An Empirical Study on Job Skills and Expectations in the field of Data Science

TEAM – DD

MAITRI SHAH

MONISH PHATARPEKAR

YANGYANG JIA

ERIC REYNOSO

JAUN RIVERA GONZALEZ

CALIFORNIA STATE UNIVERSITY, LOS ANGELES

ABSTRACT

The division of labor in formal organizations has important consequences for the distribution of opportunities and rewards. The present research is intended to contribute to the understanding of how well IT/IS job opportunities are, how job-choice decisions are made and the role of effective and ineffective recruiting practices in that process, what the most skill sets for data analysts. The issues are examined by tracking job seekers through the INDEED job search. The analysis can be considered an initial feasibility test for a larger comparative cross-country project that would aim to understand labor demand in the USA. The research provides our findings towards the job issues and aim to help job seekers develop crucial skills for a given occupation. Some implications of these results for studies of job organizations are discussed.

Keywords: IT/IS, Job Opportunities, Skill-sets, Data Science, Job seekers

1.INTRODUCTION

The emergence of client/server computing, the dominance of IT/IS companies and California's recovery from recession combine to make California IT/IS job market hot. With the current turmoil in the U.S. economy, it is more necessary than ever to ensure that graduates and employees have the skills necessary to compete in the job market. In high demand are programmers and data analysts experienced in relational technologies. More and more computers are also seeking application with computer technical skillsets like Python, R and C++.

Do large firms have a pattern of utilizing certain types of IT/IS skills more than others? Do data analysts need to more than one set of IT/IS skills for Fortune 500 companies? Understanding the set of skills required for IT/IS jobs is of importance to professionals as it governs their employability as well as their decisions to seek professional certifications. Meanwhile, the trend of job outsourcing means that certain types of jobs are kept in-house in firms while other types are contracted out. It thus becomes paramount for professionals to know the skill set that is required to preserve their employability in the job market. This ongoing study evaluates the skills required for IT/IS jobs by analyzing a broad set of online job descriptions based on INDEED (American worldwide employment-related search engine for job listings).

An initial set of data has been gathered through an automated retrieval of jobs descriptions from INDEED job search site. This data set will be analyzed job market in California compared to other states, the skillsets required for data analysts. The salary comparison of job from similar positions but different small and large-sized firms is one expected result of this study. The contribution of this study is expected to be the determination of patterns of skills required for IT/IS

jobs, a non-judgmental and quantitatively based definition of IT/IS jobs, and the investigation of differences in hiring patterns between firms of varying sizes.

To acquire and retain today's increasingly empowered customers, companies need to harness the insights in their data to personalize experiences at scale. Data scientists are crucial in turning the massive amount of data companies capture into action. They've always been in high demand, but until recently, only large enterprises and digital natives were willing to make the significant investment. Now, almost everyone is.

2. REVIEW OF LITERATURE

Several studies have highlighted the discontinuity between skills applications have and the specific needs of the employers. According to a YouGov survey for instance, 52% of employers said "none" or "few" recruits with a university degree were "work ready" (YouGov, 2013). Similarly, Harris (2013) argued that only around 25% of fresh graduate interviewees were employable.

Arum and Roksa (2010; 2014), Barberet al., (2013), and Gergen and Rego (2014) confirm these views. They argued that a mismatch of skills exist between university graduates and the needs of the industry (Gergen & Rego, 2014). Another survey found that specific industries, such as in the Information Technology (IT) sector, had a difficult time finding qualified applicants and 94% had to turn down candidates, at least in part, due to a deficiency in specific skills areas (Belkin, 2015).

An international survey conducted on graduates, educators, and employers from 9 countries identified a significant disconnect between the world of education and employment (Mourshed et al., 2014). It also noted that education providers had an inflated confidence regarding the relevance of what they were teaching. Whilst fewer than half of their surveyed students and employers believed that graduates were adequately prepared for entry-level positions, education providers were much more optimistic as 72% of them believed the new graduates were ready to work (Mourshed et al., 2014).

In the information era, enormous amounts of data have become available on hand to decision makers. Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques (Alexander Gude 2019).

Due to the rapid growth of such data, solutions need to be studied and provided in order to handle and extract value and knowledge from these datasets. Furthermore, decision makers need to be able to gain valuable insights from such varied and rapidly changing data, ranging from daily transactions to customer interactions and social network data. Such value can be provided using big data analytics, which is the application of advanced analytics techniques on big data.

With technology reaching new heights and a majority of the population having access to an internet connection, there's no denying that Big Data and data analytics have become hot topics in recent years – and a growing need. According to IBM, the number of jobs for data professionals in the U.S will increase to 2,720,000 by 2020 (CusterCharlie, C, 2019).

LinkedIn examined data from millions of member profiles, job openings, and salaries, ranking roles based on the following weighted five factors: Salary, career advancement, number

of job openings in the US, year-over-year growth in job openings, and widespread regional availability. The list features the jobs that had the highest combined score across these five factors.

But the skillset and job opportunities within data science go beyond the tech and digital spaces. What Skills Are Required for a Job in Data related jobs? 1 Python which is currently one of the most commonly used programming languages 2 SQL (Structured Query Language) as Working with data sources is a necessary aspect of data 3 Data Visualization Skills: Knowing how to visualize data and communicate results is a huge competitive edge for job seekers. Data scientists have a median base salary of $130,000, and saw 56% more job openings this year than last, according to the report. There are now more than 4,000 data science job openings nationwide (Rayome, A. D 2019).

Data related job titles are Data scientists who collect and analyze data and communicate actionable insights. Data scientists, data engineers and business analysts are known among the most sought-after positions in USA. These positions are challenging – and rewarding, with an average salary of $91,494. The demand for data analytics experts with technical backgrounds is at an all-time high. Then comes, Data engineers who often focus on larger datasets and are tasked with optimizing the infrastructure surrounding different data analytics processes. Then comes Data Analytics whose primary role is to deliver insights to a company to help their business (McKinsey Global Institute, 2011).

3. METHODOLOGY

In line with producing an analytical narrative in our data findings, the team as concurred in following a specific methodology to fully understand the captured data, what the data set is and its findings may represented, and what might possibility have be excluded from data and

thus effect the analytical and quantitative results. This set is an extraction of job posting records from the Indeed Job posting website given to use for assessment by our class advisor, Professor Ming Wang at Cal State University.

This section will explain the complete steps taken in regards of Indeed Data Analysis research paper.

3.1 Data Acquisition

We acquired the data set from the kaggle site. To help job hunters (including us) to better understand the job market, we scraped Indeed website and collected information of more than 400+ data related jobs around the United States. The information that we collected is: Company Name, Position Name, Location, Job type, Job title, Number of stars, Number of Reviews of the Company, Employee Skills, Top 9 skills, Top States for data jobs in United States.

In line with producing an analytical narrative in our data findings, the team as concurred in following a specific methodology to fully understand the captured data, what the data set and its findings may represented, and what might possibility be excluded from data and thus affect the analytical and quantitative results. Prior to beginning computational analysis of the data, the team began to assess the given data set. This set is an extraction of job posting records from the Indeed Job posting website given to use for assessment by our class advisor, Professor Ming Wang at Cal State University. The data had been extracted, and transformed into a readable CSV file, viewable by a spreadsheet application. Each record represents a job posting entry. The data set is not from the most current online posting representation from the Indeed Job Posting site.

INDEED DATA ANALYSIS

3.2 Dataset Details:

The dataset file is of csv format with a size of 18.14MB and it contains total of 5716 records. The skills have been parsed and categorized in their own fields as follows: Python, SQL, Machine Learning, R, Hadoop, Tableau, SAS, Spark, Java, and Others. Though these are the most common technology skills presently in the data analytics market, we considered utilizing this to set a standard, as opposed to sampling a large data set for all common technology skills sets that is much less practical to assess. The location of job was also categorized into separate fields in which each state currently has a technology center, e.g. Silicon Valley, Silicon Hills. These states include California, New York, Virginia, Texas, Maryland, Illinois, Washington, Massachusetts, District of Columbia, and North Carolina. Lastly, the employer's industry is categorized into one of major industries that has begun hiring and applying data analyst, data science, or data architects in a wide range through their business operations. These industries are Consulting and Business Services, Internet and Software, Banks and Financial Services, Health Care, and Insurance.

Each record is found with the following information:

Data Attribute	Description
Job_Title	Job Title is an the Indeed Job Portal
Link	Indeed Website Link
Queried_Salary	Indeed Salary Amount Quantity
Job_Type	Type of Job as in Indeed Job Portal
Skill	An Array of Job Skill sets for the particular Job
No_of_Skills	Total number of skill sets from in the job array
Company	Hiring Company Name
No_of_Reviews	Number of Reviews
No_of_Stars	Average numbers of stars per review
Date_Since_Posted	Days posted on Indeed Job Site
Description	Full Description of Job (HTML Format)
Location	State Location of Job
Company_Revenue	Quantity of Hiring Company's Annual Revenue
Company_Employees	Quantity of Employees from Hiring Company

INDEED DATA ANALYSIS

Company_Industry	Industry of Hiring Company
Skill Set Attributes	Description
python	True or False entry of skill set required for job
sql	True or False entry of skill set required for job
machine learning	True or False entry of skill set required for job
r	True or False entry of skill set required for job
hadoop	True or False entry of skill set required for job
tableau	True or False entry of skill set required for job
sas	True or False entry of skill set required for job
spark	True or False entry of skill set required for job
java	True or False entry of skill set required for job
Others	True or False entry of skill set required for job

Table 1

Job State Location Attributes	Description
CA	True or False entry of Job Location
NY	True or False entry of Job Location
VA	True or False entry of Job Location
TX	True or False entry of Job Location
MA	True or False entry of Job Location
IL	True or False entry of Job Location
WA	True or False entry of Job Location
MD	True or False entry of Job Location
DC	True or False entry of Job Location
NC	True or False entry of Job Location
Other_states	True or False entry of Job Location

Table 2

Job Industry Attributes	Description
Consulting and Business Services	True or False entry of Job Industry
Internet and Software	True or False entry of Job Industry
Banks and Financial Services	True or False entry of Job Industry
Health Care	True or False entry of Job Industry
Insurance	True or False entry of Job Industry
Other_industries	True or False entry of Job Industry

Table 3

3.3 Data Cleaning

Data cleansing is the process of detecting and correcting inaccurate records from the database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

In the Job_Title field of the dataset, special characters, NULL values as well as Empty rows was are found which is inaccurate for the analysis. For our research paper, we wanted to extract the "Job_Title" information. Therefore, we cleared out those rows from the dataset.

3.4 Analysis:

For the descriptive part of data analysis, we have used Expert Analytics which is a toolset of the SAP BusinessObjects Predictive Analytics application also we have used Tableau which is one of the most preferred tools for data analysis. As the Indeed job dataset thrives upon the skills first question is based to it.

Research Question 1: What are the top 3 job titles for data related jobs?

As per the current trend in data and its related jobs, it's important for us to know which the top 5 skills are needed for it.

With this our next research question is what the top 5 skills for match this top 3 job titles.

Research Question 2: What are the top 5 skills for data related jobs?

It's important to know the top 5 skills which can help to for data related job.

Along with this question our next research question was framed related to salary for data related job.

Research Question 3: Is the salary dependent on Job title or the company?

Our research is based to determine what salary offered to the different job titles and is it depended on company size/name.

After this we came up with our next research question which compares top 3 data related job titles offered in California with other states.

Research Question 4: List of Top 10 high paying Companies which for Data Scientist, Data Analyst and Data Engineer.

We found it more interesting to get to know companies which offer great packages to Data Scientist, Data Analyst and Data Engineer and that is how we came to research question 4.

Research Question 5: Data related Job opportunities are more in California compared to other states?

Being in university situated in California, we found it fruitful to have a visualization which shows the comparison of job opportunities between California and other states.

Research Question 6: Which industry has the maximum number of reviews given by users?

With so many industries being available, users who gave more reviews to many industries, so we have compared them all.

Research Question 7: Overall rating for Top 10 companies which has more than 10,000 ratings.

There are so many companies which has more than 10,000 but here we picked top 10 such companies whose ratings are given.

4. Data Analysis Results

By using Expert Analytics of the SAP Predictive Analytics tool for the descriptive and predictive analysis, here are our data analysis results.

Research Question 1: What are the top 3 job titles for data related jobs?

The finding to research question 1 is shown in Figure 1 which illustrates the top 3 data related job titles which includes data scientist, data analyst and data engineer.

As found in review of article, we got the same in our visualization after the analysis of the dataset (CusterCharlie, C, 2019).

Individually job title is found in the total of:

- Data Scientist: 2,543

- Data Analyst: 1,793

- Data Engineer: 1,379

So, we see that 'Data Scientist' job title has the highest number of jobs in the market.
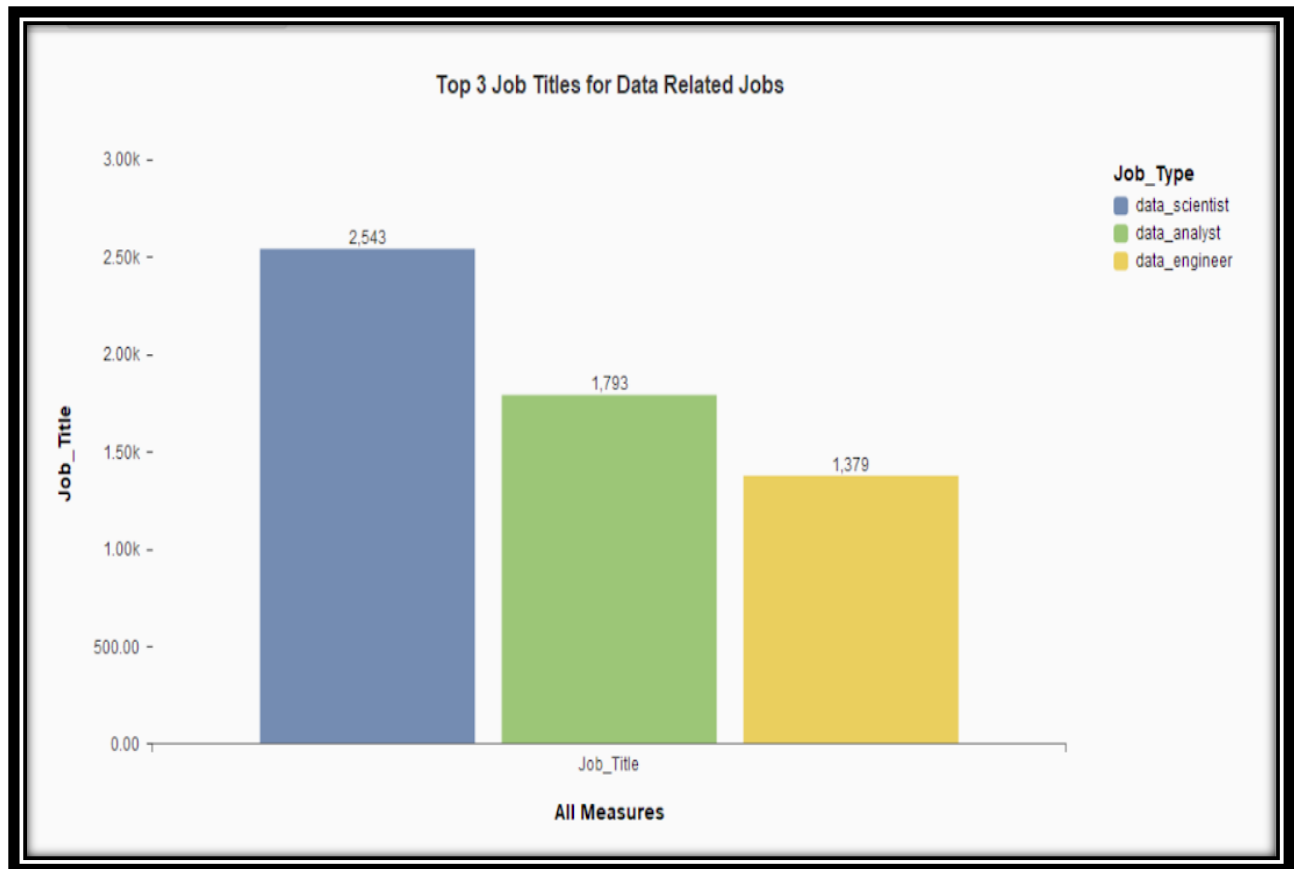
Figure 1

Research Question 2: What are the top 5 skills for data related jobs?

The finding to research question 2 which is in continuation of research question 1 and it is shown in figure 2. Top 5 skills for data related jobs found are Hadoop, python, sas, sql and tableau. Though the requirements for each of this skill differ from job title to job title, the total number represented on the bar is as per the records found for these skills by different companies which can be inferred from figure 2.

- For Data Analyst, most required skill is SQL with total of 1,044

- For Data Engineer, most required skill is python with total of 901 and near to is SQL with 867.

- For Data Scientist, most required skill is python with total of 1,912.
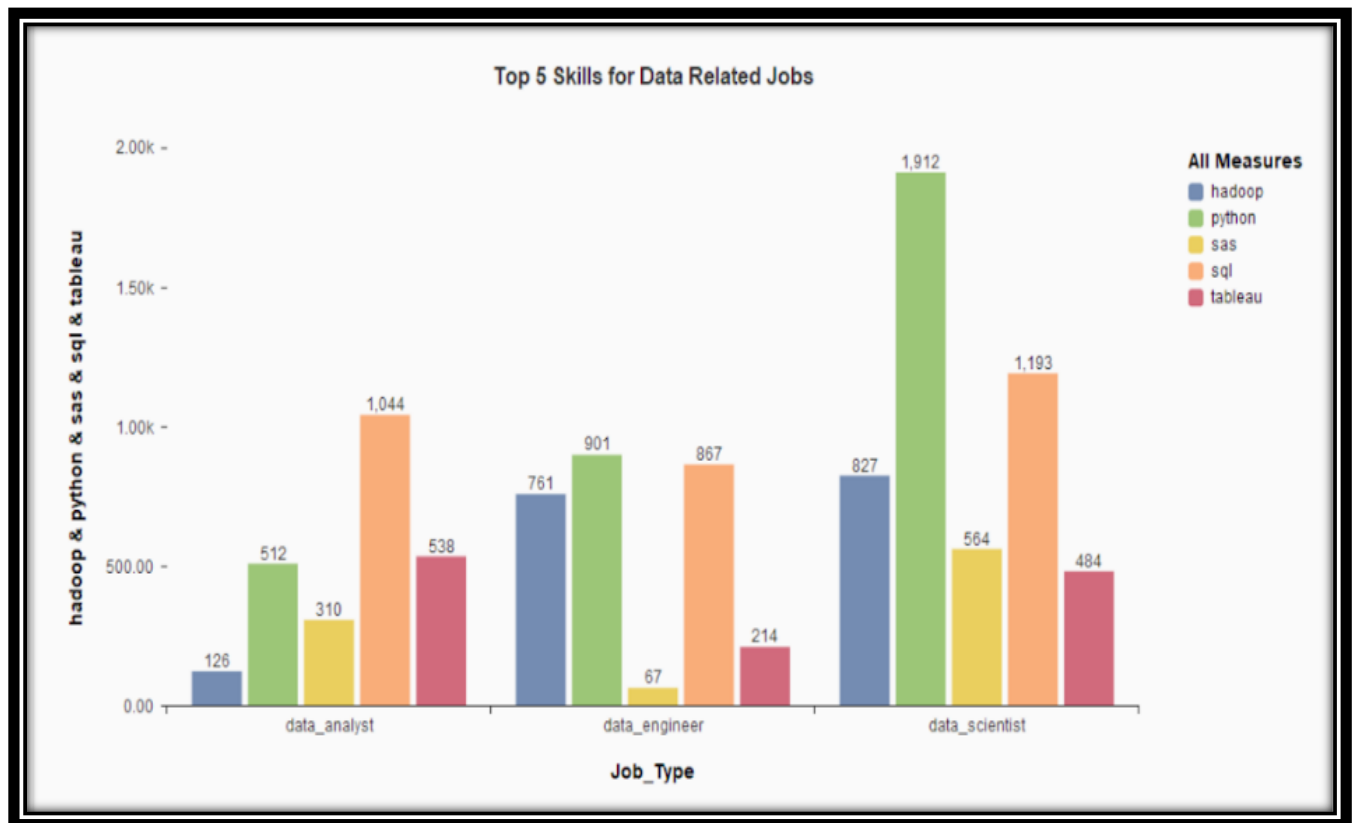
Figure 2

Research Question 3: Is the salary dependent on Job title or the company?

In response to research Question 3 which is in support to research Question 2. Yes, salary offered is depended on job title and not the company which can be inferred from figure 3. Each of the job title has been offered following package:

Data Scientist: 275 companies offered salary of 100000 - 119999, 217 companies offered salary of 120000 - 139999, 101 companies offered salary of 140000-159999, 84 companies offered salary of 80000-99999, 79 companies offered salary of 120000-139999

Data Analyst: 207 companies offered salary of <80000

Data Engineer: 118 companies offered between 120000-139999, 59 companies offered between
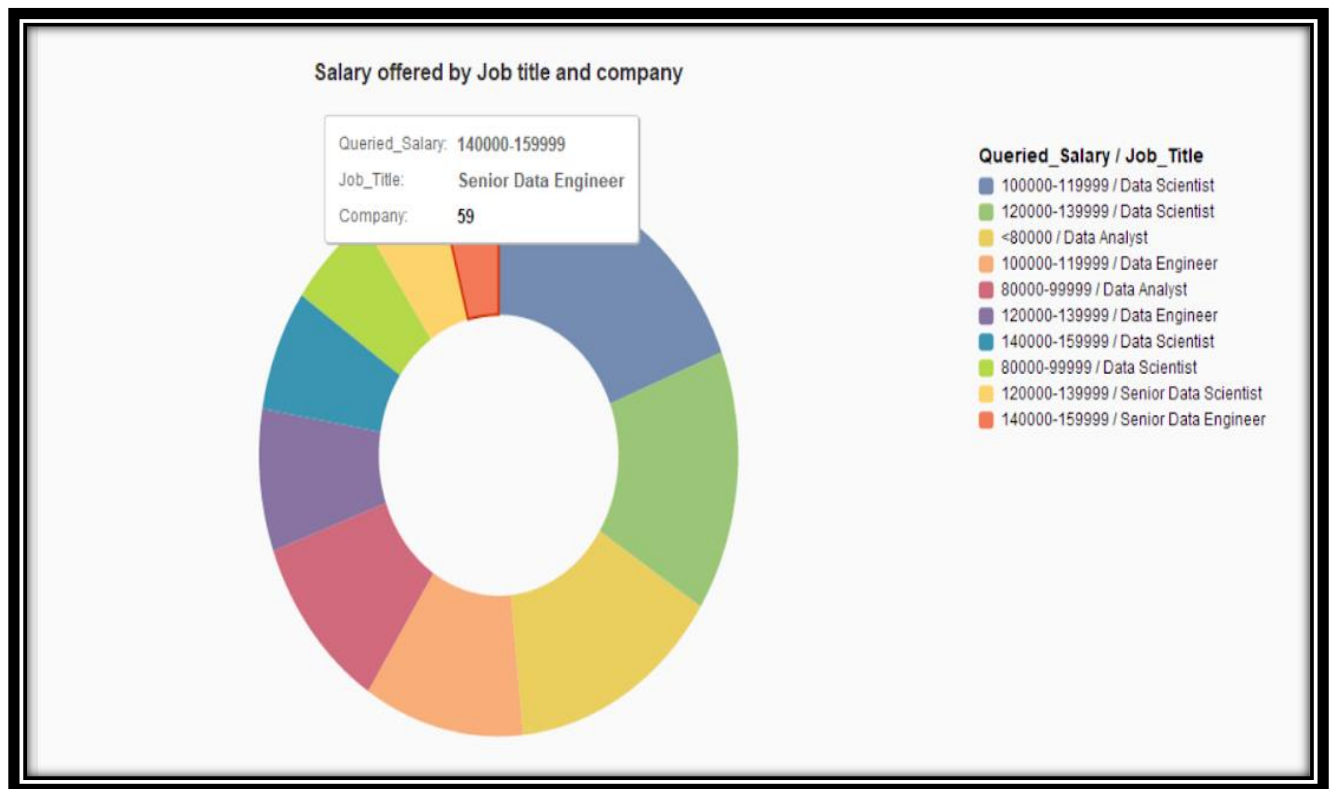
140000-159999

Figure 3

Research Question 4: List of Top 10 high paying Companies which for Data Scientist, Data

Analyst and Data Engineer

In response to Research Question 4, we have gathered 3 different visualizations for Data Scientist,

Data Analyst and Data Engineer which illustrates top 10 high paying companies.

Figure 4.1 shows top 10 high paying companies name for Data Scientist along with total number

of people working on that position with the same package. This includes:

- BoozAllen hamilton with package of $120000-$139999 and have a total of 60 employees

- BoozAllen hamilton with package of $100000-$119999 and have a total of 24 employees

- BoozAllen hamilton with package of $140000-$159999 and have a total of 23 employees

- Facebook with package of $140000-$159999 and have a total of 21 employees

- Walmart with package of $140000-$159999 and have a total of 21 employees

- Walmart Ecommerce with package of $140000-$159999 and have a total of 17 employees

- IMB with package of $120000-$139999 and have a total of 16 employees

- Payette Group with package of $140000-$159999 and have a total of 15 employees

- KPMG LLP with package of $100000-$119999 and have a total of 14 employees

- Walmart with package of $120000-$139999 and have a total of 14 employees

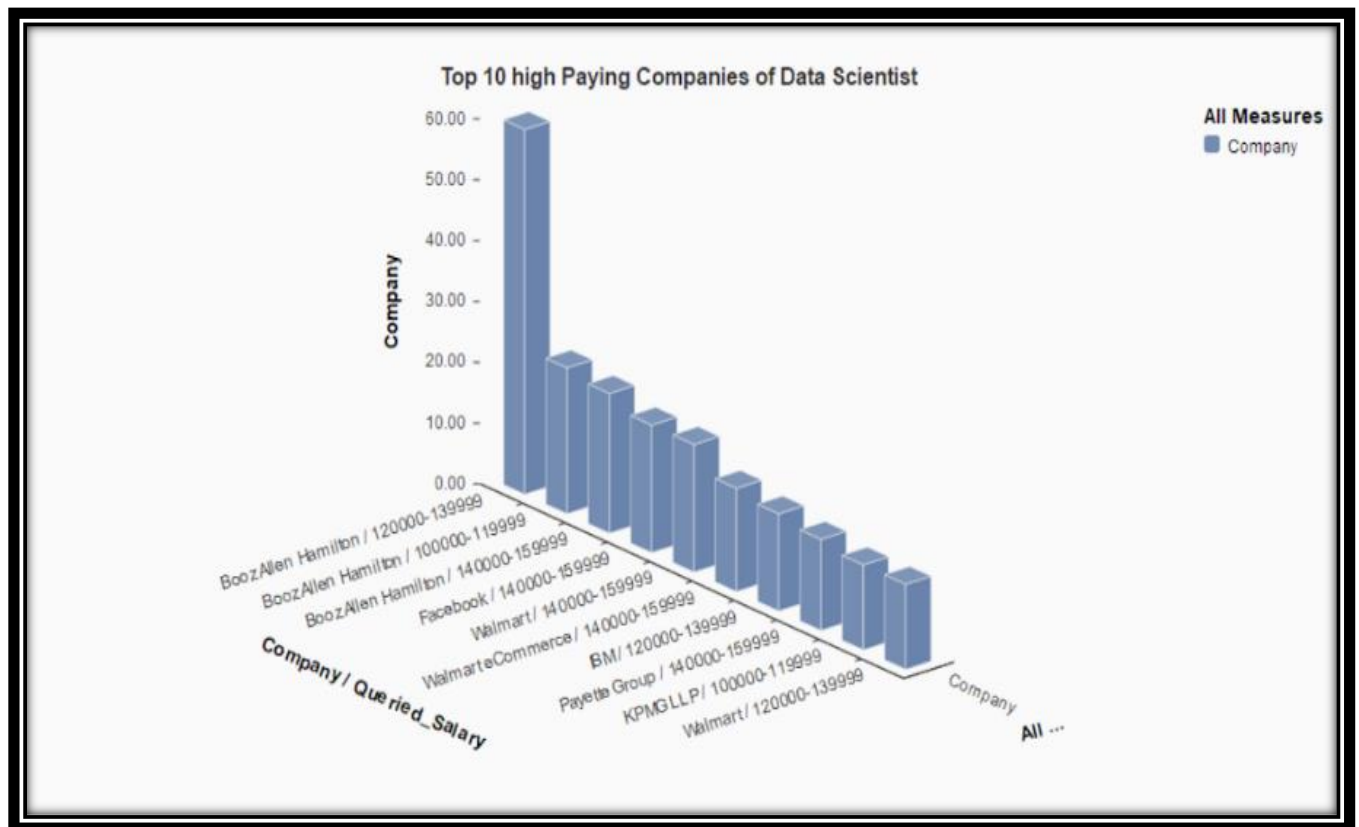There are companies repeating because they offered the same job title but with different package.

Figure 4.1

Figure 4.2 shows the top 10 high paying companies name for Data Engineer along with total

number of people working on that position with the same package.

This includes:

- Hamham with package of more than $160000 and have a total of 19 employees

- Allstate with package of $100000-$119999 and have a total of 15 employees

- Capital One with package of $140000-$159999 and have a total of 14 employees

- Capgemini with package of $100000-$119999 and have a total of 12 employees

- Capital One with package of $100000-$119999 and have a total of 9 employees

- KPMG LLP with package of $100000-$119999 and have a total of 9 employees

- Workbridge Associates with package of more than $160000 and have a total of 8 employees

- Booz Allen Hamilton with package of $100000-$119999 and have a total of 7 employees

- Hamham with package of $140000-$159999 and have a total of 7 employees

- Jobspring Partners with package of more than $160000 and have a total of 7 employees

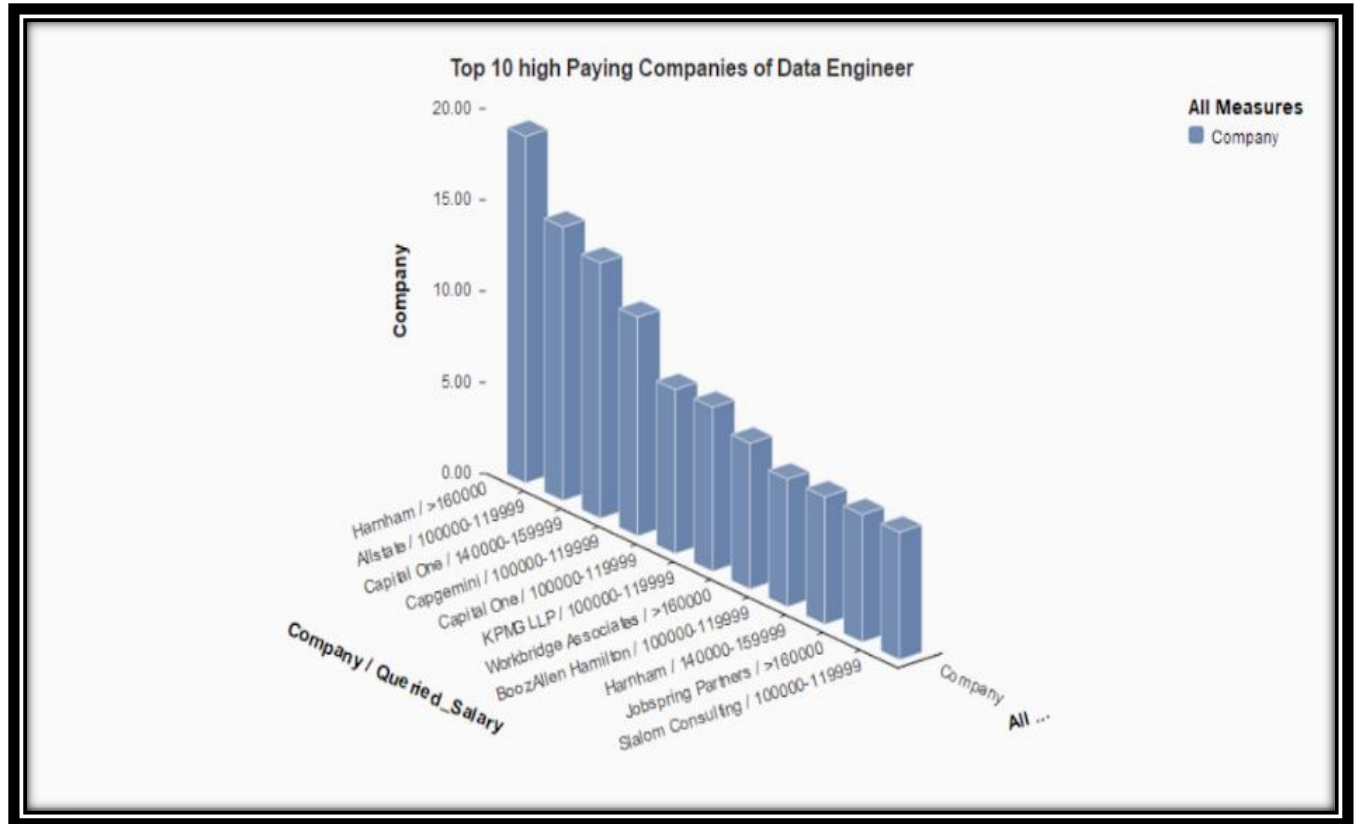- Slalom Consulting with package of $100000-$119999 and have a total of 7 employees

Figure 4.2

Figure 4.3 shows the top 10 high paying companies names for Data Analyst along with total number of people working on that position with the same package.

This includes:

- Hamham with package of more than $160000 and have a total of 13 employees

- Capgemini with package of $100000-$119999 and have a total of 11 employees

- Uber with package of $100000-$119999 and have a total of 8 employees

- USAA with package of $140000-$159999 and have a total of 7 employees

- Emblem Health with package of $100000-$119999 and have a total of 6 employees

- Accenture with package of more than $160000 and have a total of 5 employees

- Booz Allen Hamilton with package of $100000-$119999 and have a total of 5 employees

- Facebook with package of $100000-$119999 and have a total of 5 employees

- NBC Universal with package of $100000-$119999 and have a total of 5 employees

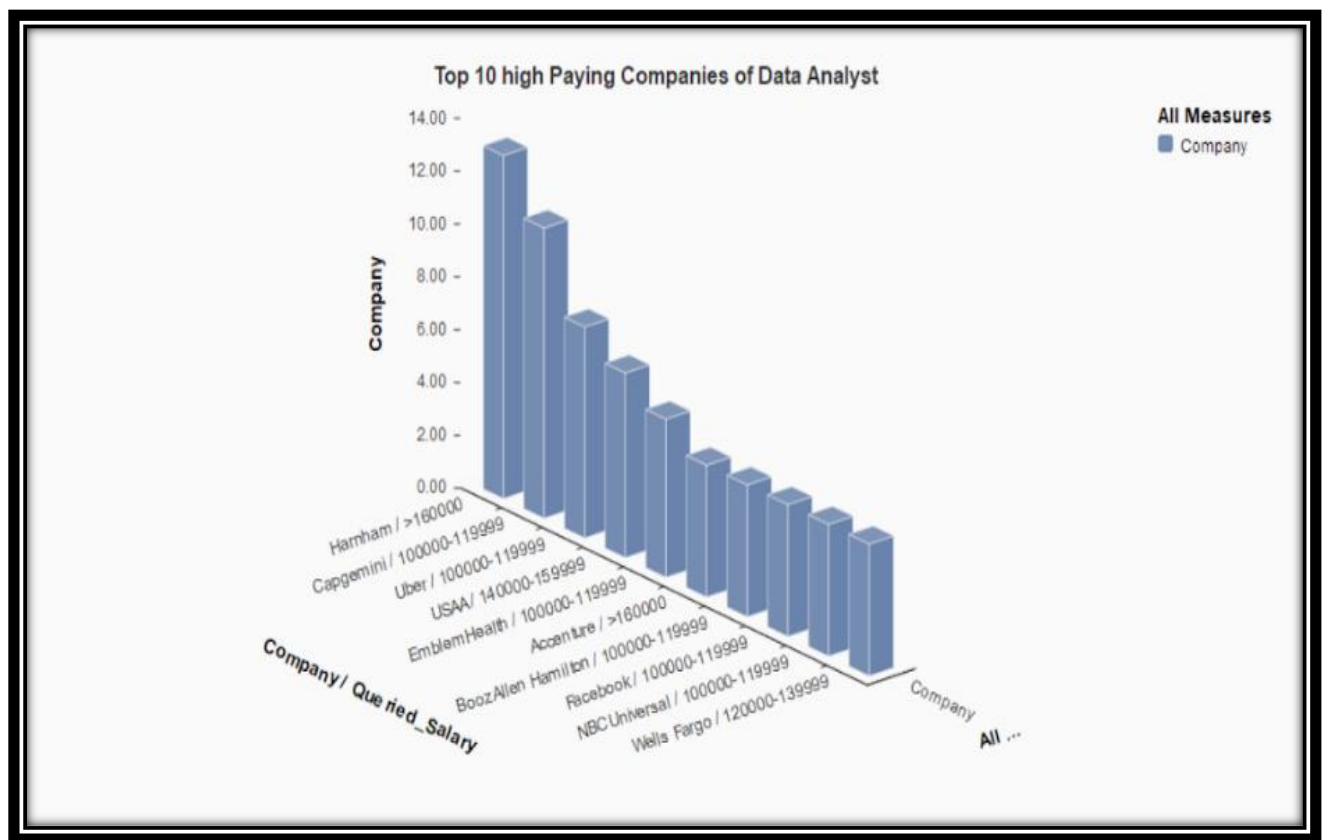- Wells Fargo with package of $120000-$139999 and have a total of 5 employees

Figure 4.3

Research Question 5: Data related Job opportunities are more in California compared to other states?

In response to research Question 4, the answer is NO as when considering top 3 data related jobs as seen in figure 1, opportunities in California with other states is not more for each of the job titles which can be seen from figure 5.

- For data scientist, we found more jobs in California (723) than other states (674).

- For data analyst, in other states (625) has almost double job options than California (376).

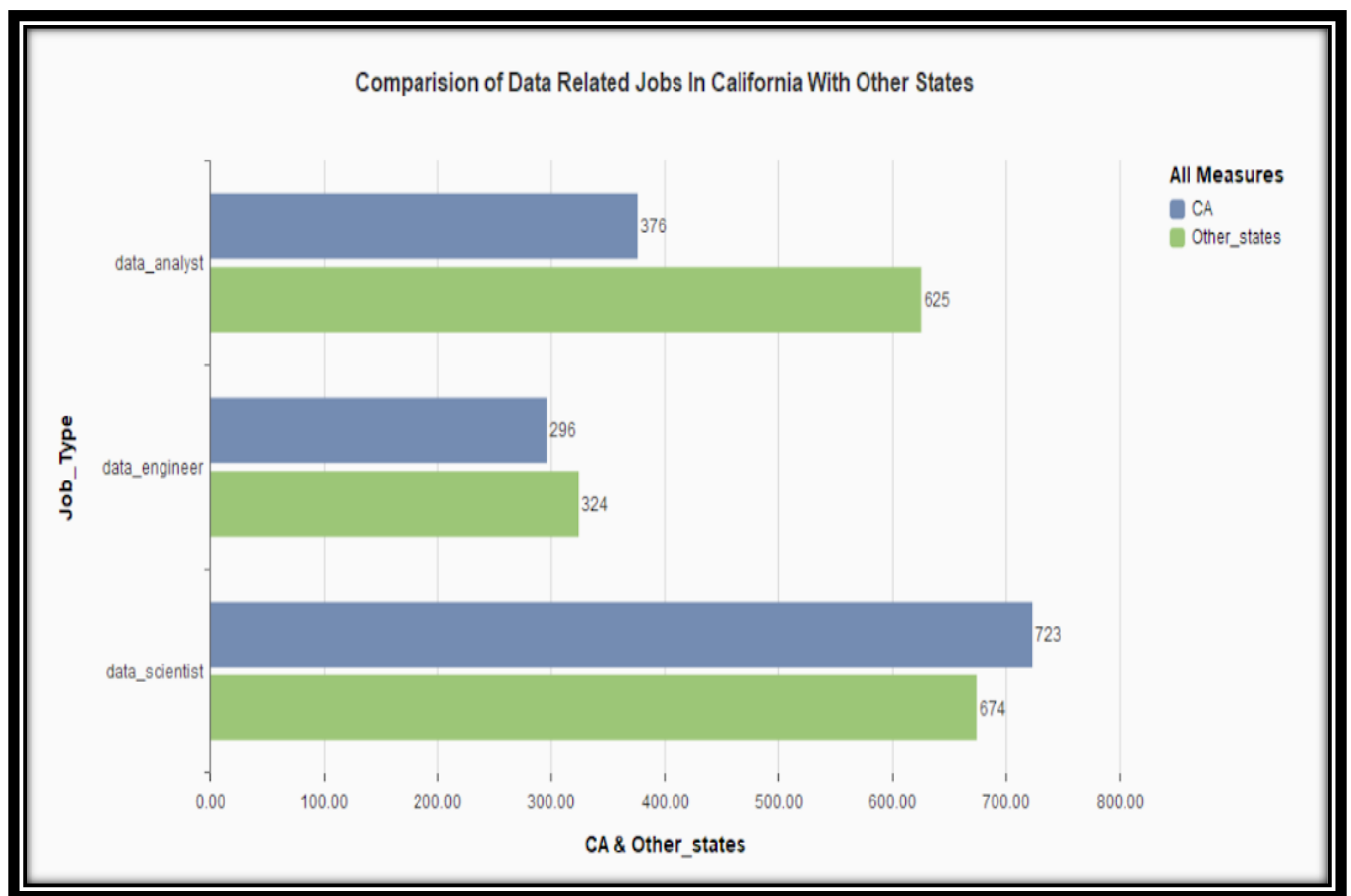- For data engineer, other states (324) have more options than California (296).

Figure 5

Research Question 6: Which industry has the maximum number of reviews given by users?

As we can see in figure, We are looking for top 9 companies in terms of Revenue. For this visualization we have used Tableau to get the visualization. We can see in figure how we are categorizing top 9 companies depending on their revenue. The top 9 companies have been mentioned:

- Aerospace with a revenue of $2,60,331

- Banks and Financial Services with a revenue of $26,95,127

- Consulting and Business Services with a revenue of $28,28,969

- Educations and schools with a revenue of $67,330

- Healthcare with a revenue of $4,58,003

- Industrial Manufcturing with a revenue of $3,19,159

- Insurance with a revenue of $5,10,283

- Internet and Software with a revenue of $7,48,971
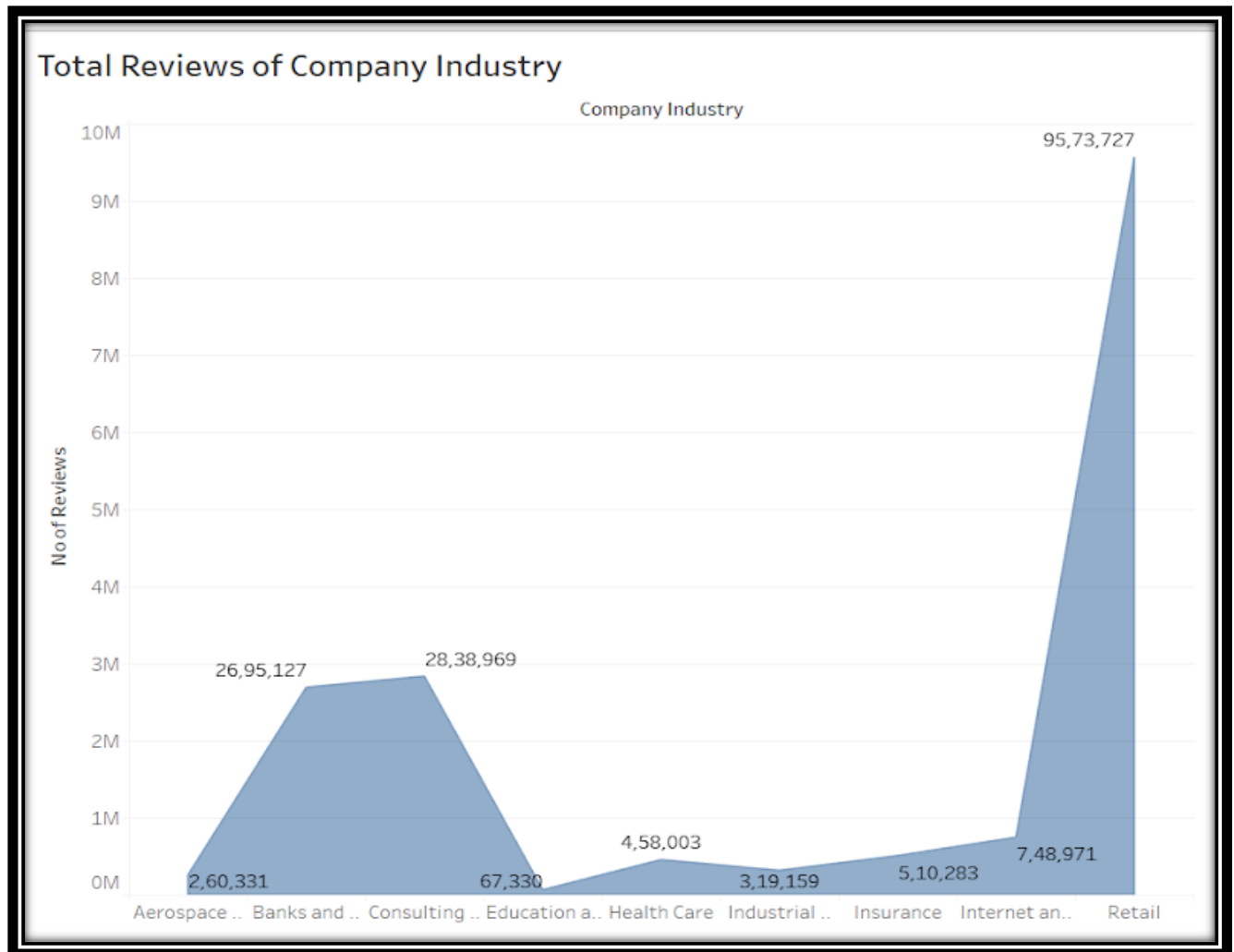
- Retail with a revenue of $95,73,727

Figure 6

Research Question 7: Overall rating for Top 10 companies which has more than 10,000 reviews.

In response to research question 5, top 10 companies based on reviews (>10,000) has been gathered. This review is given by users who are using Indeed Job portal.

Figure 10 which is illustrated in the form of line chart shows the top 10 companies names along with their rating (out of 5) which is given by users.

- Accenture and Cognizant have been found with highest rating which is 4/5

- Citi, IBM, JP Morgan chase and Version is found with rating of 3.9/5

- The Home depot, Wells Fargo is found with rating of 3.8/5

- Comcast with 3.7/5

- Walmart with 3.6/5
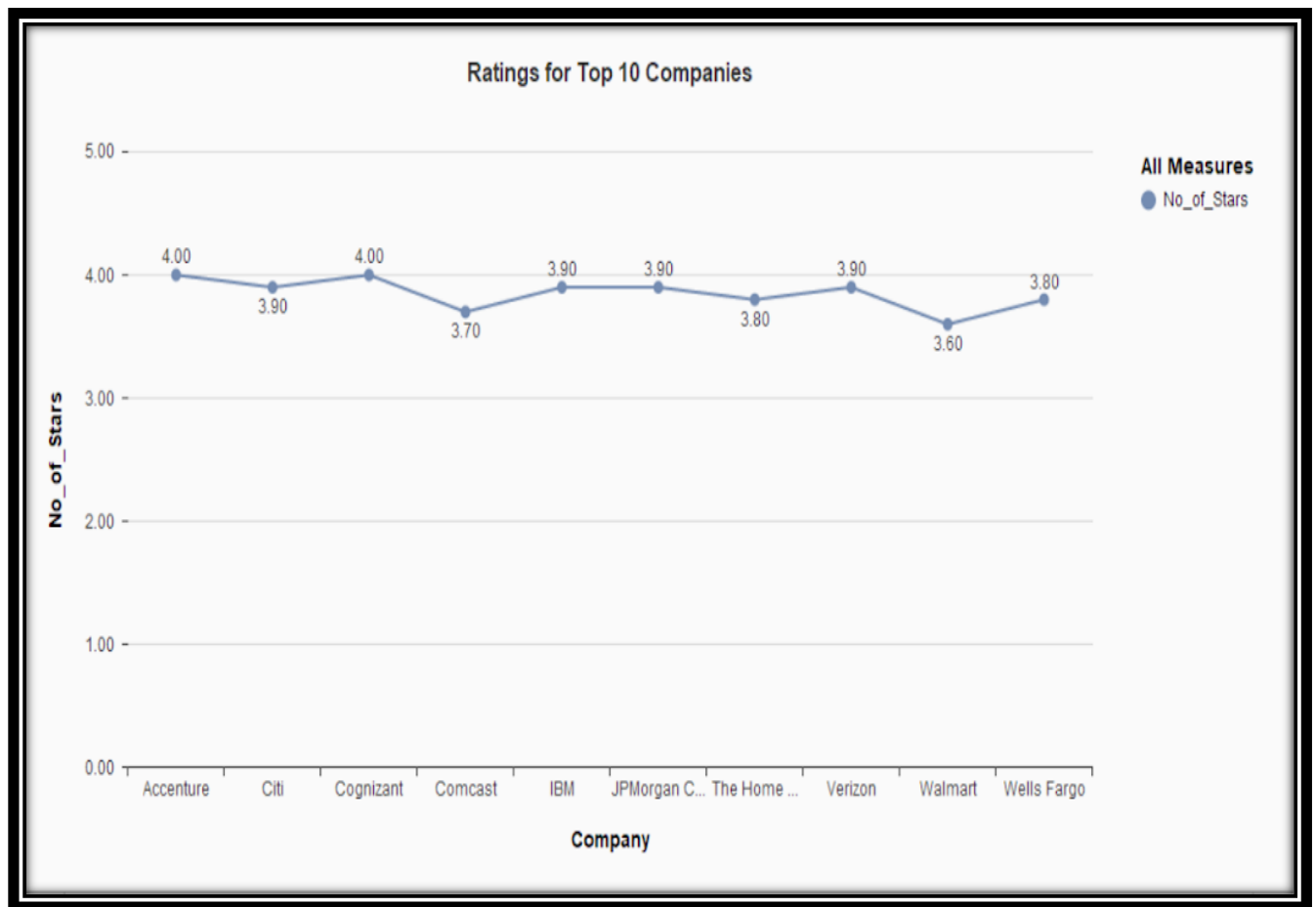
Figure 7

5. Conclusion

The database of Indeed allows anyone who is seeking for a job to apply using the keywords that they have set. This database can be considered as a first step of allowing many unemployed individuals to apply using the Indeed database. The database is divided into fields that will enable a more natural search. Some of the felids that were included were; the title of the position, the skills need to concur the position, the state of the position that is located, and the Salary that the position comes in. The Indeed job dataset has shown any position available for individuals to apply towards the selected cities of CA, NY, VA, TX, MA, IL, WA, MD, DC, and NC. Based on the research, it has provided an insight of job issues and how it can help for a job seeker to gain and practice their crucial skills for a particular job position.

This would allow an individual to have gain knowledge and improve the quality of workers who have the skills in SQL, Machine Learning, R, Hadoop, Tableau, SAS, Spark, Java, and other computer skills need to complete the job. The development of an Indeed database is critical for the exploration of the job description and their related Salary. The database with the Indeed job title will help the event of an empirical model to help determine the factors with the highest numbers of hires. We can also create a report with the applicant related skills to show the percentage of the skills in need.

After analyzing the several reports presented in the above sections, we can determine that there is a high number of Data Scientist with the following skills need to complete the job. Those skills are SQL and other skills related the company performance. As mentioned in the Review of Literature, there have been several studies that have put a spotlight to the readiness of an individual who is ready to join a workforce team. According to Roksa, most of the skills that are taught in

the University were not similar to the Information Technology sector. Thus, those skills determine the Salary of the job title. The minimum number of skills need to land a job is having more than five skills. The full-time job requires an individual to have at least one coding language. The queried Salary is less than 80,000 thousand dollars for a Data Scientist position. 1 in 4 jobs require 10 or more job skills, and the number of stars is above 4.0, and the position is located in MO, TX, OR, DX, GA, and ID. This particular data shows that the information about the job position is not situated around NY. Most of the jobs are in need in California. The job opportunities are higher in CA compare to other states. The reason is based on the report that shows in the graph, in Silicon Valley has a higher number compared to other cities in CA. One in ten offer benefits to their employees.

An essential factor to consider when one applies for an Indeed job listing is to check the benefit of the job and the current Salary and skills. The skills are connected to the wage. Therefore, there is a direct connection and link to the benefits and salary status. The job listing on Indeed currently holds to 5,000 thousand jobs. The client and the recruiters have a direct link to the listing who have an account with Indeed and their logins provide them with those posting and the prospective benefits and Salary.

Overall, seeing all of the skills, benefits, and Salary listed on the Indeed.com   website. It will allow the person to apply for the jobs of the information that is provided on the site. However, if the information is not listed on the website or the post, there is a margin of 14% that the perspective will not even look at the job. One of the reasons could be if the information is not listed, then the percentage of the information could be less reliable.

Bibliography:

Miroslav. B. (2016). The surprisingly exclusive nature of medium- and low-skilled jobs: Evidence from a Slovak job portal"

Wilson.L 2019. "California dreaming." Computerworld, 8 Apr. 1996, p. 87. Accessed 27 June 2019.

Harris, S. September 12. 2013 "Half of employers say graduates are 'not up to the job' job': Findings fuel fears universities fail to equip students with life skills"..

Mourshed, M., Patel, J., & Suder, K. 2014 "Education to employment: Getting Europe's youth into work"

Archived at the World Data Center for Paleoclimatology, www.ngdc.noaa.gov/wdc/usa/paleo.html.

McKinsey Global Institute, June 2011 "Big data: The next frontier for innovation, competition, and productivity"

Thomas H. Davenport, D.J. Patil, October 2012 "Data Scientist: The Sexiest Job of the 21st Century", Harvard Business Review

Burtch Works "Big Data Career Tips" December 2014 http://www.burtchworks.com/big-data-analyst-salary/big-data-career-tips/

Sampaio, A. (n.d.). Web Portal Application Development Technologies. Encyclopedia of Portal Technologies and Applications, 1131-1137. doi:10.4018/978-1-59140-989-2.ch184

Mansourvar, M., & Yasin, N. B. (2014). Development of a Job Web Portal to Improve Education Quality. International Journal of Computer Theory and Engineering,43-46. doi:10.7763/ijcte.2014.v6.834

Passy, J. (2019, May 03). Here's the No. 1 state for jobs in the U.S. - and it's not California or New York. Retrieved from https://www.marketwatch.com/story/heres-the-no-1-state-for-jobs-in-the-us-and-its-not-california-or-new-york-2019-01-31

Jeffdanielsca. (2019, March 22). California accounted for nearly three-quarters of US jobs added in February. Retrieved from https://www.cnbc.com/2019/03/22/california-accounted-for-nearly-three-quarters-of-us-jobs-added-in-february.html

2019 Survey Results: Salary Negotiation in Data Science & Analytics. (2019, June 20). Retrieved from https://www.burtchworks.com/2019/05/06/2019-survey-results-salary-negotiation-in-data-science-analytics/

Alexander Gude. (2019, March 26). Data Science Salaries. Retrieved from https://alexgude.com/blog/data-science-salaries/

Retrieved from https://www.bizjournals.com/bizwomen/news/profiles-strategies/2016/10/you-should-negotiate-for-a-job-title-heres-why.html?page=all

How Does a Job Title Affect Salary Negotiation? (2017, December 05). Retrieved from https://www.payscale.com/salary-negotiation-guide/job-title-affect-salary-negotiation

/@dimitrilinde. (2017, July 06). Scraping and Classifying Indeed Job Postings for Data Occupations, Part 1. Retrieved from https://medium.com/@dimitrilinde/scraping-and-classifying-indeed-job-postings-for-data-occupations-part-1-22c948860e61

50 Free Job Posting Sites to Help You Hire Faster and Smarter | Hire by Google. (n.d.). Retrieved

from https://hire.google.com/articles/free-job-posting-sites/

Monica, J. O., Danielle, Karen, Lara, Rick, Cynthia, . . . Avondale. (n.d.). Indeed. Retrieved from

https://www.consumeraffairs.com/employment/indeed.html

CusterCharlie, C. (2019, June 10). 10 High Paying Data Analytics Jobs. Retrieved from

https://www.dataquest.io/blog/10-data-analytics-jobs/

Rayome, A. D. (n.d.). Why data scientist is the most promising job of 2019. Retrieved from

https://www.techrepublic.com/article/why-data-scientist-is-the-most-promising-job-of-2019/