# CS 210 Project Step:2

Yağız Uçar 31204

Stock Market Dataset:

      Dataset includes the prices and trading volumes of various financial assets such as natural gas, crude, oil, copper, Bitcoin, Berkshire, Netflix, Amazon, Meta, and gold. It contains dates along with relevant prices and volume data. Financial analysis, academic research, behavioral finance, developing financial products and machine learning and data science applications could be given as an instance of possible areas of usage of this dataset.


Bitcoin Trading Volume from 2022-01-01 to 2023-06-22
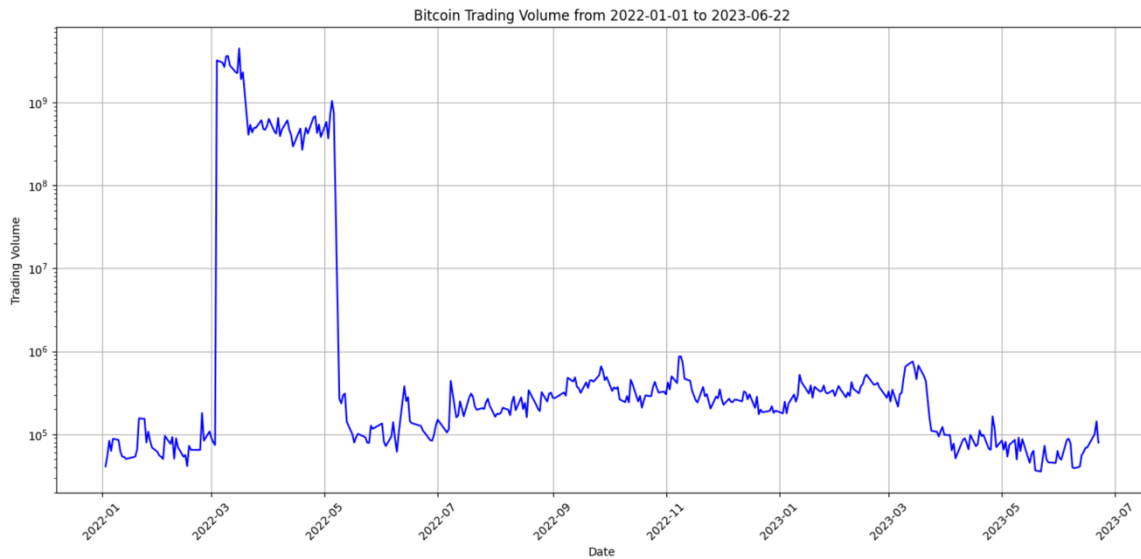
- The graph shows the changes in the Bitcoin trading volume from the dates between 2022/01/01 and 2023/06/22.
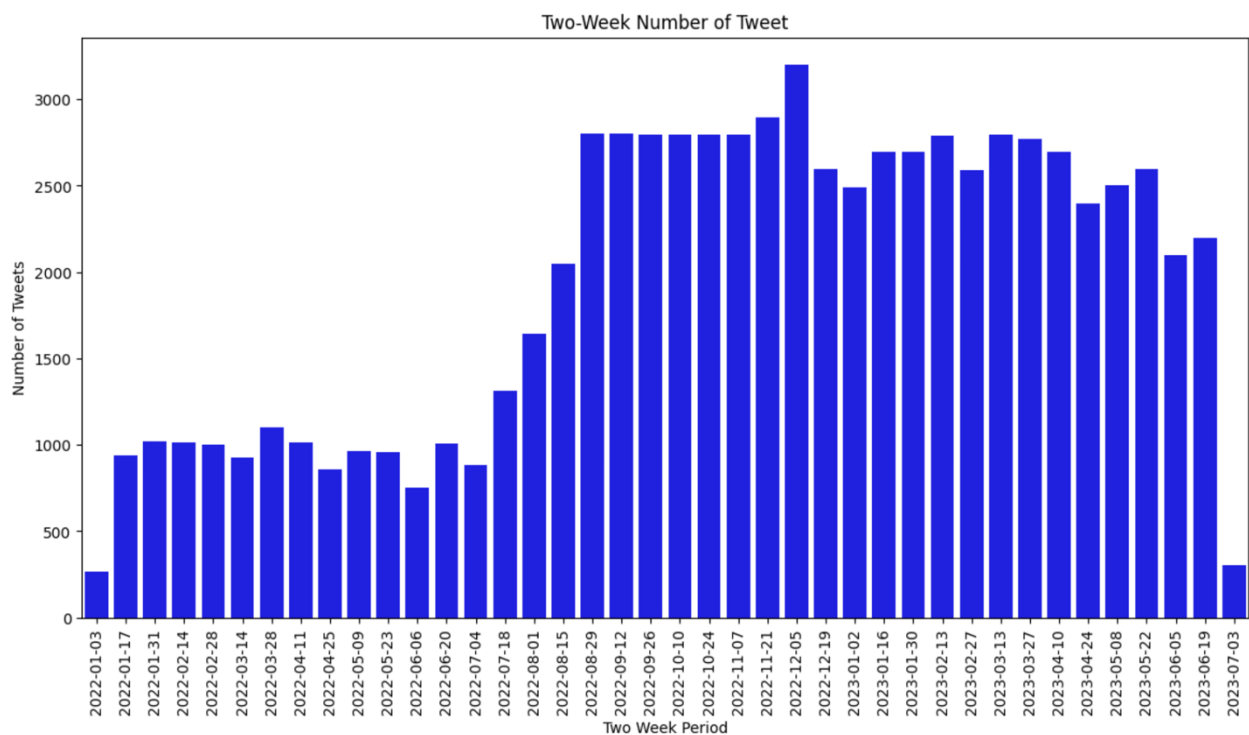
Bitcoin Price from 2022-01-01 to 2023-06-22
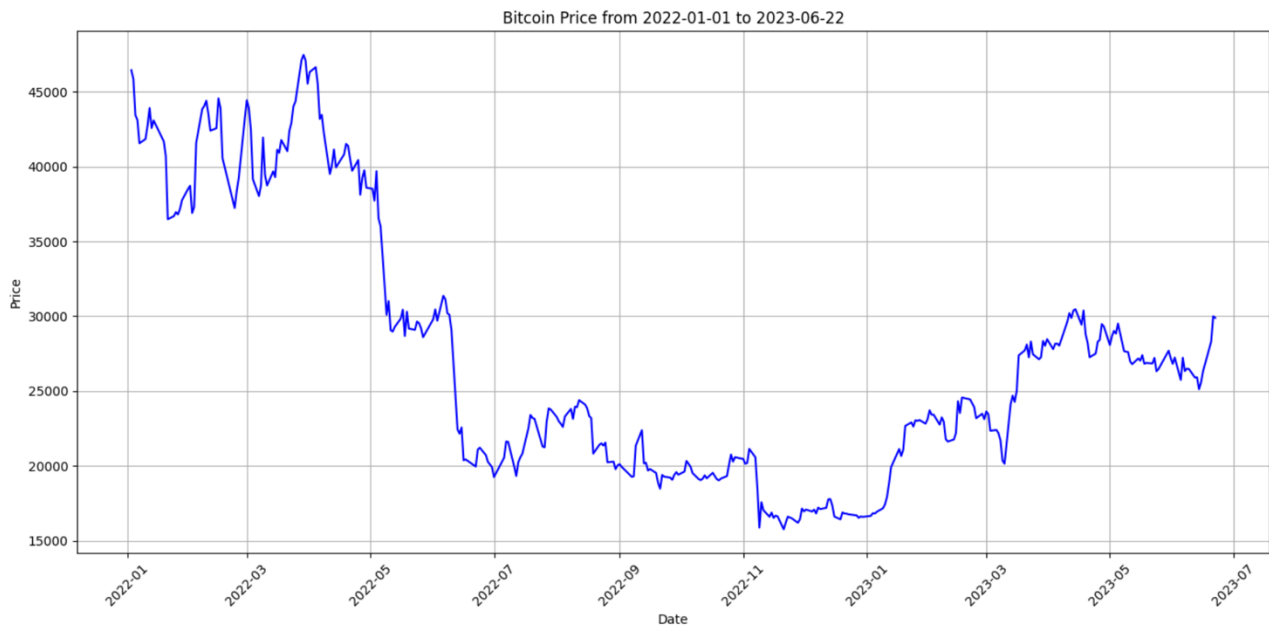
- The graph shows the changes in the Bitcoin prices from the dates between 2022/01/01 and 2023/06/22.
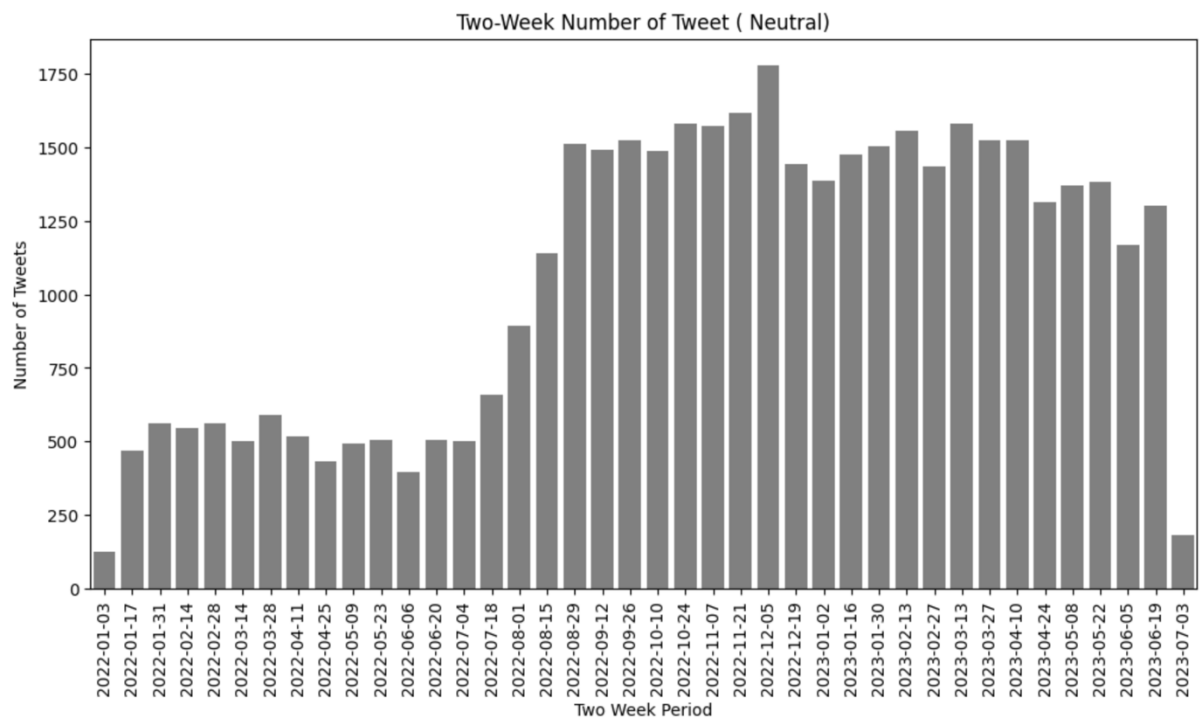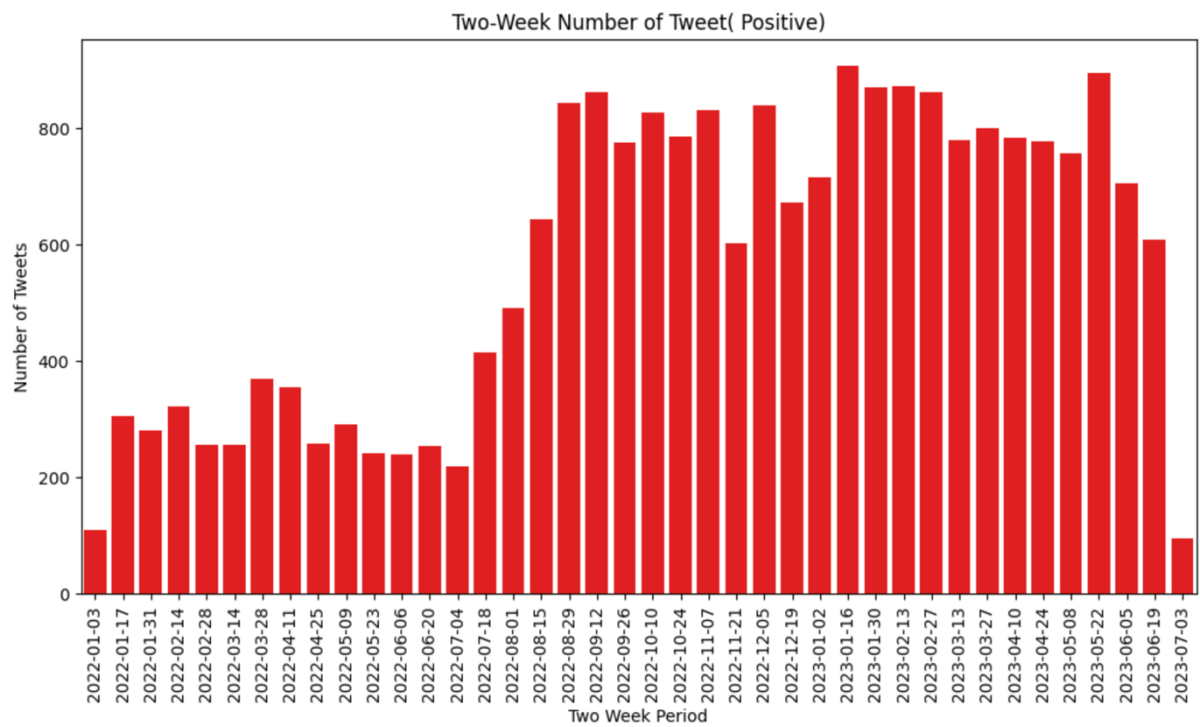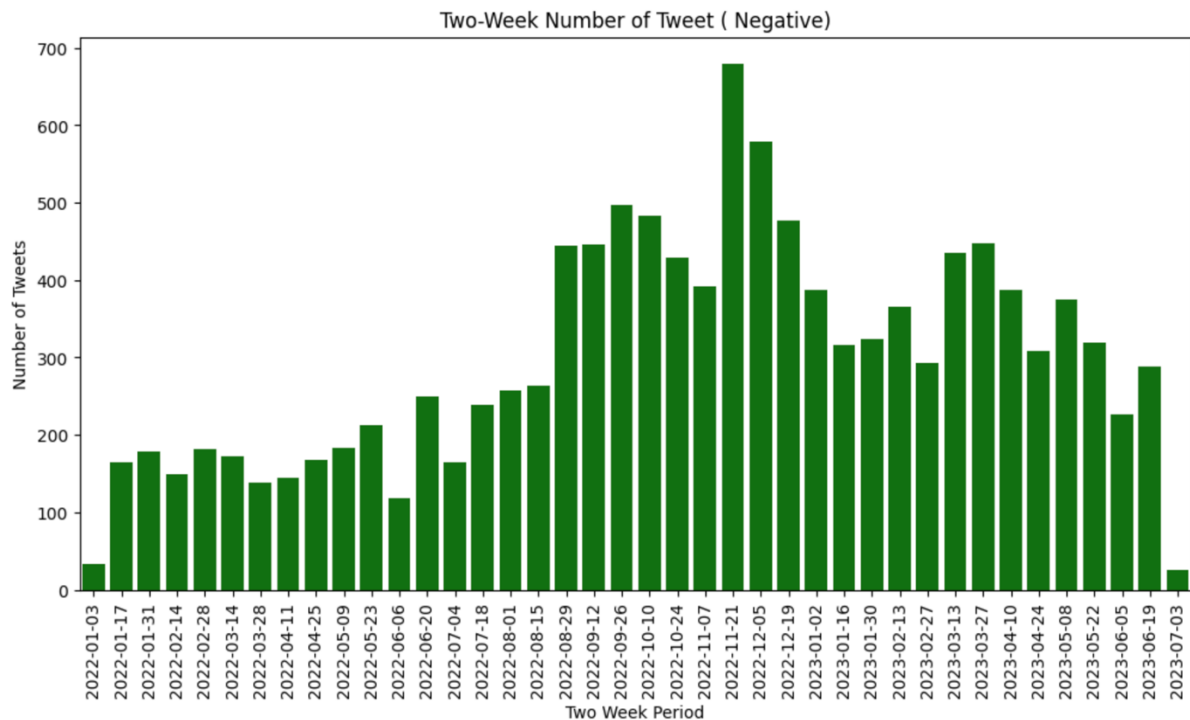
Social Media Dataset:

Dataset includes the present an analytic complication of tweets collected from various users. Moreover, it includes not only the tweets but also details such as the number of likes, retweets, quote counts, and sentiments analysis labels. In dataset sentiment analysis is categorized as a "Positive", "Negative" and "Neutral" for every tweet. In the data set content of tweets provides rich source of data reflecting the sentiment and conversational trends on social media. This dataset can be utilized to understand social media dynamics and user interaction.

Tweet Type
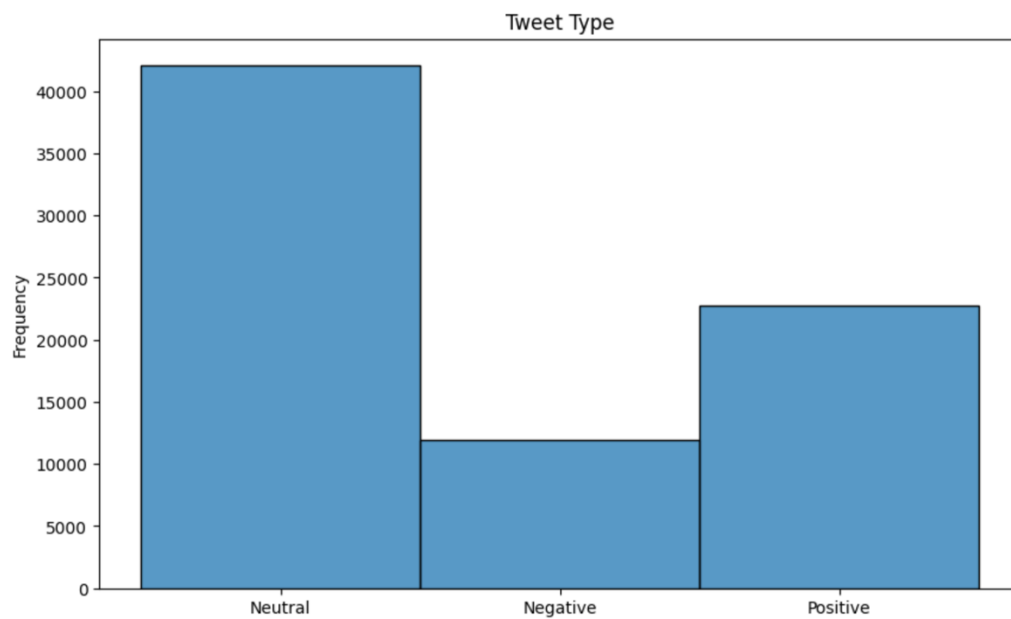
- The graph shows the frequency of neutral, positive, and negative tweets which are related with the Bitcoin.

Two-Week Number of Tweet

Two-Week Number of Tweet(Positive)

Two-Week Number of Tweet (Neutral)

Two-Week Number of Tweet (Negative)

- These 4 graphs above divide the dates between 2022/01/01 and 2023/06/22 into 2-week periods and shows the number of negative, positive, neutral, and total number of tweets about bitcoin during these 2-week periods.

Formulating Hypothesis and Hypothesis Testing:

The graphs which were visualized based on the datasets lead to understanding the information contained how datasets will used during the hypothesis testing process.

Formulating Hypothesis:

- Null Hypothesis (H0):

  Social media activities (tweets) have no effect on both Bitcoin's price and trading volume.

- Alternative Hypothesis (HA):

  Social media activities (tweets) have effect on both Bitcoin's price and trading volume.

Hypothesis Testing:

- Finding Correlation Matrix:

In this step the correlation between Bitcoin Price, Bitcoin trading volume, and the total number of tweets from 2022/01/01 to 2023/06/22, was analyzed. After, merging the datasets, a correlation matrix was created, and heatmap was generated.


Correlation Matrix Heatmap
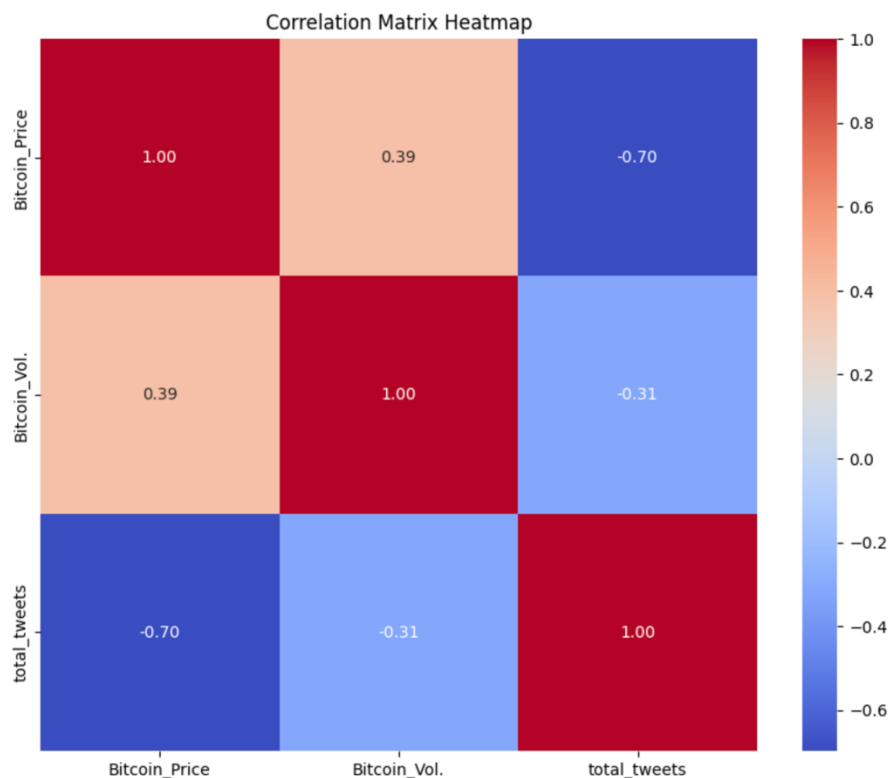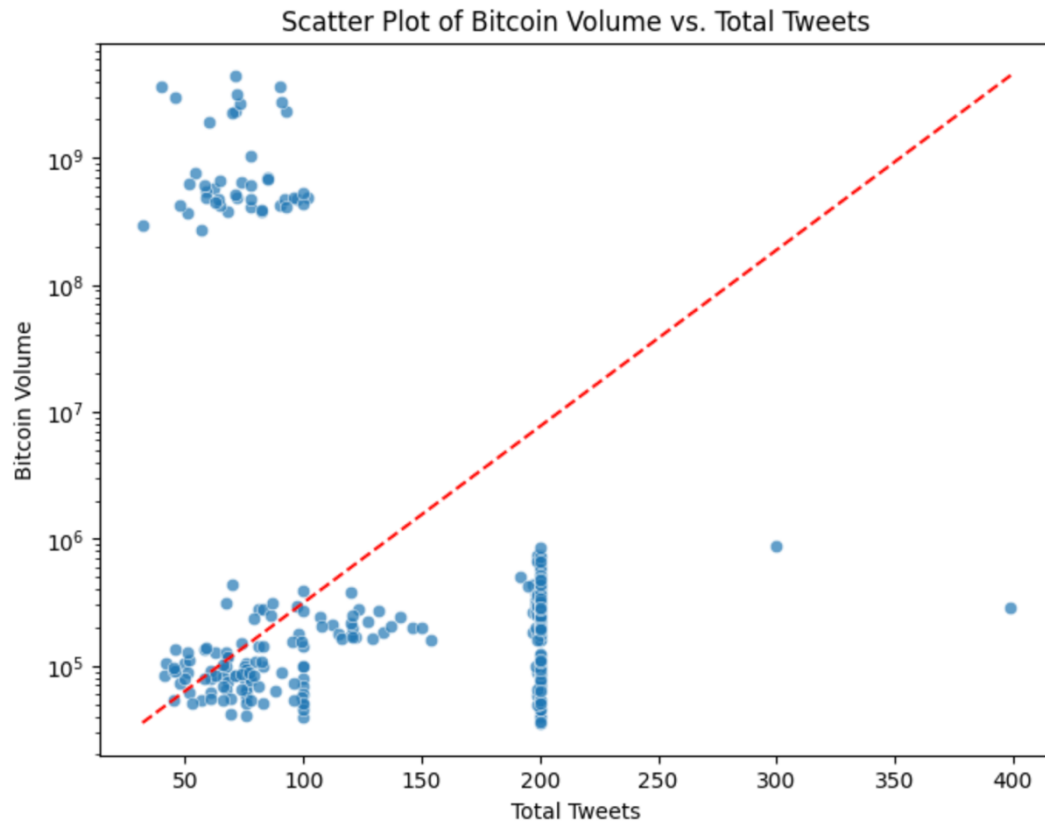
The result showed that:

- A strong negative correlation was found between Bitcoin Price and Total number of tweets. (-0.70)
- A negative correlation was found between Bitcoin Trading Volume and total number of tweets. (-0.31)

- Correlation Test:

Correlation test was conducted to examine the relationship between "Bitcoin price and Total number of tweets" and "Bitcoin trading volume and Total number of tweets" from 2022/01/01 to 2023/06/22.

A scatter plot was created to visualize both bitcoin price and number of tweets relationship and bitcoin trading volume and number of tweets. Correlation coefficient shown and p-values were calculated.

Scatter Plot of Bitcoin Volume vs. Total Tweets
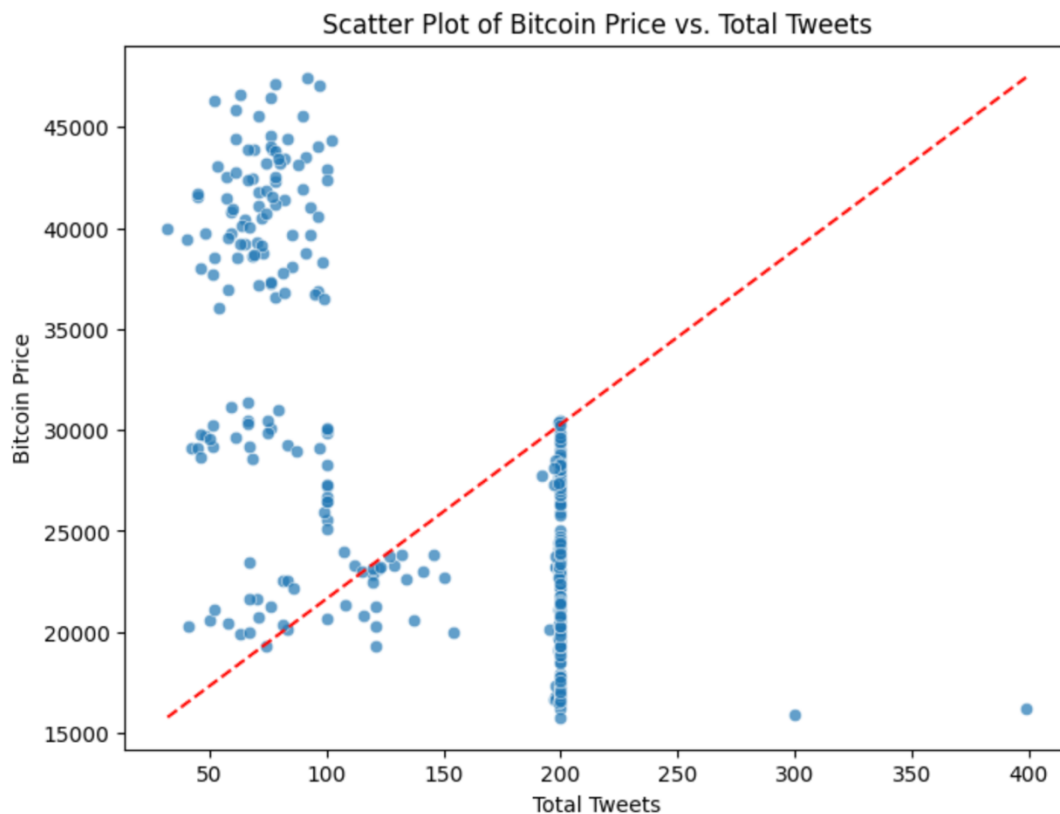
The result showed that:

- Both p-values were calculated less than the 0.05 threshold.
- The null hypothesis is rejected.
- Strong negative correlation between Bitcoin Price and number of tweets.
- Negative correlation between Bitcoin Trading Volume and number of tweets.

Linear Regression Model for Prediction:

Combining Bitcoin price and number of tweets data, a regression model was created. A "Linear Regression Model" was used to make a prediction for bitcoin price based on number of tweets. Mean Squared Error (MSE) was calculated, and prediction result visualized.

Linear Regression: Prediction of Bitcoin Price based on Total Tweets
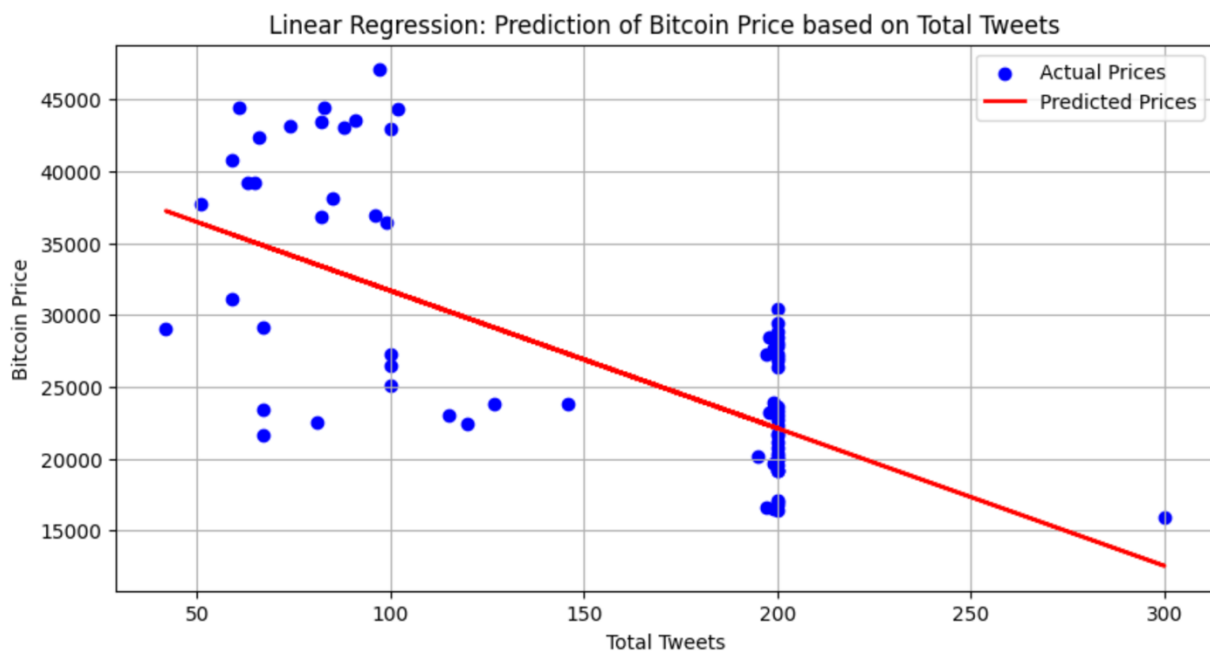
Combining both "Bitcoin price and number of tweets" and "Bitcoin Volume and number of tweets" data, a regression model was created based on dates. Mean Squared Error (MSE) was calculated, and prediction result visualized.

Prediction of Bitcoin Volume Based on Days and Total Tweets

Prediction of Bitcoin Price Based on Days and Total Tweets