

Team 7 Technical Write up: Data Science Jobs

What are the top industries?

When reviewing the Indeed dataset, we extracted three specific columns related to company and industry--Company Industry, Company, Number of Stars. With the new created data frame, we are able to group by industry, then count the total number of entries per industry. Limiting the series to top five industries, we concluded that Consulting and Business Services (30.5%), Internet and Software (26.9%), Banks and Financial Services (19.1%), Health Care (14.1%), and Insurance (9.4%) fell within this limit.

What are the top companies within the industries?

Building off of the data frame used to extract the top five industries, we are able to list company name as a string. Average company rating was the integer then used the filter these companies in ascending order. We created bar charts to show to average company rating for the five companies within that industry. Harnham, Google, Capital One, Kaiser Permanete, and USAA showed the highest ratings across all industries.

Where are they hiring?

The analysis of the locations data that based on the Indeed dataset, shows California in the top states with 1395 open Data Science positions which represents 34% out of the top 10 locations, followed by New York state with nearly 15%, while the remaining states show close percentages (between 3% and 8%) for the Data Science open positions. The lead of both California and New York for the number of positions available can be explained by the number of companies that are recruiting in these two states. For example, most of data science hiring companies in California are IT ones such as Google, Facebook, Uber while those are hiring in New York are both IT and Finance firms.

What skills are they looking for?

Based on an analysis of the skills data provided within the Indeed Dataset, Python is the top rated skill required to obtain a position within the Data Science field, leading with 26% of the top 5 skills rated, amongst Data Scientist, Data Engineers and Data Analyst, as a collective. Python is followed by SQL with 24%. SQL is leading, at 33.6%, in regards to the primary language needed to obtain a position as a Data Analyst. SQL and Python are nearly even, with SQL at 38% and Python at 39%, in regards to the primary language needed to become a Data Engineer. Lastly, the top two skills needed to become a Data Scientist is Python with 33.5%, and R with 27% of top five skills required.

What salary can we expect?

Starting from the analysis and exploration, that based on the Indeed data set, there are more postings relevant to a Data Scientist position. 44% of the job postings were towards data scientist. 47% of the jobs had a salary range between 100k-140k. Diving deeper into the salary data reveals that Data Scientists and Engineers command the higher portion of the salary range distribution, while Data Analyst are in the lower portion. There may be other factors that contribute to Scientist and Engineers commanding higher salaries such as degree/experience, but one factor included in our data set is the number of skills. When comparing the average number of skills a job posting requires vs the job type, it was found that the job postings for the Data Scientist and Engineers require more skills.