**CS 210 Project Phase I – Ege Tan - 30977**

**Description:**

>In the past, electric vehicle manufacturers such as Tesla have benefited from rising oil prices. Higher oil prices increase the operating costs of traditional combustion engine vehicles, making EVs more financially attractive. This dynamic suggests that interest in electric vehicles may increase during periods of high oil costs, which could boost Tesla's value due to expected sales increases. However, there are other important factors beyond direct economic factors that influence the relationship between oil prices and Tesla's market success. One of these factors is market sentiment. Rising oil prices can boost investor confidence in alternative fuel sources and renewable energy, which can create a more positive outlook for businesses like Tesla. Such positive sentiment can lead to increased share values even without the immediate financial performance of the company. In this context, I aim to delve deeper into the complex relationship between Brent crude oil prices and Tesla's share performance, focusing on the years 2019 to 2022. The motivation for this comparison stems from the evolving dynamics in the energy and automotive industries, particularly as the global economy shifts towards electric vehicles and renewable energy. By examining oil price fluctuations in combination with Tesla's market capitalization and trading volumes, I aim to uncover valuable insights into broader economic and environmental trends. The datasets underpinning the project, consisting of annual Brent crude oil prices and Tesla's share prices and trading volumes, provide a glimpse into the historical trends of the global oil market and reflect Tesla's stock market performance.

**Explore Data:**

Data Loading and Initial Processing: I started with loading two datasets to notebook, which one containing Tesla's stock prices and another one featuring Brent crude oil prices over a specified period.

Data Cleaning: Early steps involve cleaning the data, which includes handling missing values and ensuring data types are appropriate for analysis (converting dates from strings to a datetime format).

Data Merging: There is a step that merges the two datasets on the date column so that each date has both Tesla's stock price and the corresponding oil price. This merged dataframe allows me direct comparisons and further analysis between these two variables.

Descriptive Statistics and Visualization: I included steps from recitation 6 and recitation 9 for calculating descriptive statistics to understand the central tendencies, dispersion, and shape of the distributions of the data. The steps include plotting time series graphs or histograms to visually assess trends and fluctuations in Tesla's stock price relative to oil prices.

Data Analysis: Correlation analysis and regression models, to explore the relationships between oil prices and Tesla's stock prices. This helped in understanding whether there's a statistically significant association between rising oil prices and Tesla's market valuation.

Interpretation and Conclusion: Final sections involve interpreting the results from the statistical tests and models, discussing the implications of the findings, and possibly suggesting areas for further research or the limitations of the current analysis.

**Explanation:**

A screenshot of a graph

Description automatically generated  
>By opening us stock market data in google collab, I deleted the unrequired columns for my hypothesis. Moreover, there were some Na values, after merging my both data according to their dates. I used ‘bfill’ method of pandas that fills the Na’s with bellow row values. I also deleted some unmatched date rows when I merged two datasets. First dataset had so many columns but the ones that I need were Date and Tesla\_Price columns, that is why I deleted all other columns from my dataset. Moreover, my second dataset was including only Date and Price columns, I did not change anything from second dataset. Afterwards I merged them according to dates with outer join. Then as I mentioned before, the nan values occured because of unmatched dates in both datasets. I filled these NaN valued rows by values from previous rows. As a result, I successfully merged two datasets.

Change int tesla stock prices from 2019 to 2022:

A green line graph with numbers

Description automatically generated

Change in brent oil prices from 2019 to 2022:

A green line graph with numbers and a line

Description automatically generated

Frequency of brent oil prices from 2019 to 2022:

A graph of a graph

Description automatically generated with medium confidence

Frequency of Tesla stock prices from 2019 to 2022:

A graph of a number of blue bars

Description automatically generated with medium confidence

**Formulate Hypothesis and perform Hypothesis testing:**

* In the past, electric vehicle manufacturers such as Tesla were gaining from increased oil costs. The rising cost of oil is making traditional combustion engine vehicles to be more expensive to operate, which increases the appeal of electric vehicles from a financial standpoint.
* After merging two datasets according to their dates, some null values occurred. I filled these values by ‘bfill’ method in python. In that way I handled with null values in merged dataset.

Correlation graphs for Tesla stock price and brent petrol price:

A graph of different sizes of graphs

Description automatically generated with medium confidence

Null Hypothesis (H₀): There is no relationship between Brent oil prices and Tesla's stock price; any observed correlation is due to random chance.

Alternative Hypothesis (Hₐ): There is a positive relationship between Brent oil prices and Tesla's stock price, suggesting that increases in oil prices lead to increases in Tesla's stock price.

From hypothesis tests that we covered in recitation the most adequate one is Pearson value for my hypothesis so I 'll consider the p-value for the regression coefficient to determine whether to reject the null hypothesis. If the p-value is less than the typical alpha level of 0.05, we'll reject the null hypothesis in favor of the alternative.

>> PearsonRResult (statistic=0.5788910855374884, pvalue=7.132760243490819e-84)

As the correlation coefficient, I obtained 0.57 which states that the correlation between brent petrol prices and Tesla stock prices is positive. From the p-value, less than the significance level (0.05), we can see that our test is significant.

Interpretation: The p-value is extremely small (far less than 0.05), which provides strong evidence against the null hypothesis. Therefore, I reject the null hypothesis and accept the alternative hypothesis that there is a positive relationship between Brent oil prices and Tesla's stock price. This suggests that increases in oil prices are associated with increases in Tesla's stock price. This result confirms my earlier correlation and regression analysis, reinforcing the hypothesis that higher oil prices may enhance the financial appeal of electric vehicles, such as those produced by Tesla.

**Build a single linear regression model for prediction:**

Perform a simple linear regression to further investigate the relationship, with oil prices as the independent variable and Tesla's stock performance as the dependent variable. While building a single linear regression model, I also calculated some additional values such as r square value, mean squared error (mse), error rate and accuracy rate.

R-squared (R²) value (0.3351): This value explains how much of the variability in Tesla's stock price can be explained by changes in oil prices. An R² of 0.3351 suggests that approximately 33.51% of the variability in Tesla's stock price can be accounted for by the model that uses oil prices as a predictor.

Mean Squared Error (MSE) (9188.3200): This measures the average squared difference between the observed actual outcomes and the outcomes predicted by the model. An MSE of 9188.3200 suggests that the model's predictions are, on average, 9188.3200 units squared away from the actual value of Tesla's stock price.

Error Rate (0.6649): Typically, the error rate is used in classification problems to indicate the proportion of incorrect predictions, indicating the proportion of the total variation in the dependent variable (Tesla's stock price) that is not explained by the independent variable (oil prices). This implies that about 66.49% of the variation in Tesla's stock price is not explained by changes in oil prices.

Accuracy Rate (0.3351): Like the error rate, this metric is generally used in the context of classification. It indicates that about 33.51% of the variation in Tesla's stock price is explained by the model.

Relation graph of Tesla Price and Brent oil price:

A graph showing a price

Description automatically generated with medium confidence

Regression graph of Tesla Price and Brent oil price:

A red line with blue dots

Description automatically generated

Conclusion: These values suggest that while there is some relationship between oil prices and Tesla's stock prices, a significant portion of Tesla's stock price movements is influenced by factors other than oil prices. This could motivate further investigation into other variables that might better explain the fluctuations in Tesla's stock price.

Dataset Link 1: https://www.kaggle.com/datasets/saketk511/2019-2024-us-stock-market-data

Dataset Link 2: https://www.kaggle.com/datasets/mabusalah/brent-oil-prices