GICS SECTOR PERFORMANCE

3/1/2023

SUMMARY

By utilizing three datasets, data preparation was conducted to ensure a clean dataset for visualizing the data. Some of the transformations conducted to the dataset included reformatting data, converting dates to date time, dropping duplicates, splitting columns, and replacing values that were outliers or incorrect. These datasets were then loaded into individual tables in a SQLite database. While preparing to load the data, it was found that the website source had significant changes to the Wikipedia table since the previous milestone. Adjustments were made to ensure that all variables were collected appropriately.

After three tables were created, a merged table was created with all three sources for the purpose of data visualization. Certain columns were selected to organize the data and to prevent duplicate columns. The data was merged to connect to the S&P 500 records that were to be analyzed for the GICS Sector performance. The final dataset included 500 records.

The first visualization created showed the revenue change by GICS sector as a box plot. The end results showed the distribution of percentage revenue change for each sector and potential outliers included. The next visualization utilized 30 year increments (triennial) for the year that the S&P 500 company was founded to determine if the date the company was established had any pattern on how the average profit percent changed. It was found that company that were established 2010 or after had the highest negative decrease in profit. The third visualization shows a scatterplot between assets and market value to determine relationship. There was a lot of variability with the chart, but most companies were in the bottom left-hand corner. The fourth visualization showed a histogram on the closing value of the company stock. Most stock prices were less than $400. The final visualization showed a scatterplot of the year the company was founded and the number of employees. Most companies had 500,000 or less employees.

ETHICAL IMPLICATIONS

All datasets were sourced from independent organizations and are subject to change by those institutions. Historical financial and stock data are provided by the companies listed in the S&P 500 and are subject to ethical practices or could face legal issues. After merging the data, it was found that there were missing values for some companies from the various sources. The visualizations created were subject to the data that was available after the merge.