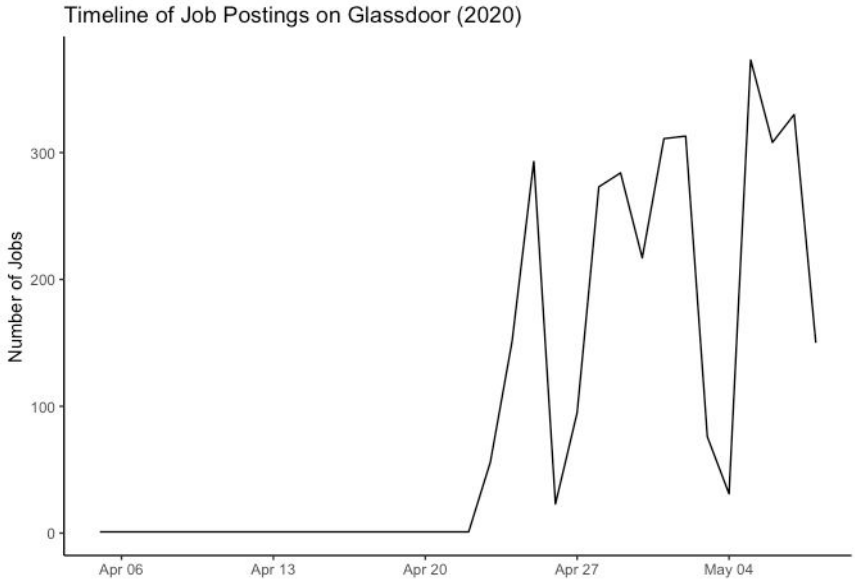


Figure 17

With increased job posting data from separate areas of the country, we could extend our analysis to compare and contrast different regions of the country to each other (Northeast vs Southeast, etc.). With a wider range of locations, our findings would be applicable to a greater amount of individuals on the job market.

An additional limitation for our data is the nature of company ratings provided by Glassdoor. The job ratings data is representative of the entire company, not for a specific department of a company. Our analysis presented findings based on company ratings which may not be entirely accurate for the subset of job postings that we are interested in within these companies. Ratings for a given company may be skewed by extremely positive or negative reviews from a company department that does not have to do with data science, such as customer service. A recommendation for Glassdoor that may help aid job seekers, companies, and also potential future analysis would be to have individual department ratings within companies, as this would remove influence by other entities within the organization.

Despite these limitations, our analysis can be used to provide insights for data science job applicants as to where they should start when conducting their job search as well as what to expect during their search. Our findings can also provide recommendations to Glassdoor as to how they might be able to improve their user experience - by updating company rating to be department-specific. Our ideas for improvement upon this project would provide insights for a larger population of job applicants and also improve the validity of our findings.