The dataset of the jobs and salaries of data scientists was found on Jobs and Salaries in Data Science (kaggle.com) and analyzed through the R studio and the provided CSV file. The dataset has twelve categories: work years, job title, job category, salary currency, salary, salary in USD, employee residence, employment type, experience level, work setting, company location, and company size. Among these categories, only work years, salary, and salary in USD are numerical data, and the rest are non-numerical. These data can help analyze the trend of data scientists' salaries and other information such as the job category and distribution of experience level.

One of the most common ways to categorize employees is to group them according to their experience level. The data scientists are separated into entry-level, mid-level, senior, and executive levels. However, the general trend of the salary of data scientists is approximately the same over the past four years (2020 - 2023), which was increasing.
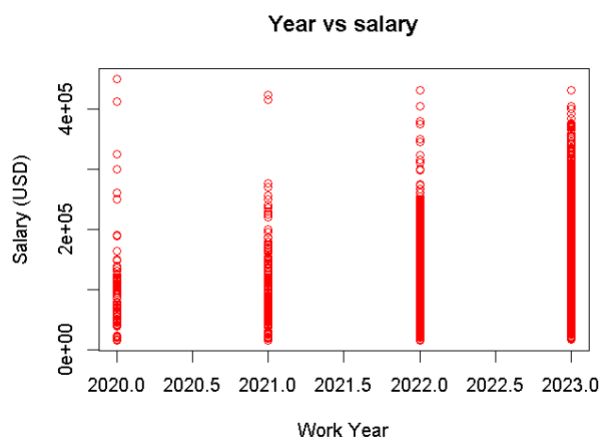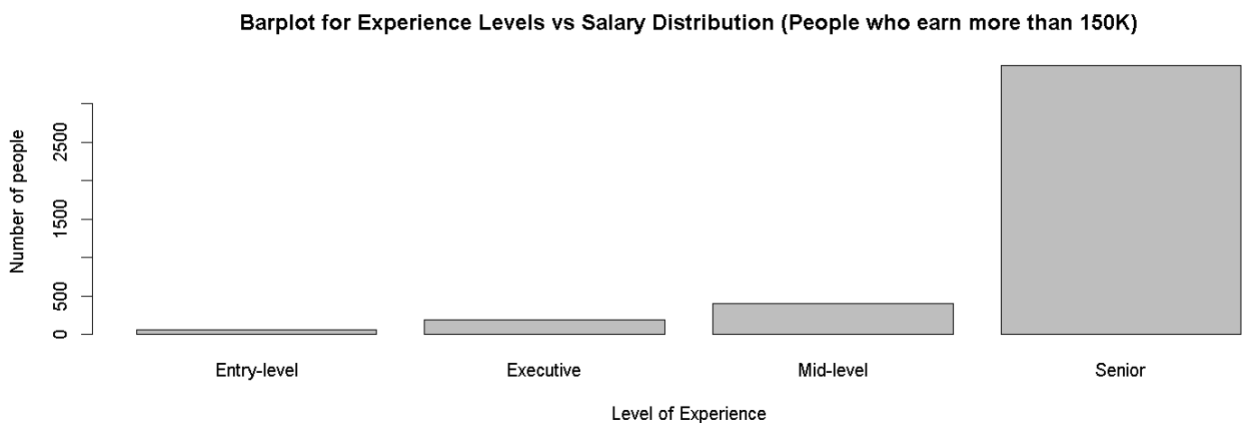

For example, the scattered plot above shows the changes in salary between 2020 and 2023, and the salary was increasing gradually. Meanwhile, more people earn more than 200K (USD) in 2023 than in the past three years. According to this graph, we cannot see any evidence showing the adverse effects of the recession due to COVID-19. However, we can see that data scientists earned the least money in 2021 when checking a single category. In this case, the average senior

data scientist made $144489.9 in 2020, $132104.4 in 2021, $148284.1 in 2022, and 165699.6 in 2023. After 2021, the salary increased because of the recovery of the economy and need and demand. The same situation happened to entry-level and mid-level data scientists. Apart from this, the salary of the executive-level data scientists remained the same in 2021 but in 2023. The average executives' salaries are $179958.2 in 2020, $186128.0 in 2021, $191059.1 in 2022, and $189496.0 in 2023. The decrease may be related to the layoffs of the tech companies recently since there are fewer executive-level data scientists, and a smaller number of decreases can lead to a noticeable change in the average salary of executives, as shown in the graph below.

Barplot for Experience Levels Distribution

Boxplot of work years vs Salary (Executive)

Boxplot of work years vs Salary (Senior)

In addition, the analysis shows that senior data scientists who work in medium-sized companies earn as much as the senior scientists who work in large companies, which are

$163240.1 in large companies, $162660.7 in medium-sized companies, and $114544.4 in small companies. Meanwhile, the average salaries of senior data scientists during the past four years (2020-2023) is $162356.13, meaning that most senior-level scientists work in medium-sized or large companies. At the same time, that can also allude to the fact that small companies are disappearing or they have fewer demands than before. This leads to a huge number of senior data scientists can earn more than other levels of data scientists on average.

Boxplot of company size vs Salary (Senior)

Barplot for Experience Levels vs Salary Distribution (People who earn more than 150K)

Moreover, we can predict the salaries of different levels of data scientists using the data we have for the past four years. We can know their experience levels by checking their salaries, and

the accuracy rate is approximately 72.1%: if the salary is below 35000, then it might be entry-level; if between 35000 and 68200, it might be mid-level; if between 68200 and 319000, it might be senior data scientist; if above 319000, then it might be executive. Since the average salary of the executive is close to that of senior data scientists, which is $189462.91 and $162356.13, respectively, it may be hard to find the exact range of executives by looking at the salary. In addition, there are more senior data scientists than executives, which makes it harder to predict the range.

In conclusion, the dataset provides information about data scientists' salaries, work settings, work years, etc. The patterns, such as a decrease in salary in 2021 and an increase in 2022, can be analyzed by comparing the number of producing plots. The rest of the analysis can be found in the R file attached after importing the data from Kaggle.com.

Work cited:

https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data?resource=download