PROJECT REVIEW

AIMS & OBJECTIVES

The aims and objectives of this project was centered around predicting which types of jobs are most likely to be automated by Artificial Intelligence (AI). This was to be achieved by finding a pattern among professions that have a high probability of AI automation. A suitable dataset – Probability of AI Automation in 2030 – was sourced from Kaggle.

Through rigorous data analysis, a pattern was identified by using the values of mean salary in the workplace and an index representing exposure of AI tools in that field. The initial hypothesis of this project stated a correlation between automation probability and the individual's mean salary and level of education. This was further verified by plotting several bar charts. It showed that the dataset did not have an equal distribution of people having different years of experience working in the field. Furthermore, the level of education impacted the mean salary of the individual, but no further relation could be made with respect to automation probability.

Therefore, another parameter was considered to measure probability of AI automation, namely the index of AI exposure in the workplace. By using these values, several line plots were made to find patterns in automation probability. From these graphs, some definitive conclusions were drawn; mean salary and AI exposure did affect the probability of automation by AI.

DATA ANALYSIS TECHNIQUES

After finding a suitable dataset, there were several transformations done on to remove outlier data. The dataset was checked to find any duplicated values and null values. These values were then removed from the dataset.

To check for outlier data, a statistical method was used, known as the z-score. It measures how many standard deviations the particular data-point lies outside the mean range of the dataset. The calculation for z-score is done by implementing a statistical library in python, scipy.stats which has a predefined function to calculate the z-score.

PROJECT ANALYSIS

During the preliminary analysis stage, a bar chart was drawn representing the variations in salary by the years of experience. This bar graph was, however, unhelpful as there was no consideration to the different kinds of professions present in the dataset. From these results, it was more practical to identify trends in a singular profession. By querying the dataset, twenty unique professions were found to be available.

The profession of a financial analyst was taken, and the trends in salary by years of experience was checked again. From this graph, it was made evident that there were an unequal number of people with different years of experience. This was further verified by comparing salary to educational level.

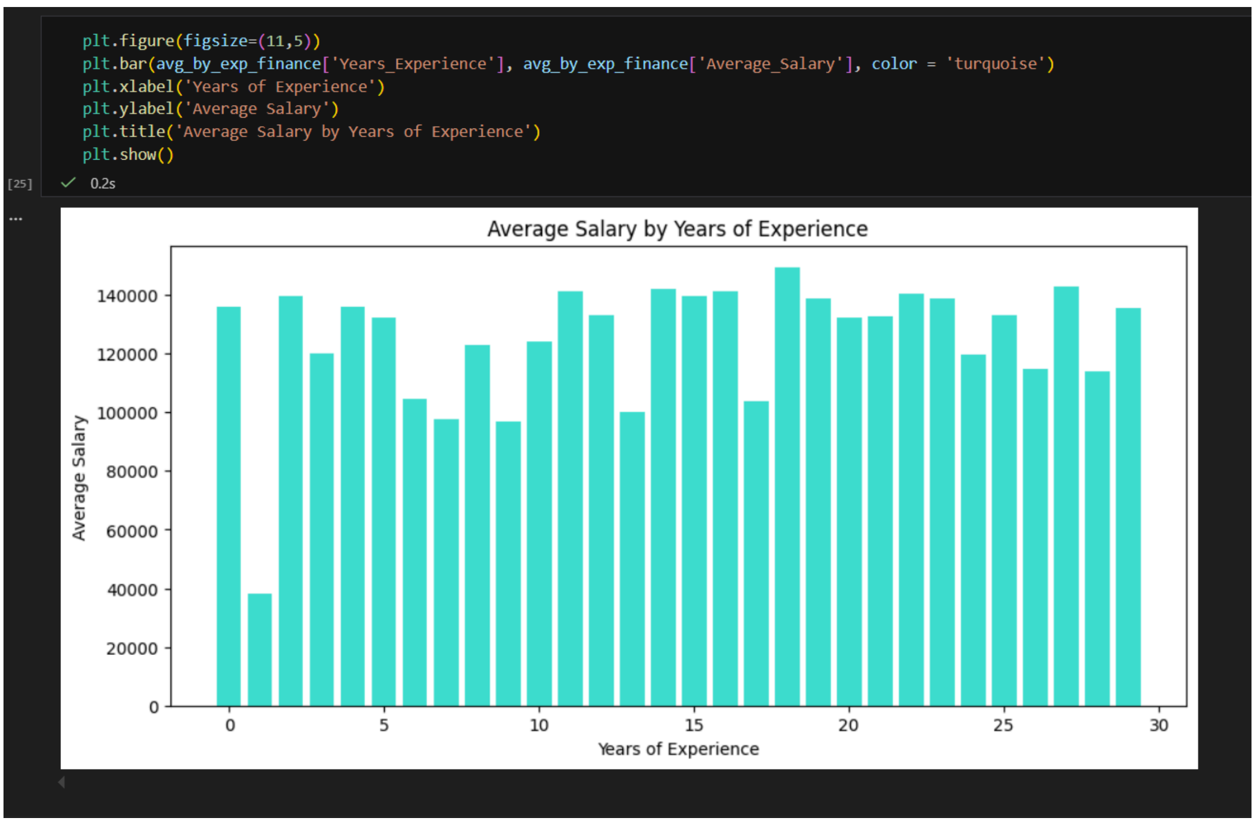

Figure 1. Bar Graph representing average salary by years of experience for a Financial Analyst

```
filtered_finance['Education_Level'].value_counts()
```

```
Education_Level
Bachelor's     56
High School    32
PhD            32
Master's       31
Name: count, dtype: int64
```

Figure 2. Code snippet showing the frequency of people with different educational qualifications

From these results, it was clear that identifying trends across different professions would be more efficient. In the next stage of analysis, mean salary and AI exposure index were plotted against job titles. These graphs were compared to the line graph representing automation probability vs. jobs, to find which factors affected the AI automation probability for the future. The detailed analysis process led to the conclusions that both AI exposure index and average salary both influenced the probability of jobs being automated by AI in the future.

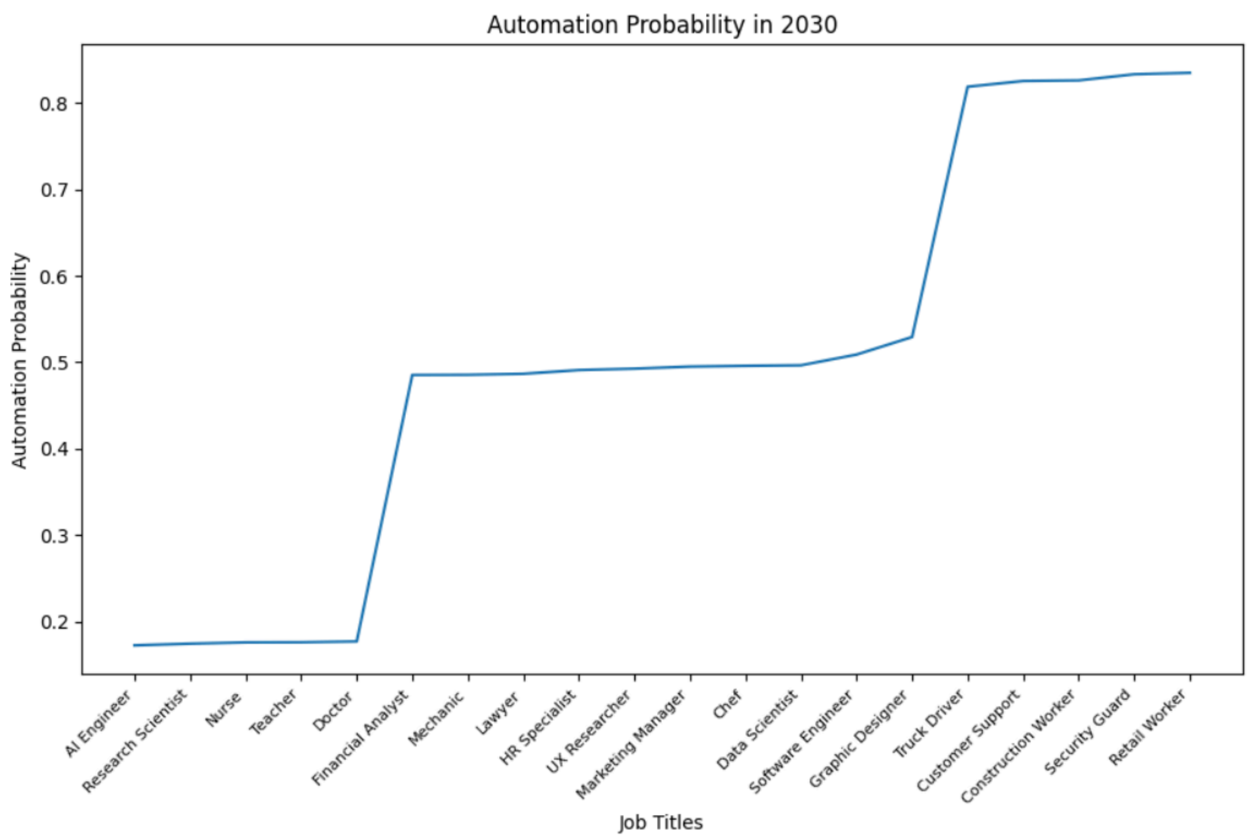

Figure 3. Line graph depicting AI automation probability vs. professions

FUTURE DEVELOPMENTS

It was analysed that the education level of a person affected their salary per annum. Similarly, the relationship between mean salary and automation probability could be further studied. Within each profession, different people can have different work assigned to them based on their educational levels. If salary is affected by the level of education, then it is entirely possible that automation probability and mean salary are closely related.

The dataset also contained values for ten different skills in each profession for every individual. This data was not relevant to find patterns in automation probability, however it can be analysed to find skills essential for each profession.

```
plt.figure(figsize=(8,6))
sns.heatmap(jobs_by_skill)
plt.ylabel(None)
plt.show()
```

0.5s

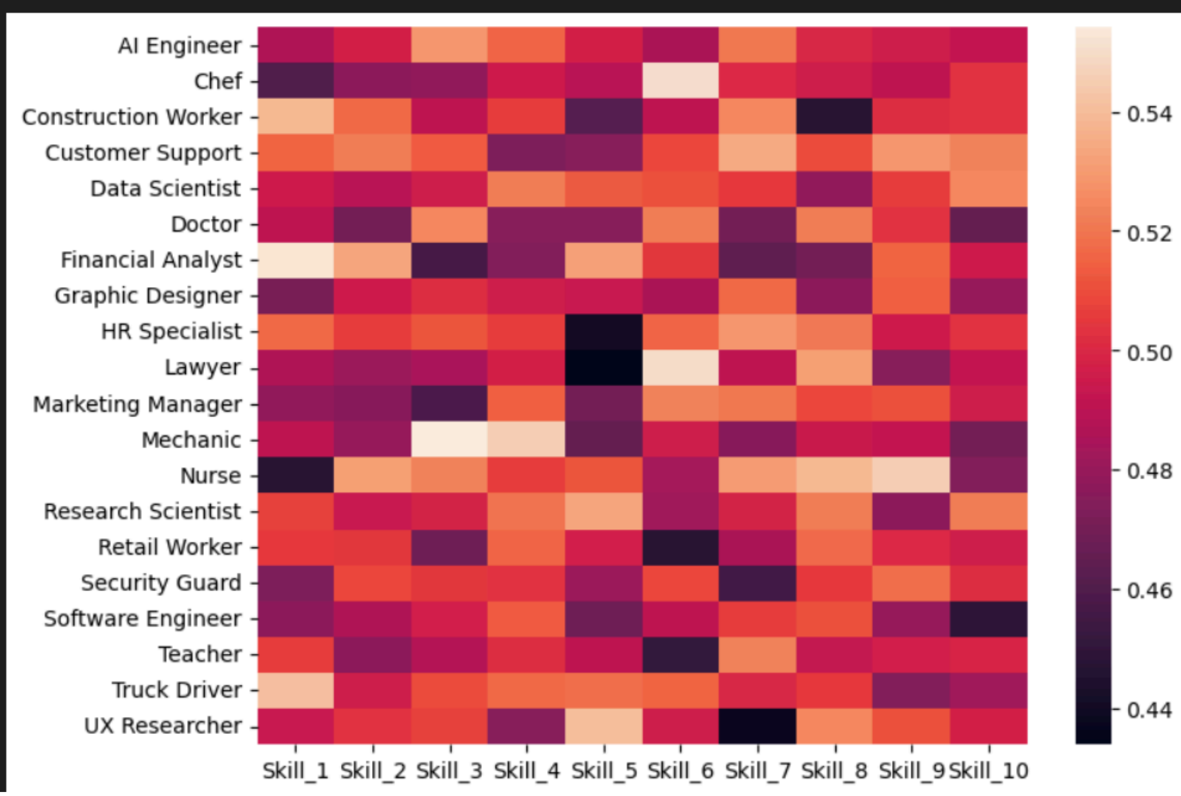

Figure 4. A heatmap showing the proficiency of skills in each profession