LAB 17

Gift Ajayi

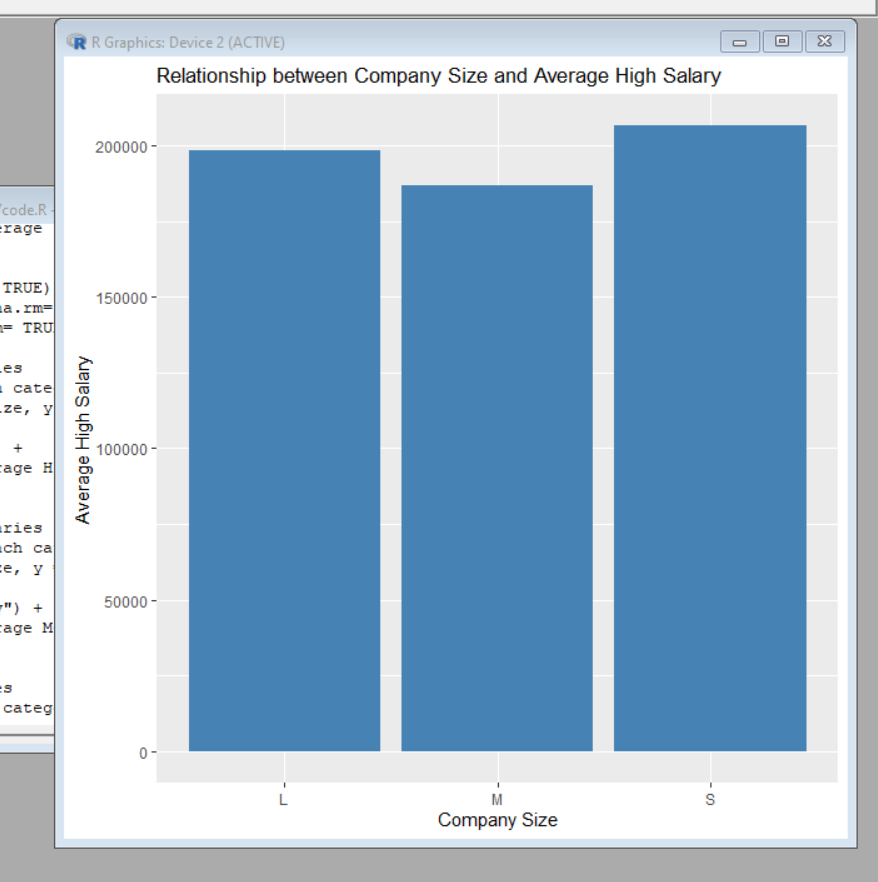
Computation For Scientific App

July – 10 – 2023

For this analysis, I used a dataset obtained from Kaggle, which provides information on the salaries of different types of data scientists based on their roles and the companies they work for. Initially, I wrote a Python code to calculate the percentiles of salaries in USD. I then created a new column to determine the salary level for each value in USD, considering whether it falls within the range of the three major percentiles (25th, 50th, and 75th). Additionally, I created three new columns, namely low salary, medium salary, and high salary, which assign the corresponding values from the salary in USD column to their respective ranges.

After adding the necessary tables, I exported the data as a CSV file and converted it into an Excel file for further analysis in RStudio. In R, I imported the data and conducted an analysis by calculating the averages of the low, medium, and high salaries. To ensure accurate calculations, I used "na.rm = TRUE" to handle any empty spaces in the data. I then grouped the averages by company sizes.

Next, I created plots where I visualized the relationship between company size and the averages of low, medium, and high salaries. Upon examining all three graphs, I reached the conclusion that small companies tend to be higher-paying employers compared to the other two company sizes. Additionally, medium-sized companies have a greater number of low-paying salaries. Regarding medium salaries, there seems to be minimal variation across all company sizes.



A screenshot of a computer screen

Description automatically generated

A screenshot of a graph

Description automatically generated

References:

Randomarnab. (2023, April 13). *Data science salaries 2023 💸*. Kaggle. https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023?resource=download

YouTube. (2023, February 1). *Learn R in 39 minutes*. YouTube. https://www.youtube.com/watch?v=yZ0bV2Afkjc&ab\_channel=EquitableEquations

YouTube. (2018, October 25). *Complete python pandas data science tutorial! (reading CSV/Excel files, sorting, filtering, Groupby)*. YouTube. https://www.youtube.com/watch?v=vmEHCJofslg&ab\_channel=KeithGalli