Salary Data Analysis

DS 311: Technologies in Data Analytics

Professor Norman Lo

Justin San Andres, Seyoung Kim, Viviana Huang Zheng

Executive Summary

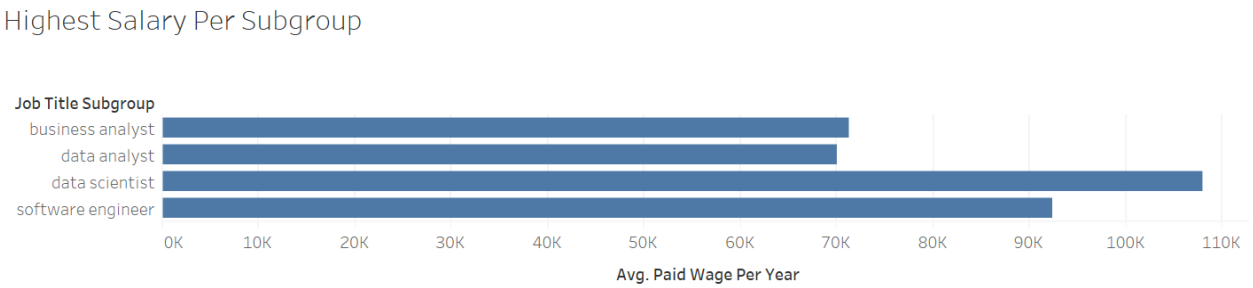
The United States has many companies that are looking for workers outside of the US. There are also many immigrants coming to the US looking for a job. In this labor market, it's important to be able to identify specific job-titles, locations, and companies of jobs that workers would want in order to cater towards the market. A big factor that we would base this off of is the salaries that workers would make in specific job positions. Using publicly available data compiled by the US Department of Labor's Office of Foreign Labor that provides information about job titles, job title subgroups, work state, paid wage per year, and employer names, etcetera, team The-Code-Tastic-Four analyzes the salary data set in order to identify which jobs, companies, and where would be the best place for people looking into working in the US.

Exploratory Data Analysis

This report will answer three main objectives related to the salaries of jobs related to data. In order to do so, our team has built various codes, tables, and charts to provide evidence for our findings using Python and Tableau.

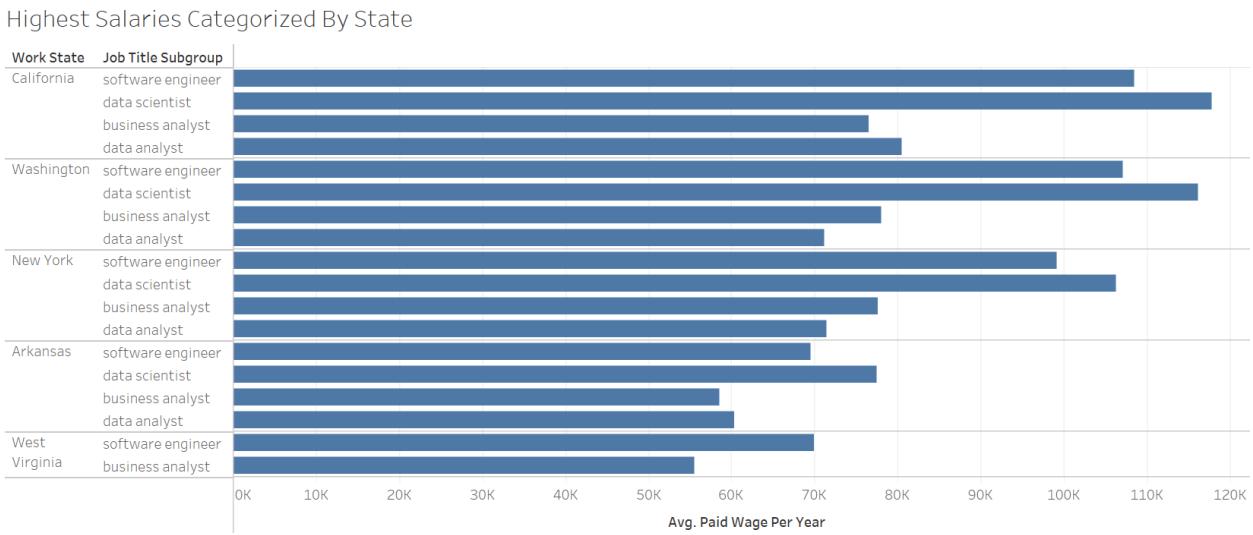
- Objective One -

The first objective we explored was how certain salaries compare to each other depending on the specific type of job. In this report, we focused on sub-type jobs related to data to cater our analysis in a more specific field. In our data set, the 4 sub-type jobs related to data include: business analyst, data analyst, data scientist, and software engineer. We wanted to compare their yearly salaries to each other to see which jobs have higher or lower salaries. We found that on average, data scientists have the highest salaries, followed by software engineers. Data analysts overall have the lowest salaries but not too far apart from business analysts.

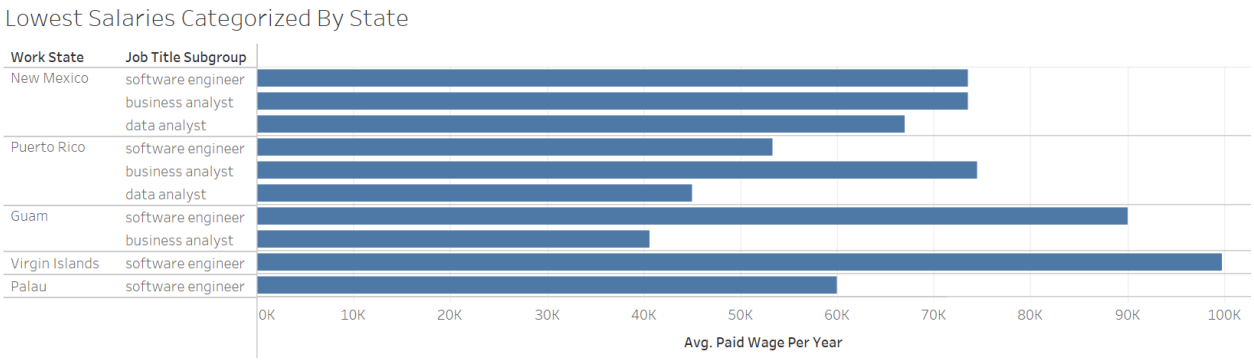


There are many different factors that can affect the salaries of a job. One factor being the location of the job. When looking into locations of jobs that pay the most, we found that California, Washington, New York,

Arkansas, and West Virginia were the states that paid these jobs the most, California being the state that pays the most. This makes sense as the top 3 states: California, Washington, and New York have one of the highest cost of living in comparison to other states. With a higher cost of living, it's necessary for people working in those areas to have a higher wage to be able to be located in those locations. Although different locations have varying costs of living, we found that data scientist was still the job that has the highest salaries. This result changes once we start looking into places that have overall lower salaries.

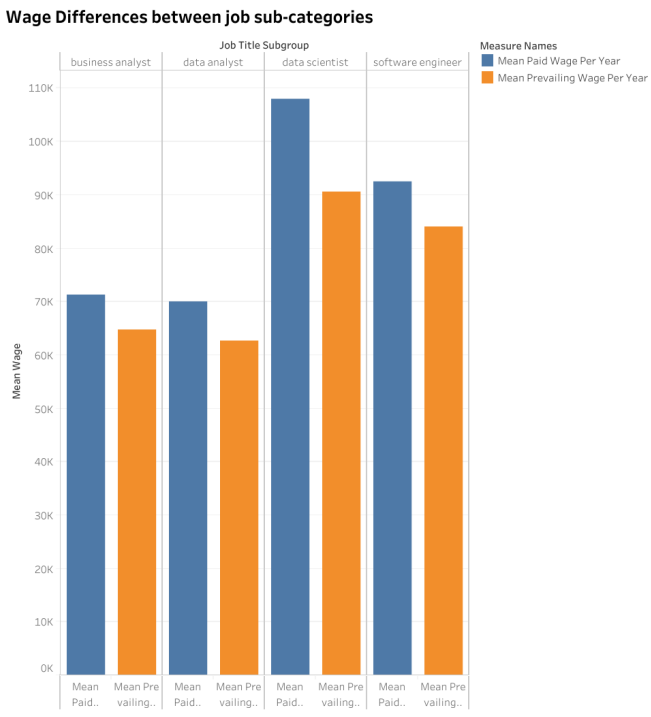


We included locations outside of the United States including US territories to see if there would be a big difference in salaries. The locations being evaluated were: New Mexico, Puerto Rico, Guam, Virginia, and Palau. We found that in all of the following locations, all did not have a category for data scientist, indicating that there are no data scientist job opportunities in those areas. We found that in the majority of these locations, software engineers had the highest salaries. This makes sense as software engineers were the second highest paid jobs after data scientists. We can assume that there are a lot more job opportunities as a software engineer for different locations in comparison to data scientists. People looking into working in the US as a data scientist may have a harder time finding a job in less "common" areas besides the major big states. This can be background as to why immigrants look into certain areas more than areas, some areas simply have more job opportunities available.



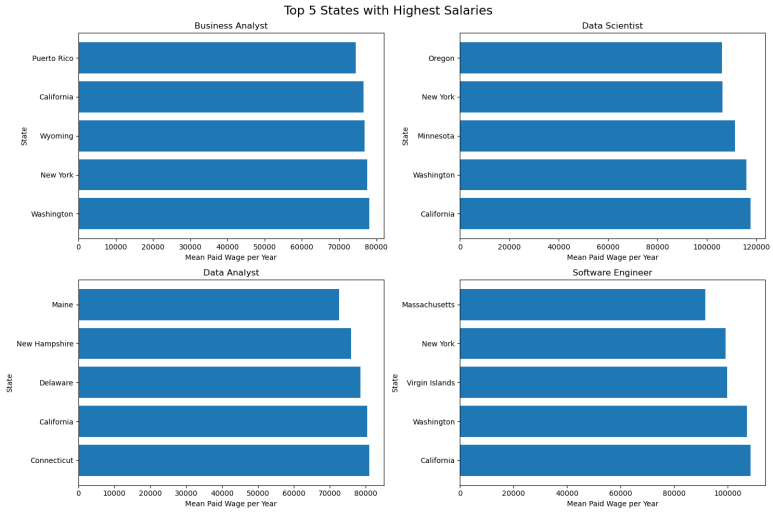
- Objective Two -

The second objective focuses on identifying the state that has the highest paying data-related salaries. In this report, we considered four job title subgroups as data-related jobs: business analyst, data analyst, data scientist, and software engineer. As we observed in the first objective, there are distinct differences in prevailing and paid wages among each subgroup. Both the prevailing wage and the paid wage indicate that data scientists have the highest salaries, followed by software engineers.



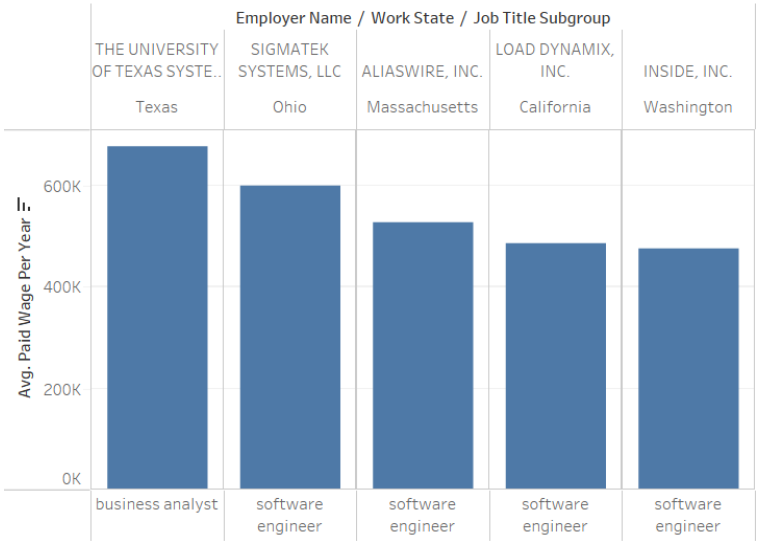
The Code-Tastic Four conducted an analysis to determine the top 5 states with the highest salaries in each subcategory. California ranked within the top 5 for all four categories, and notably, it emerged as the state with the

highest paid wage for software engineers and data scientists. Additionally, Washington and New York ranked within the top 5 for three out of the four categories, excluding data analysts. Consistent with the earlier finding that data scientists had the highest paid wage among subgroups, this graph confirms that data scientists receive the highest paid wage.



We introduced a histogram that illustrated the companies offering the highest salaries, where they are located, and for what position. All companies present here are not what someone would normally identify as a competitive company in terms of salary. The University of Texas States' average paid wage per year topped the graph at 600,000 dollars, whereas the companies ALIASWIRE and SIGMATEK show wages paid in the 500,000-600,000 dollar range.

Top 5 Companies with Highest Salaries



The Code-Tastic Four looked at the top 10 states that the hightest salaries for tech-related jobs when taking into account the Cost of Living. It was not surprising to see that states in the West harboured the highest salaries in regards to salary. The Virgin Islands and New York trail behind, showing numbers that compare relatively close to California and Washington. When we take in the Cost of Living into account, we find that Washington leads alongside states that were not present in our previous findings.

Top 10 States with Highest Salaries

Work State	
California	105,501
Washington	103,456
Virgin Islands	99,788
New York	90,709
Massachusetts	87,168
Oregon	83,450
Vermont	81,495
Maryland	81,135
Utah	80,210
Nevada	79,926

Top 10 States with Highest Salaries (after normalization)

Work State	
Washington	927.0
Alabama	858.1
Kansas	825.8
Louisiana	821.6
Georgia	821.5
Texas	817.4
Utah	810.2
Mississippi	809.6
New Mexico	804.8
Missouri	804.1

- Objective Three -

Objective three asks a series of questions, with the main question being " How do offered salaries compare to the prevailing wage?".

We first wanted to look at a generic view of the question before answering the follow-up questions in the task. The question is looking for a difference between offered salaries, which we determined was the table "Prevailing Wage Submitted", and the prevailing wage by job sub-group, which was determined to be "Paid Wage". To make our data present our findings by year, we grouped the Date table by year and achieved the following:

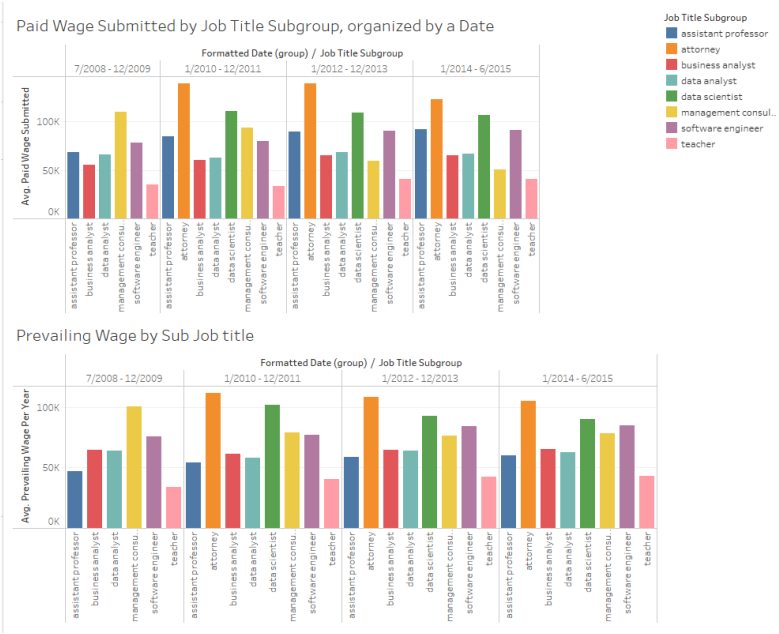


Figure 3-1

Figure 3-1 not only shows the differences in what was submitted and actually earned, but that certain sub-groups like attorneys and data scientists do not appear to until 2010, which we can conclude that factors like demand come into play, resulting in higher instances and prevalence to the dataset.

The first question to be answered from Objective three was "Are there job sub-categories that tend to get over-paid/under-paid?". We altered the parameters so that if prevailing wage submitted and the prevailing wage was equal, it would still count as being under-paid. The following table shows that though the instances, teachers and software engineers have a high chance of being overpaid by 39%, with software engineers trailing behind at 28%. Business Analysts shows the most occurrences of under-pay at 77%.

Figure 3-2

The Code-Tastic Four filtered the our salary dataset to display our findings for the next secondary objective, where it was asked about the top 5 companied that both under-pay and over-pay. We found that through our principle, it can be said that FANG companies (Google and Facebook) tend to either underpay, or give offers satisfactory to employees, on the otherhand, Fujitsu and Intel show staggering numbers when it comes to the instances of overpay, reaching well over 1000.

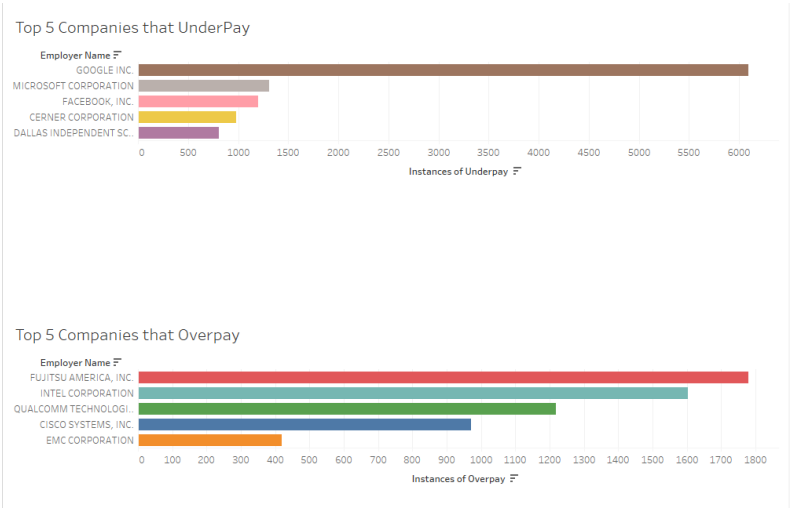


Figure 3-3

Finally, we now take a look at the final portion of our project, where we are asked to report on the affect of Cost of Living to salaries. We found that like our initial look into the objective, Teachers seem live as comfortably as those within tech jobs. We can attribute this to different living circumstances and an individuals preference to live.

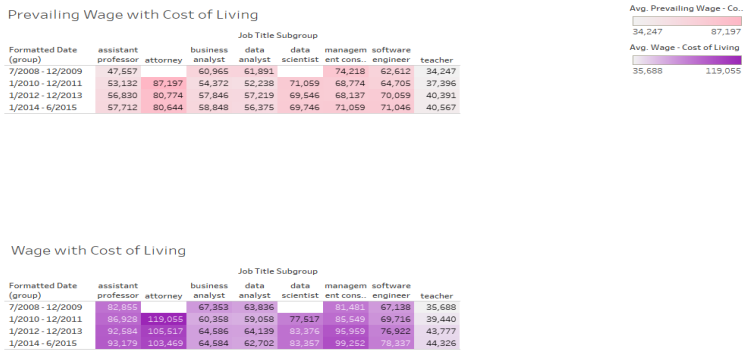


Figure 3-4

Conclusion

In Summary, we found that certain jobs that applicants apply to are weighed heavily on where they are located geographically and what applicants can expect in terms of the pay companies offer to accommodate the requirements for these applicants to be productive.

There are many known companies asking for tech roles. Netflix, as an example show the highest rate for data scientists, whereas, Intuit boasts near same numbers for data analysts. States in the West, such as California and Washington, are in high demand for people in these fields with an exception in New York boasting favorable numbers for those looking to start their careers in certain states. Tech jobs do not see much prevalence in territories owned by the U.S, with New Mexico leading the results as the lowest paying U.S. state for tech applicants.

An applicant in the tech field can find that data scientists and software engineers are attractive occupations with salaries near or above 100,000 dollars. The discrepancies in pay amongst the tech sub-groups show that data analysts show the lowest numbers as data scientists continue to tread the pack. Taking into consideration the Cost of Living, we find that Washington is the best place to work in that shows interesting numbers in the cost of living there to tech salaries.

Amongst all jobs, an attorney seems to be more lucrative than most tech jobs, with Data Scientists trailing just behind. Interestingly enough, software engineers show the most instances of overpay. It is surpsing to see some of the companies shown (i.e. Fujitsu) represented as the top 5 companies that over pay. What is interesting is that alongside Technical professions, teachers also seem to be less affected by the cost of living.

This dataset holds insight into what applicants can expect when applying for these jobs in certain aspects that affect way of life.