

Summer Data Competition

Team: Yaqiong (Juno) Cao, Dexter Nguyen, Xinying (Silvia) Sun



How to Navigate Analytics Job Search During COVID-19

October 25, 2020

Executive Summary

Are you an analytics master student who is seeking a job and have no idea what the job market looks like, especially during COVID-19? Our project aims to analyze the current online job posting situation that can help students better understand the current job market, find the most suitable job positions, and make the best preparation for these jobs. After integrating the latest data from authorized websites like LinkedIn, Glassdoor, and company official websites, we generated several insights from our multi-dimensional analysis. We encourage master students to develop their communication and problem-solving skills besides technical skills. Students should pay more attention to job postings like Financial Analyst, Data Engineer, and Data Scientist. They should also consider narrowing their search to big cities with high ratings and focusing on tech and financial companies like Amazon, which has an incredible hiring growth rate this year.

Data Integration

Using different data collection methods, including web scraping in Python and data merging in SQL and R, we consolidated data from a variety of sources, as shown in our relational schema (Figure 1).

Job Data Table & Company Workforce Table

Using keywords via scraping at [LinkedIn](#), we first collected analytics and data-oriented job posts from the top 10 companies (in revenue) of each of the top five largest industries: Information Technology, Financial Services & Investment Banking, Retail, Fast-Moving Consumer Goods (FMCG), and Media & Entertainment. Next, we limited our data to two months: August and September, a period right after graduation, which we believe will effectively reflect the job market's situation on the LinkedIn platform. We then extracted primary information such as companies and associated industries, job post's date, job titles, and job descriptions that include skills and requirements for each position. We also obtained data relating to each selected company's current hiring situation for the past 12 months to provide evidence support for further deployment. To clean the job data, we focused on three tasks. First, we removed duplicate observations resulted from multiple scraping runs. Second, we researched each column/variable, doing corrections for city names of job locations and industry classification. Last, we spent most of the time cleaning job title and allocated them into four title levels: Associate, Analyst/Engineer, Manager, Scientist/Researcher. This step would provide "good" data for a separate analysis of job titles.

Job Skills Table

For the next part, we created a subset from the LinkedIn original data set by extracting information from the job description column. This column contains skills and requirements for each analytics job. Based on our original data set, we selected 15 positions that best fit the capability and desire of analytics master graduates, including MQM students. Also, conducting text analysis in R for job descriptions, we identified the top 27 most popular skills, which means they appear in the basic/preferred qualifications and requirements of job postings most of the time. This process resulted in a subset data, including company names, job titles, and a list of skills required for each job title.

Company Info Table & Company Rating Table

Based on the raw company data that showed active hiring activities on LinkedIn, we dove deep into the financial information of these 25 companies. We collected the data about the size and market capitalization from [MorningStar](#) and analyzed the second quarter fillings of 2020 and 2019 from the investor section of each company's official website. The rating for each company was, meanwhile, obtained from [Glassdoor](#).

Job Locations Table

We built a location table with location column from the job data (a city where each job is located). Then we added such information as state, longitude, and latitude based on Google Map and GitHub

sources. Using R, we merged our raw data and the cities information table to create a new location table with unique location values that could be used in mapping visualization in the following analysis.

Deployment

Overall job postings in the US

We first analyzed the US overall job postings by using R mapping. According to the mapping (Figure 2), it is evident that companies located in east and west coasts are hiring more. Our data proves that big cities also provide more job opportunities. When we analyzed the data in-depth on the regional level (Figure 3,4, and 5), we found that the job posting of GA and NY in east coasts, WA, and CA in west coasts, MN, and IL in the Midwest is relatively high in other places in the US. It is reasonable since the fast-growing cities provide more employment opportunities, and higher-salary jobs appeal to many employees. When looking at specific cities with the largest number of job postings like Seattle, Atlanta, and New York, we found that almost 90% of job postings in Seattle are from Amazon, 70% of the job postings in Atlanta are from The Home Depot. Only New York has a different pattern when the local job postings are from a diversity of companies. This finding verifies our assumption that looking for an analytics job is easier in big cities. We also observed another insight into the hiring situation based on the number of jobs posted through the week. Diving deep into two weeks: September 14 – 30, from Figure 8, we saw the number of jobs posted reached the peak on Wednesday and Thursday, while recruiters were less likely to post new jobs during the weekend from Friday to Monday.

Hiring situation at the industry and company levels

After having a general idea of the US job market in the last two months, we wanted to learn more about the hiring situation at the industry and company levels. According to Figure 6, Information Technology & Services companies are offering more job opportunities, mostly by Amazon and IBM. There are about 50% of the job opportunities provided by the tech industry. Moreover, the job opportunities offered by Amazon account for 1/3 of the total. Meanwhile, companies in FMCG and Media & Entertainment industries like Johnson & Johnson and AT&T are way behind in hiring (Figure 7). About 1% of job opportunities are provided by the Media & Entertainment industry, and about 2% provided by FMCG. Conducting external research, we found that these industries and companies have performed in reverse directions under the impact of COVID-19, which could result in their hiring decisions.

Rating and Revenue analysis overall and within the industry

After learning about the hiring conditions, we are curious about how companies are running their business under COVID-19. Also, we analyzed the company ratings using mapping. It is reasonable that retail pharmacy companies have higher revenue recognized because of the volume of goods sold, like the case of CVS Health (Figure 11). Under the effect of COVID (Figure 12), media companies like The Walt Disney and Omnicom Media Group decreased their revenue significantly. It surprised us that Johnson & Johnson's revenue decreased when we expected it to increase revenue. By doing more research on their current 10Q, we found that a poor sale performance of the medical device, which dropped 33.9% compared to 2019Q2, is the main reason to reduce the total revenue even though the pharmaceutical sale segment increased under the effect of COVID as expected. Meanwhile, Amazon saw a massive expansion in revenue, matching its hiring growth, since their online and cloud-based business model benefits from the current pandemic. Interestingly, Truist increased its revenue by 80% as a result of the merger of BB&T and SunTrust Bank. In terms of rating, as most tech companies are located in California or Seattle, the West coast companies have a relatively higher rating than the other US (Figure 13). Employees more favor salary, benefits, company culture, and working conditions of tech companies. Therefore, we encourage analytics master students, including MQM, to focus on job search in the technology industry, financial services, and investment banking industry based on revenue and rating analysis.

Job postings at the job title level

In terms of job postings at the title levels, we selected Analyst/Engineer and Scientist/Researcher levels to do more analysis (Figure 14). Regarding Analyst/Engineer level, we saw the dominance of Financial Analyst and Data Engineer. Meanwhile, Data Scientist has been still the most popular job demanded in the Scientist/Researcher market. To leverage our insights, we picked the top 4 companies from different industries to learn more about which job titles they are looking for most. Figure 15 provided us an interesting inference that each company has a distinct demand for their top analytics job. Respectively, Financial Analyst, Data Engineer, Data Analyst, and Business Analyst are the most popular by Amazon, CVS Health, JPMorgan Chase, and The Home Depot. These companies are currently different in terms of revenue performance and business situation, as mentioned in the previous analysis.

Job skills for each title

According to Figure 16, we found that Communication, Problem-solving, Data Visualization, Data Modeling, and Critical Thinking are the top five skills for every position and every company. An analytics master curriculum like the MQM program with a good balance between general/soft skills and technical skills will definitely make a difference in a job search. Diving deep into each job title, we found that the Financial Analyst and Business Analyst requirements are somehow similar (Figure 17) since they both include Data Visualization, Microsoft office, Data Modeling, and SQL. For the Data Engineer and Data Analyst position (Figure 18), we saw a different pattern. While the Data Engineer jobs require such database skills as Excel and SQL, the Data Analyst jobs prefer more advanced programming languages like Python and Data Modeling/Machine Learning techniques. Interestingly, Data Visualization is very demanded for the Data Analyst position. Regarding Data Scientist and Data Researcher (Figure 19), they both ask for a high level of mastering the most popular languages like Python, SQL, and R. In addition, we found that Data Scientist, Business Analyst, and Data Engineer have more requirements about hard skills such as SQL, Python, R. Also, Analytics Manager usually requires 2+ years of work experience (Figure 20, 21, 22, and 23). To sum up, every position requires soft skills, but Business Analyst, Financial Analyst, Data Engineer, Data Scientist, and Data Researcher will demand more technical language skills.

Implementation

Our project aims to help analytics master students, including MQM, have a better understanding of the current context of recruitment and make sufficient preparation while seeking employment. Although we believe that our analysis is useful and efficient, it does have some limitations. First, because the job postings decreased over time by recruiters, the job data from August and September may not perfectly reflect the job market. Second, we could not have insights into the job market in some states in the US since no job posting was recorded (showed in our mapping analysis Figure 2). To verify our assumption that these areas are negatively affected by the COVID-19, we can collect the data from other hiring websites. Third, we found that Audit and Consulting firms like BIG4, McKinsey, BCG, and Bain don't have public financial information, which limited our quantitative analysis for the company revenue part. Last, using keyword in scraping job titles, we might ignore some positions that don't include "analytics" in their titles. These jobs are, however, data-oriented and should be classified as analytics jobs. For example, companies that don't have a specific audit analytics position but require auditors to learn Python and Tableau. In that case, we should try other data collection methods to obtain a wider scale of relevant data and improve analysis accuracy.

Appendix

Figure 1: Relational Schema

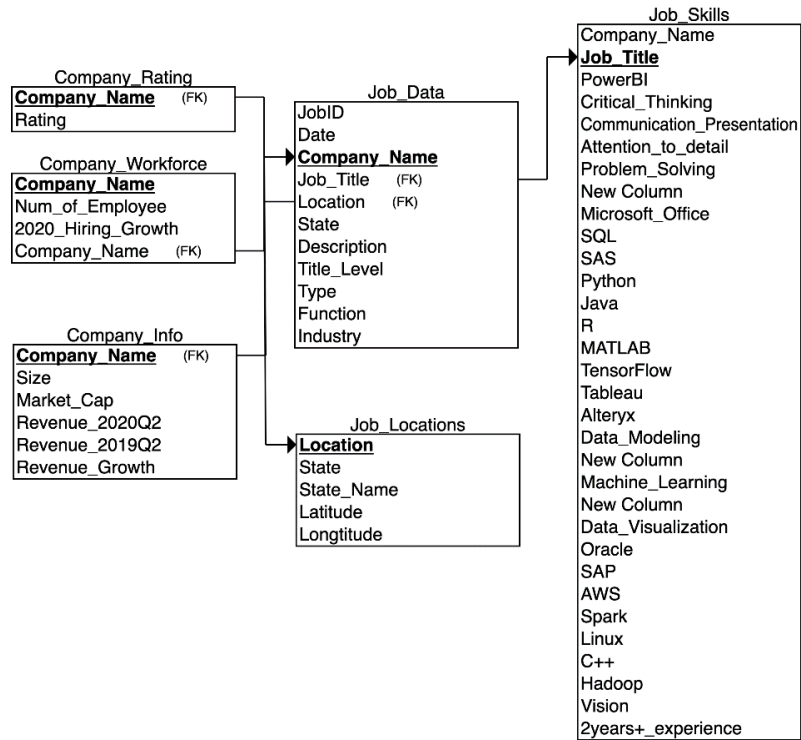


Figure 2: Job posts across the US

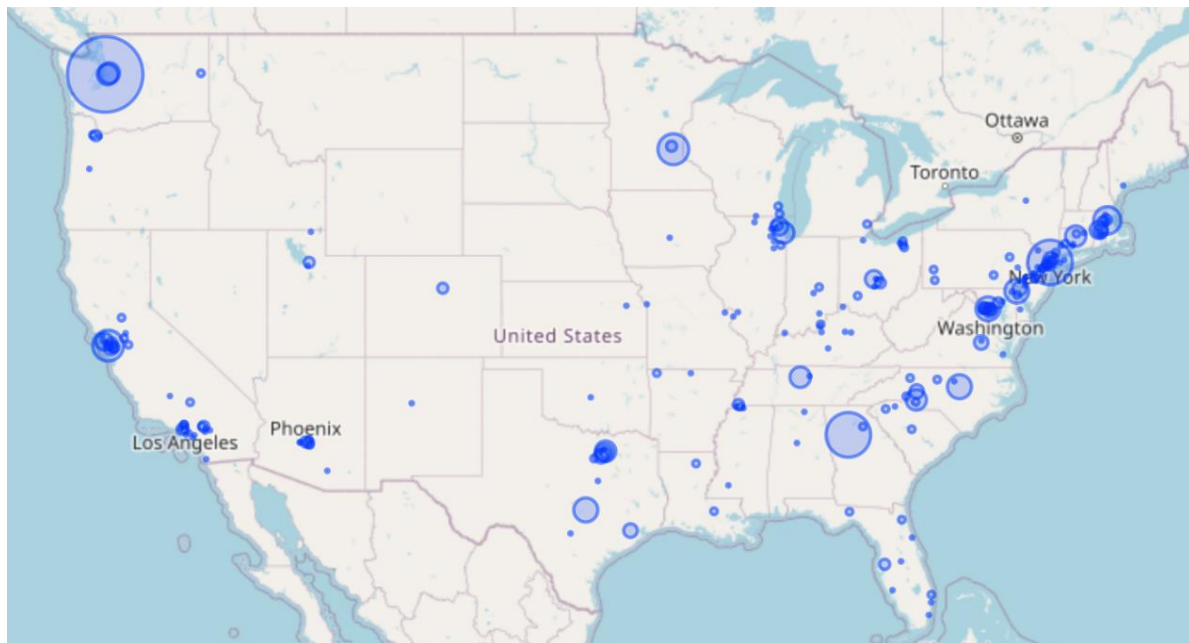


Figure 3: Job posts on the East Coast

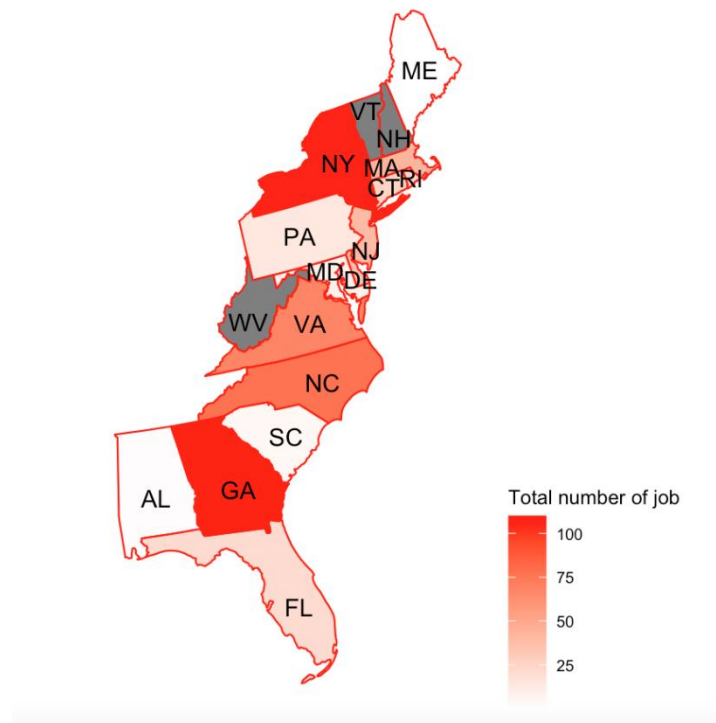


Figure 4: Job posts on the Mid-West

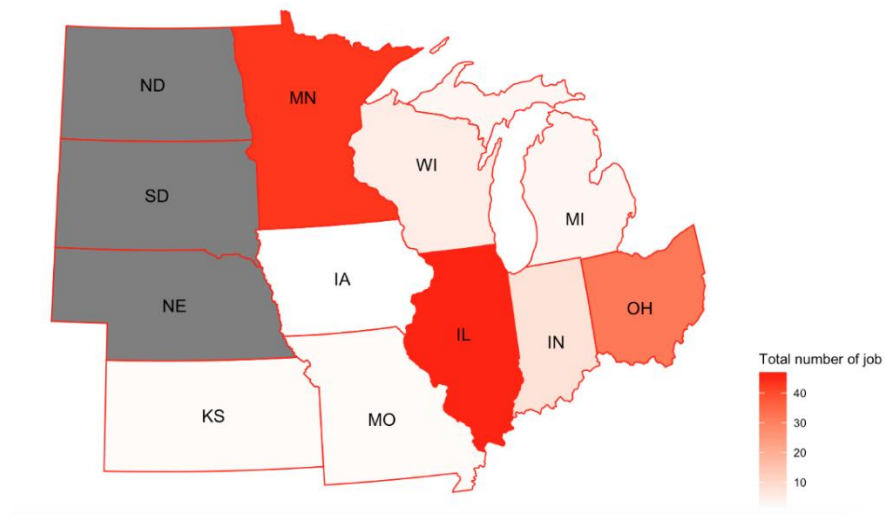


Figure 5: Job posts on the West Coast

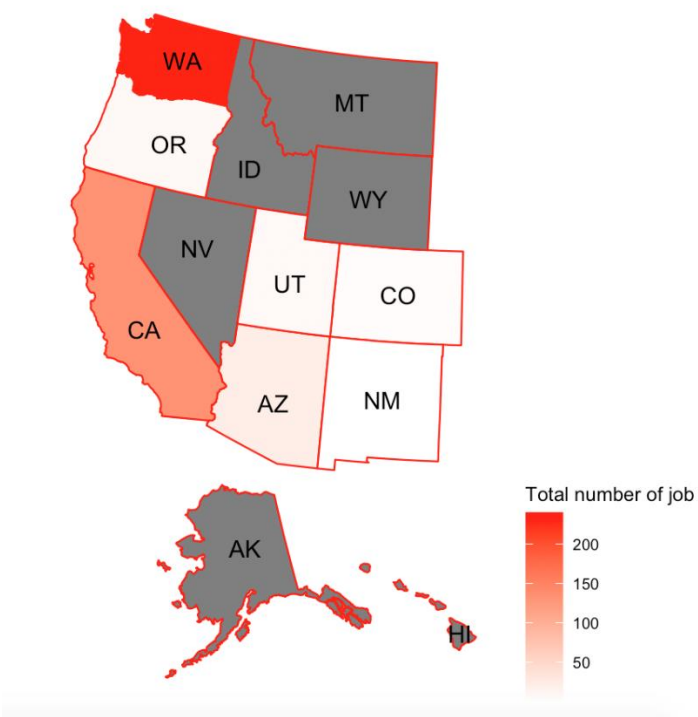


Figure 6: Job posts by industry

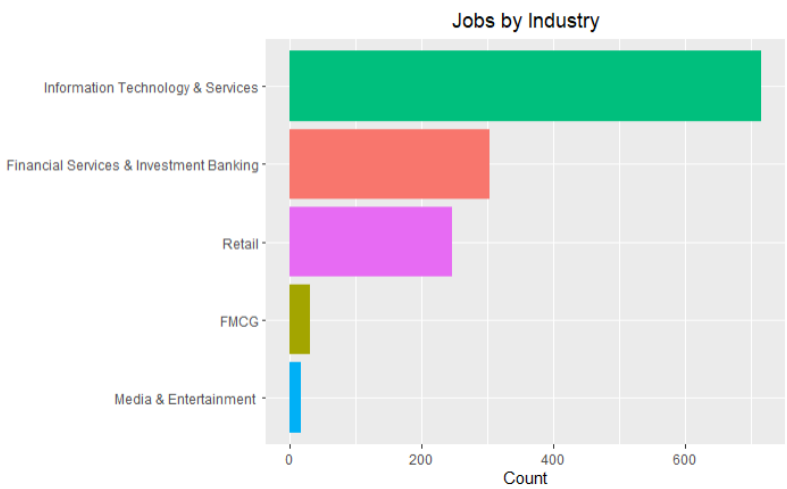


Figure 7: Job posts by industry and company

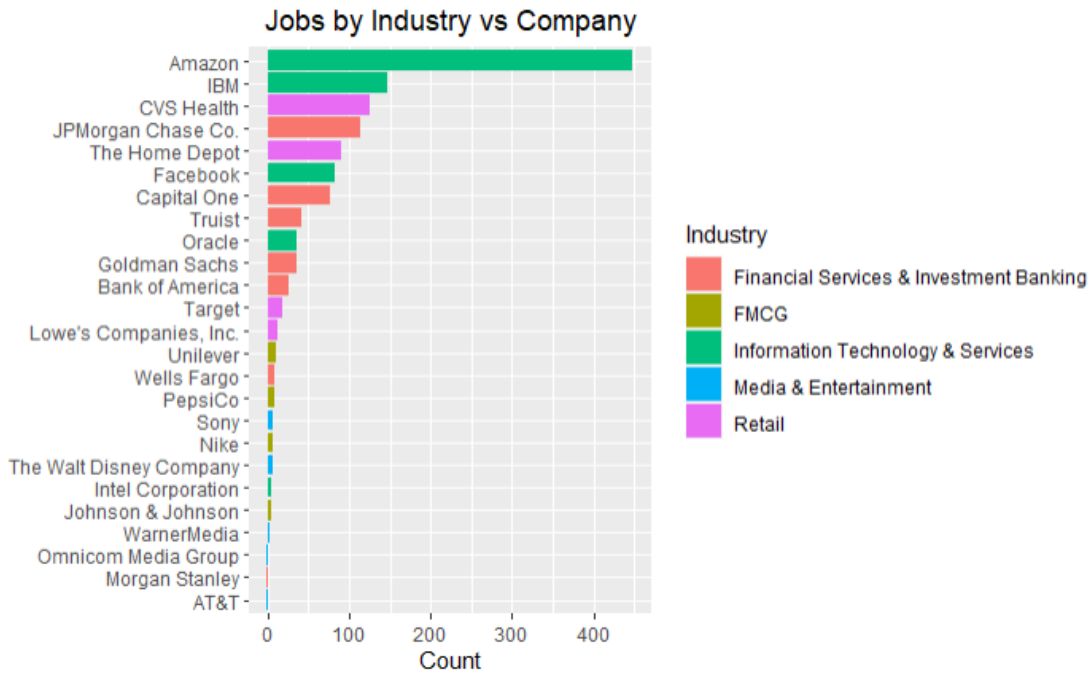


Figure 8: Job posts from September 14 to September 30

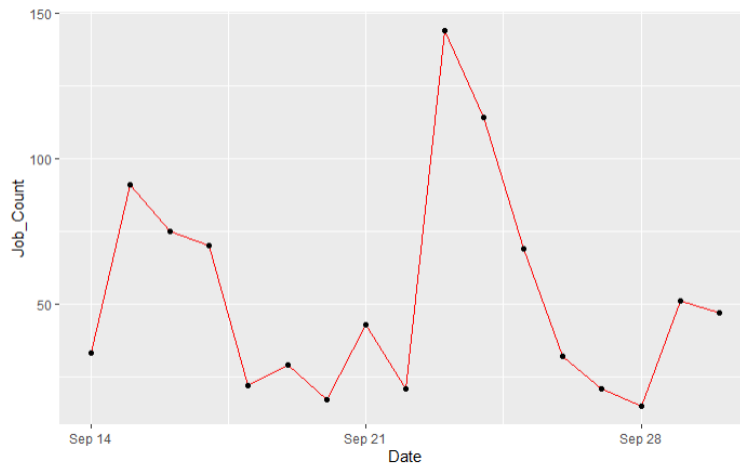


Figure 9: Number of employees by company and industry

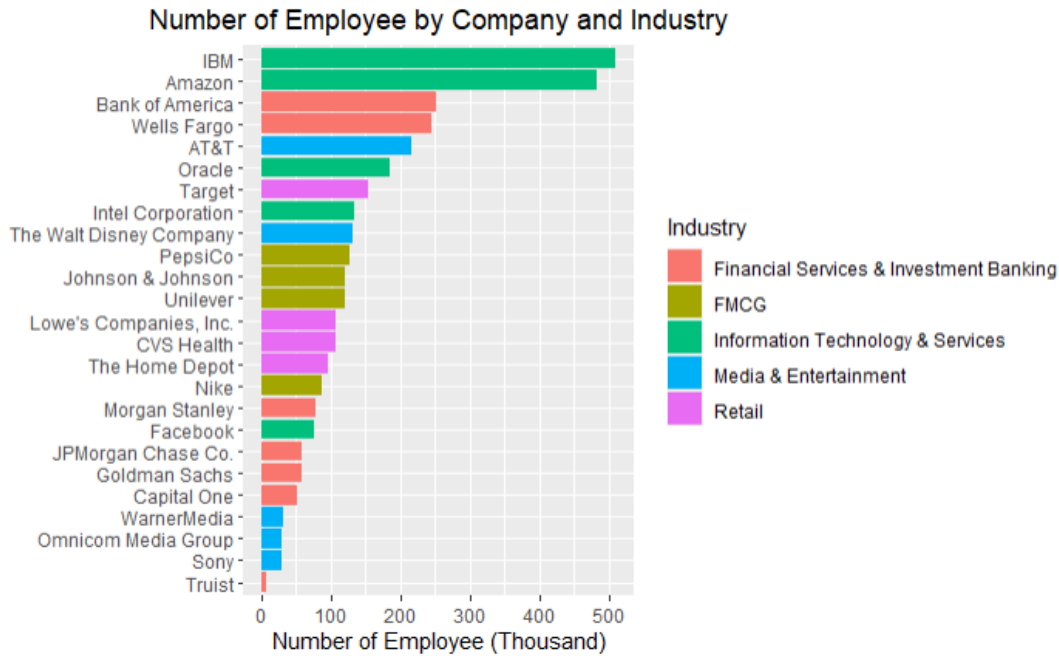


Figure 10: Hiring growth rate by company and industry

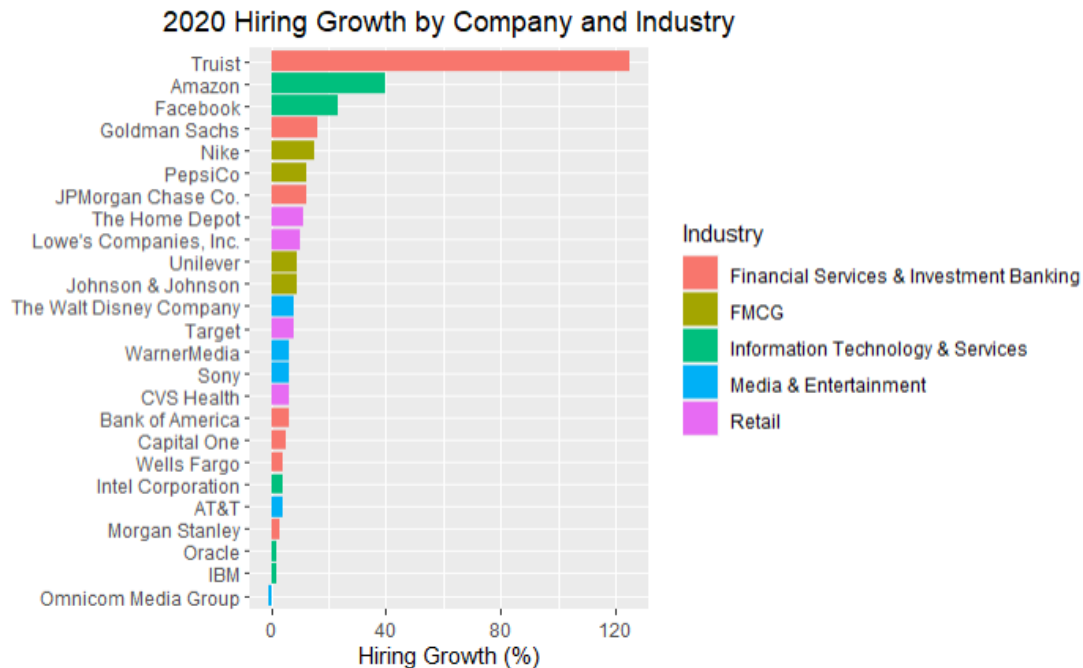


Figure 11: Revenue by company and industry in Q2. 2020

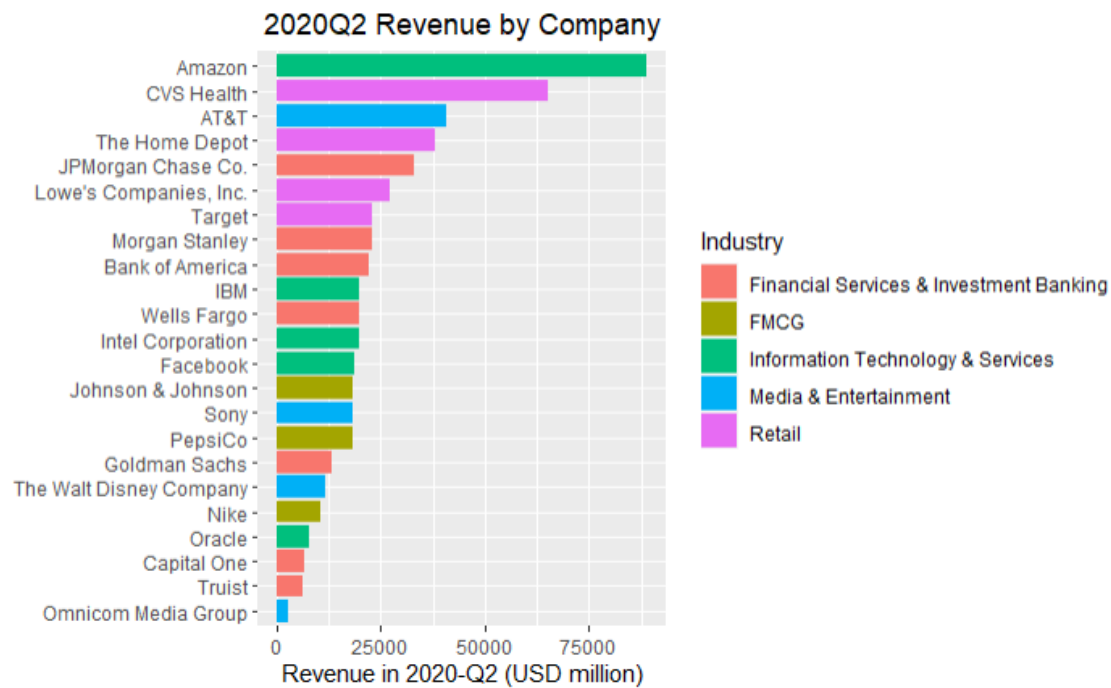


Figure 12: Revenue growth rate by company and industry in Q2. 2020

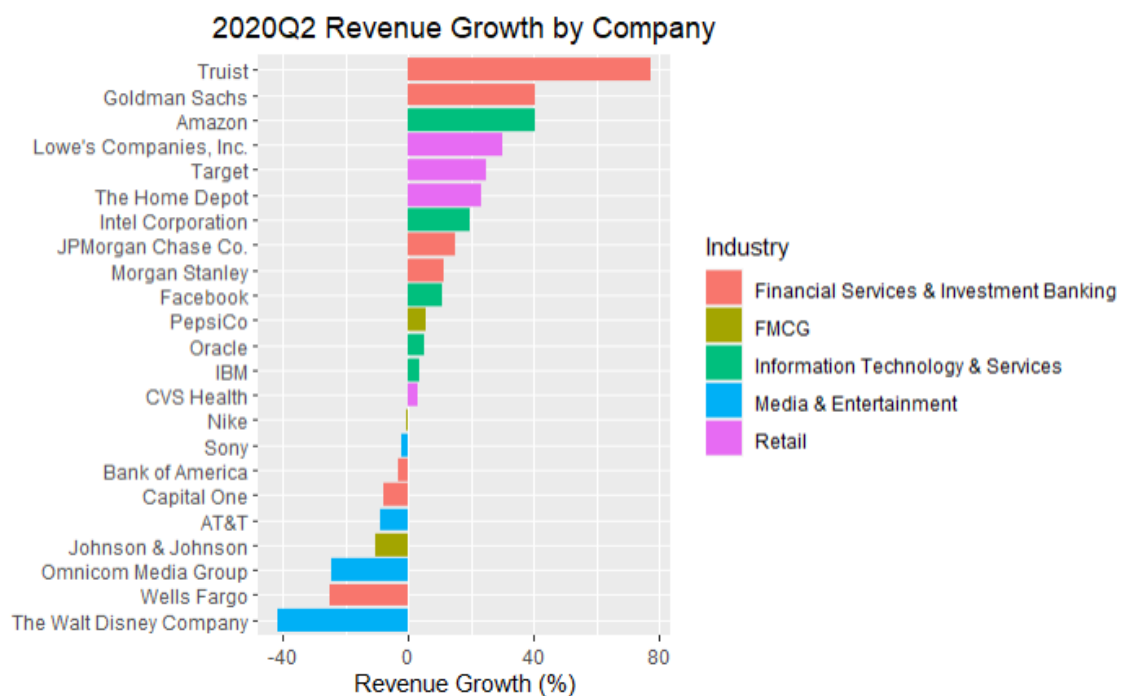


Figure 13: Company rating in the US

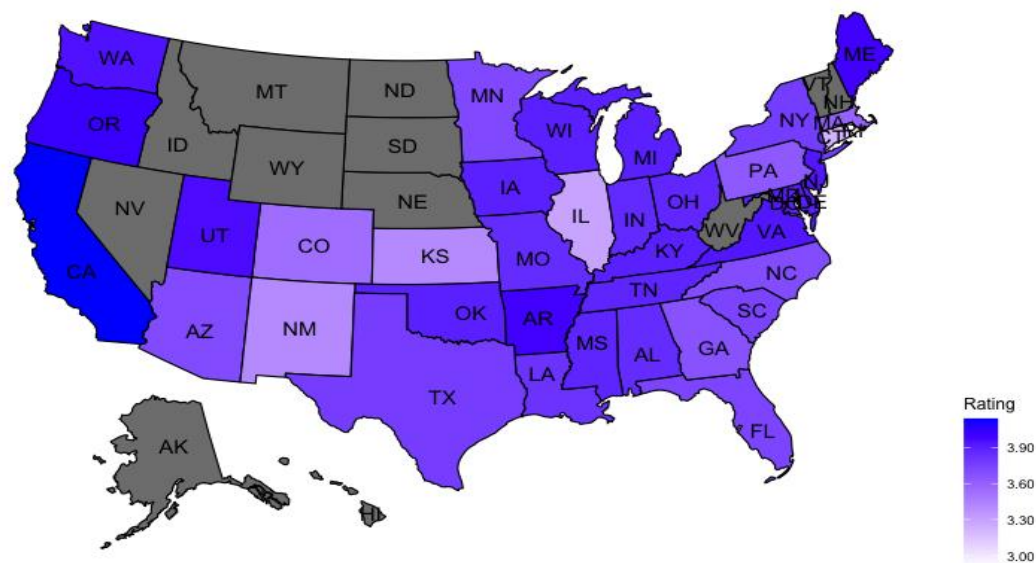


Figure 14: Number of job posts for different positions at Analyst/Engineer and Scientist/Researcher levels

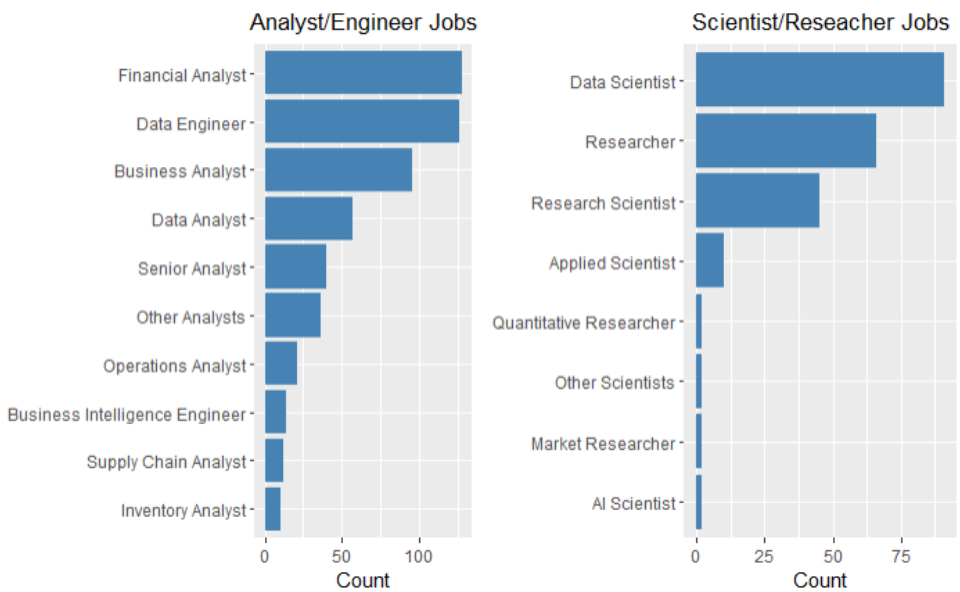


Figure 15: Number of job posts for different positions by top four companies

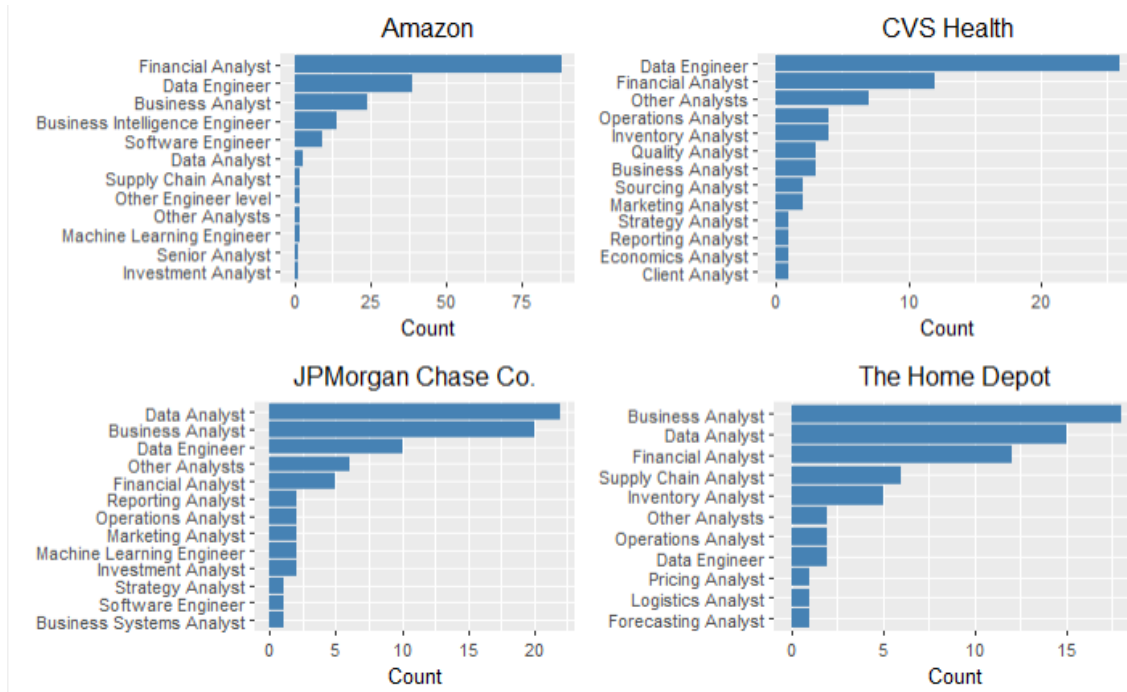


Figure 16: Top skills demanded for analytics jobs

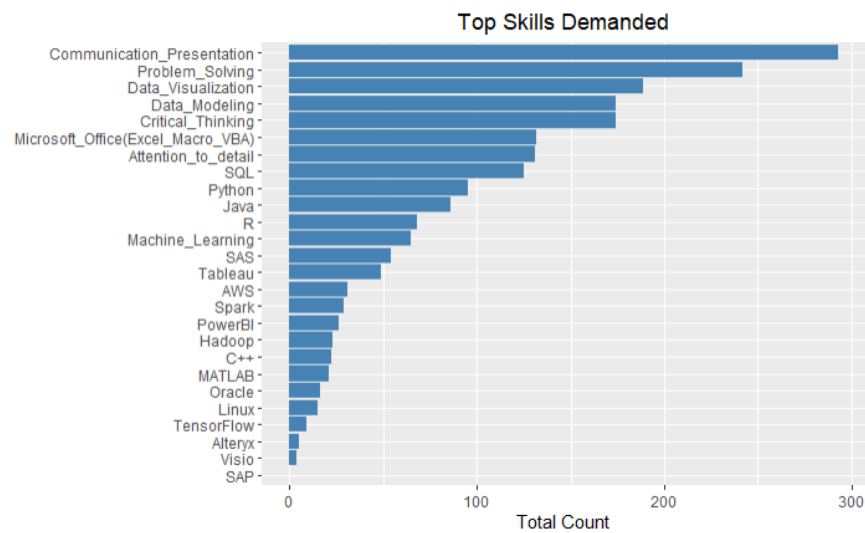


Figure 17: Skills for Financial Analyst and Business Analyst

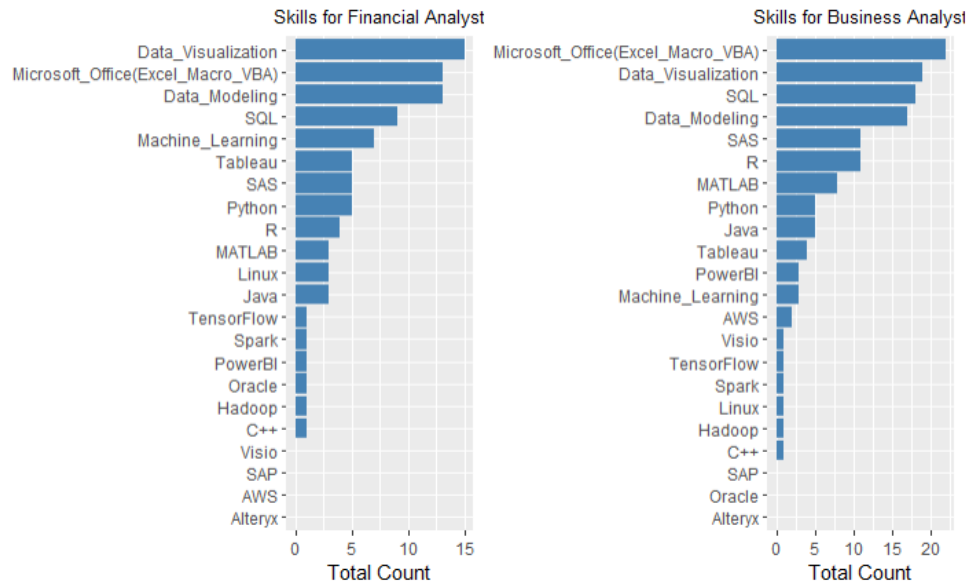


Figure 18: Skills for Data Engineer and Data Analyst

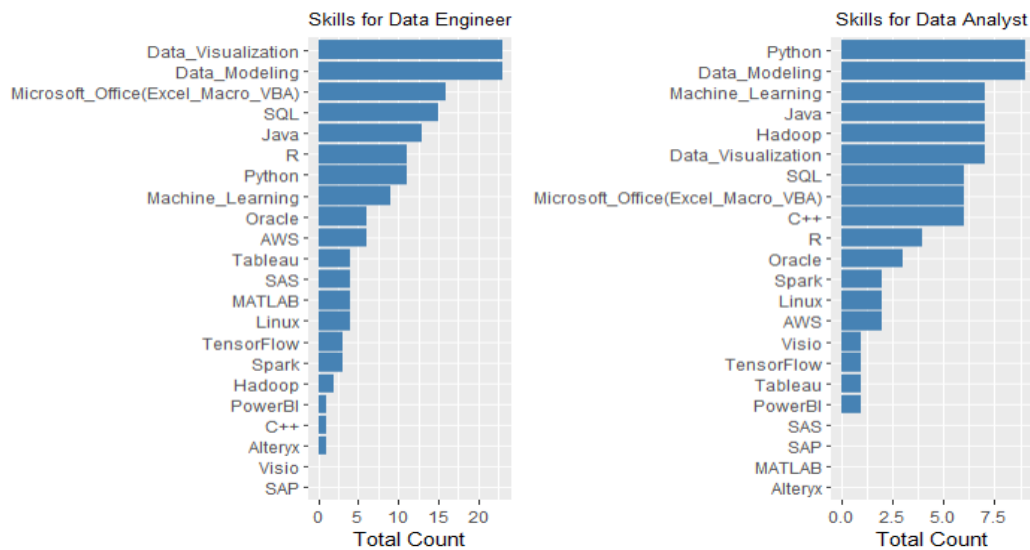


Figure 19: Skills for Data Scientist and Data Researcher

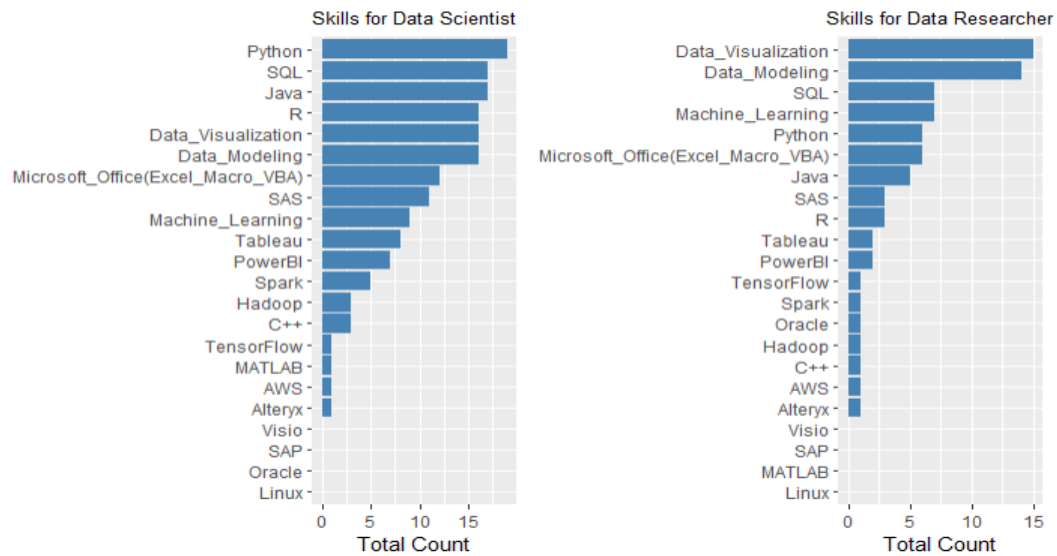


Figure 20: Ranking job titles for each skill

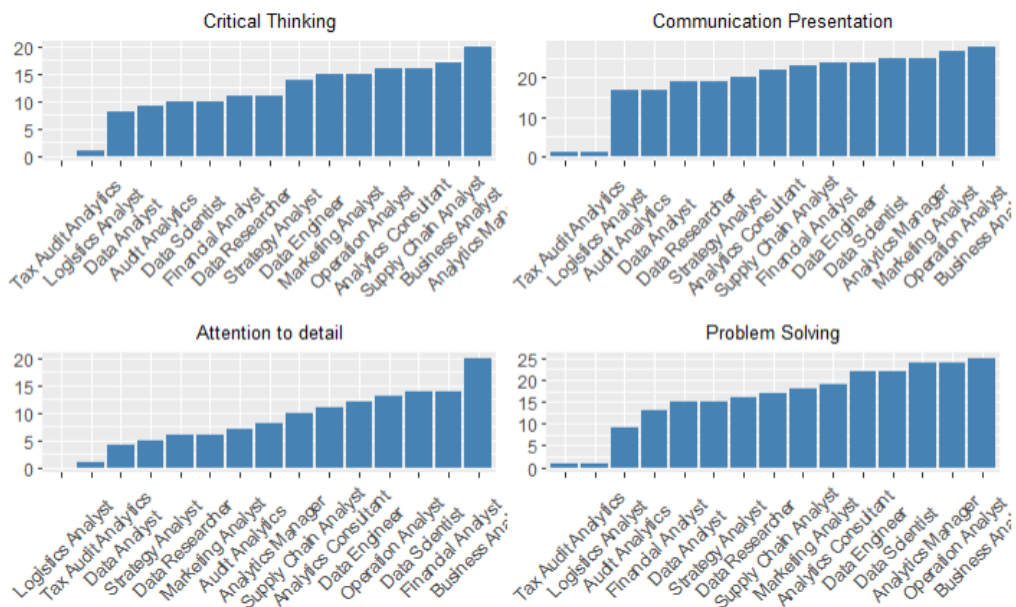


Figure 21: Ranking job titles for each skill

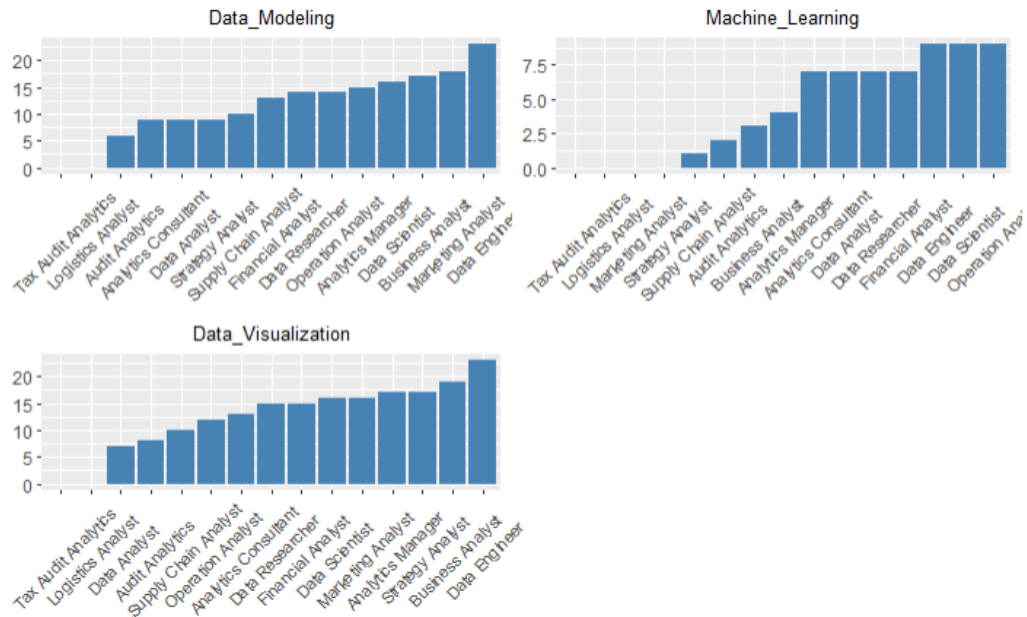


Figure 22: Ranking job titles for each skill

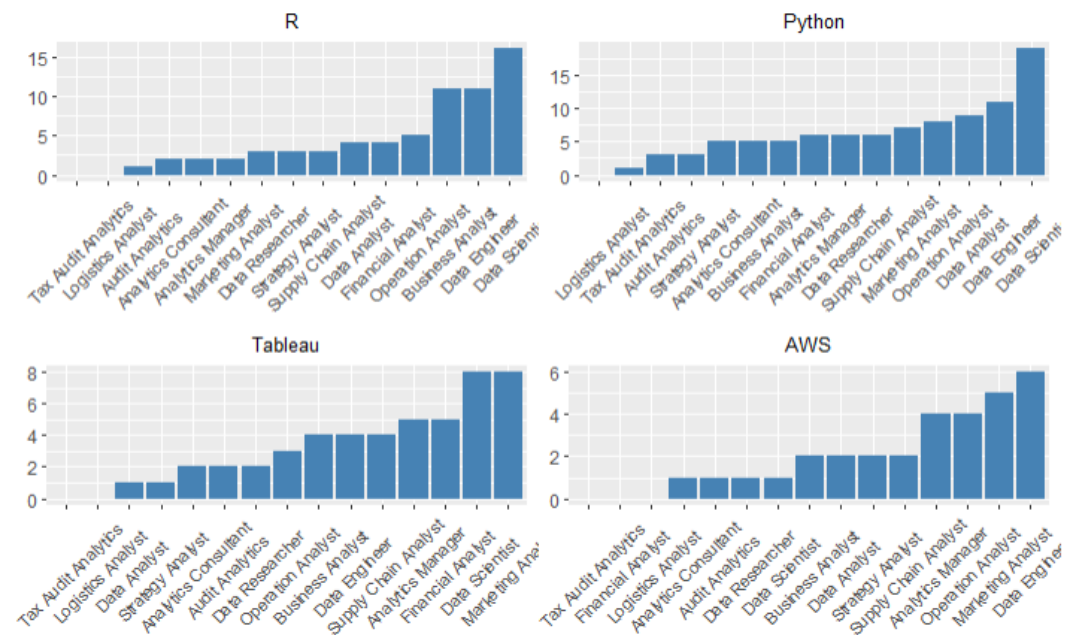


Figure 23: Ranking job titles for each skill

