Om Prajapati, Benjamin Adams, Siddakk Chatrath

Professor Sloan

CS 4100

4 December 2024

Sentiment Analysis of Elon Musk's Tweets and Their Impact on Stock Prices Using Linear Regression

Introduction

For our final project, we decided to analyze the sentiment of Elon Musk's tweets and investigate how his tweets mentioning his companies impact their respective stock prices. We were curious to know whether positive or negative messages from a prominent figure like Musk could influence stock movements in a measurable way. This question intrigued us because understanding how stock prices respond to external influences would provide valuable insights for investing.

Elon Musk was chosen as the focal point of this study due to his large audience on Twitter and his frequent references to the companies he manages. In 2022, Musk acquired Twitter and rebranded it as X; however, for simplicity, this platform will be referred to as Twitter throughout this paper. The inspiration for this project came from the paper "Sentiment Analysis with NLP on Twitter Data" by Hasan, Maliha, and Arifuzzaman. This paper highlights Twitter as a rich source of data for sentiment analysis due to its character limit and high volume of tweets. It also provides a framework for collecting, processing, and analyzing tweet sentiment, which gave us valuable insight for the methodology of our project.

By combining sentiment analysis of Musk's tweets with stock price data for Tesla, Twitter, and SpaceX, we aimed to identify whether Musk's public statements cause a measurable shift in the stock price of his companies.

Methods

Our analysis of how Elon Musk's tweets affect the stock market involved three primary steps: data collection and preprocessing, sentiment analysis, and correlation of sentiment metrics with stock prices.

Data Collection and Preprocessing

We utilized datasets sourced from Kaggle, including:

- A dataset containing all of Elon Musk's tweets in 2022.
- Datasets containing historical stock price data for Tesla and Twitter.

The tweets dataset was preprocessed to remove irrelevant content, such as emojis and special characters. The timestamps of the tweets were aligned with U.S. trading hours to ease with future stock analysis.

Sentiment Analysis

The sentiment analysis was conducted using TextBlob, a Python NLP library. Using TextBlob we assign two sentiment metrics for each tweet:

Polarity: Ranging from -1 (negative) to 1 (positive).
Subjectivity: Ranging from 0 (objective) to 1 (subjective).
These metrics were calculated for all tweets, then subsets of the entire dataset were created by the company for stock analysis.

Correlation with Stock Prices

To evaluate the relationship between tweet sentiment and stock price movements we decided to consider two possible cases, tweets which are posted during trading hours, and tweets posted outside of trading hours.

During trading hours: Price changes during trading hours were calculated as the difference between the closing and opening prices on the same day.
Outside of trading hours: For tweets posted after trading hours, price changes were calculated as the difference between the next day's opening price and the current day's closing price.

Using these calculations we populated a new dataset which contained tweets, their polarity, and their corresponding stock change. Then we used Linear regression to measure the correlation between the polarity and the stock price changes for Tesla and Twitter. The slope of the line of best fit represents the sensitivity of stock prices to changes in tweet polarity.

Alongside these methods we created visuals to represent our findings. For our NLP analysis we created interactive scatterplots for each company. These scatterplots display the Polarity and Subjectivity of a tweet and when a point is hovered over it shows the actual tweet itself. For the stock analysis, we created line graphs with lines of best fit to illustrate how the polarity of Musk's tweets correlates with stock price changes. These visuals make it easier to observe any potential relationships, highlighting trends or patterns in the data that could indicate how tweet sentiment affects market performance.

Results

The first focus of our approach was to create visualizations that could reveal potential correlations between the polarity and subjectivity of Elon Musk's tweets and the stock prices of his managed companies. To begin, we created a scatter plot of subjectivity versus polarity for all tweets in our dataset, displayed below as Figure 1. This plot offers an overview of the sentiment and subjectivity trends within the dataset, serving as a foundation for further analysis of how these metrics might be important. This visualization helps us understand the sentiment landscape of Elon Musk's tweets. Several general trends can confirm this as a relevant and usable dataset and idea. The polarity values span the entire range from -1 to 1, indicating that Elon Musk's tweets include highly positive, highly negative, and neutral sentiments. A clustering of points near the center (polarity = 0) suggests that many tweets may have a neutral sentiment. Subjectivity values also vary significantly, with tweets ranging from completely objective to highly subjective. This indicates that Musk's tweets are a mix of factual statements and personal opinions. There appears to be a denser concentration of points in the range of 0.4 to 0.6 subjectivity and around 0 polarity. This may reflect a tendency for many of Musk's tweets to lean toward neutral sentiment while being moderately subjective. There are , however, outliers with extreme polarity and subjectivity, which might represent the presence of some highly opinionated and emotionally charged statements that could potentially impact stock sentiment.

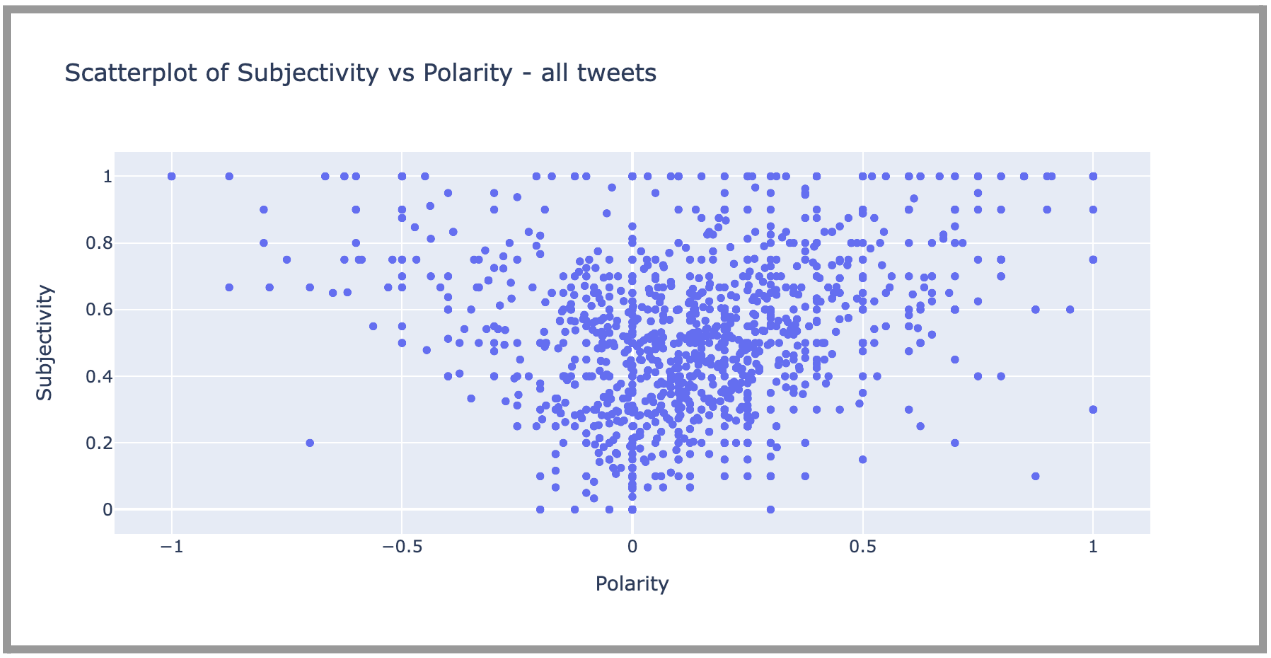

Figure 1

To further understand these trends, a visualization for subjectivity vs polarity of tweets was constructed for each company to see if any patterns could be uncovered that were unique to a company. Figure 2 illustrates the scatter plot for Twitter only, demonstrating a wide range of subjectivity values and a slightly higher concentration of positive polarity around 0.0-0.4. This would indicate that perhaps sentiment about the company in these tweets ranges more from neutral to slightly positive.

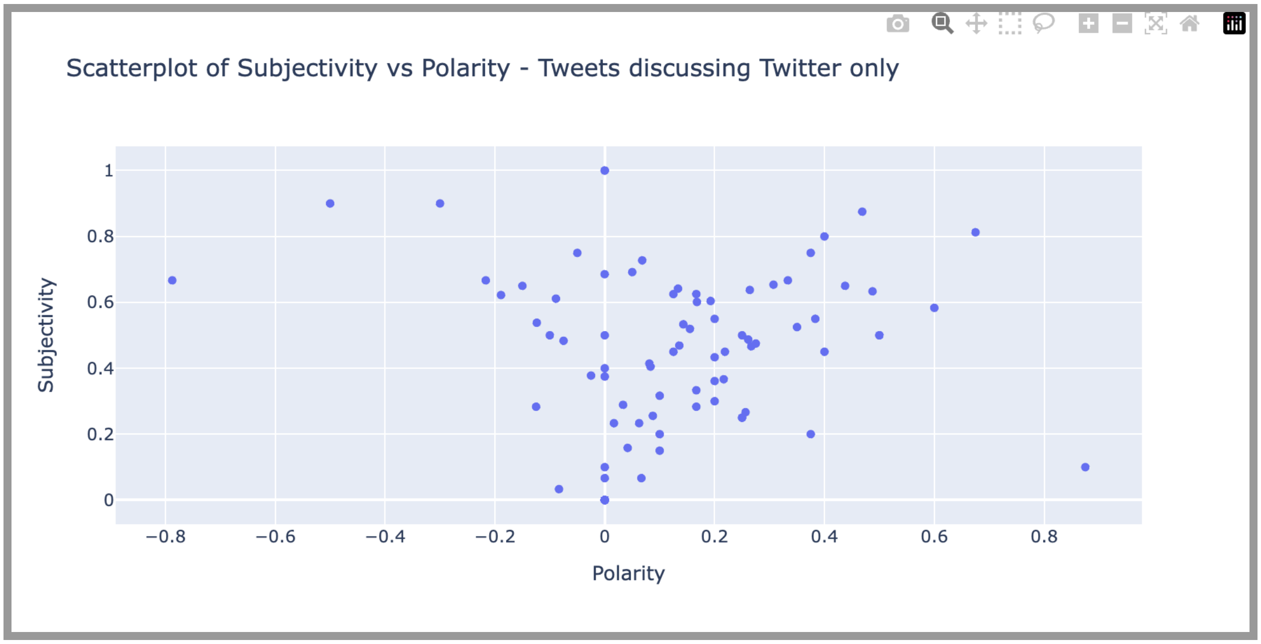

Figure 2

For SpaceX (Figure 3), the sentiment distribution displays a different pattern. The polarity remains broadly distributed with a concentration between 0.0-0.4 (which is positive), the subjectivity values tend to concentrate around 0.5 to 1.0. This clustering indicates that SpaceX discussions are often more opinion-driven, reflecting excitement or criticism regarding its technological achievements and goals. Additionally, many of the tweets mentioning SpaceX in this time frame are related to monetary support for Ukraine and/or Musk's political beliefs, which affects both subjectivity and polarity. Positive outliers with high subjectivity highlight enthusiasm for SpaceX milestones, whereas negative outliers likely point to concerns about missed targets or controversies.

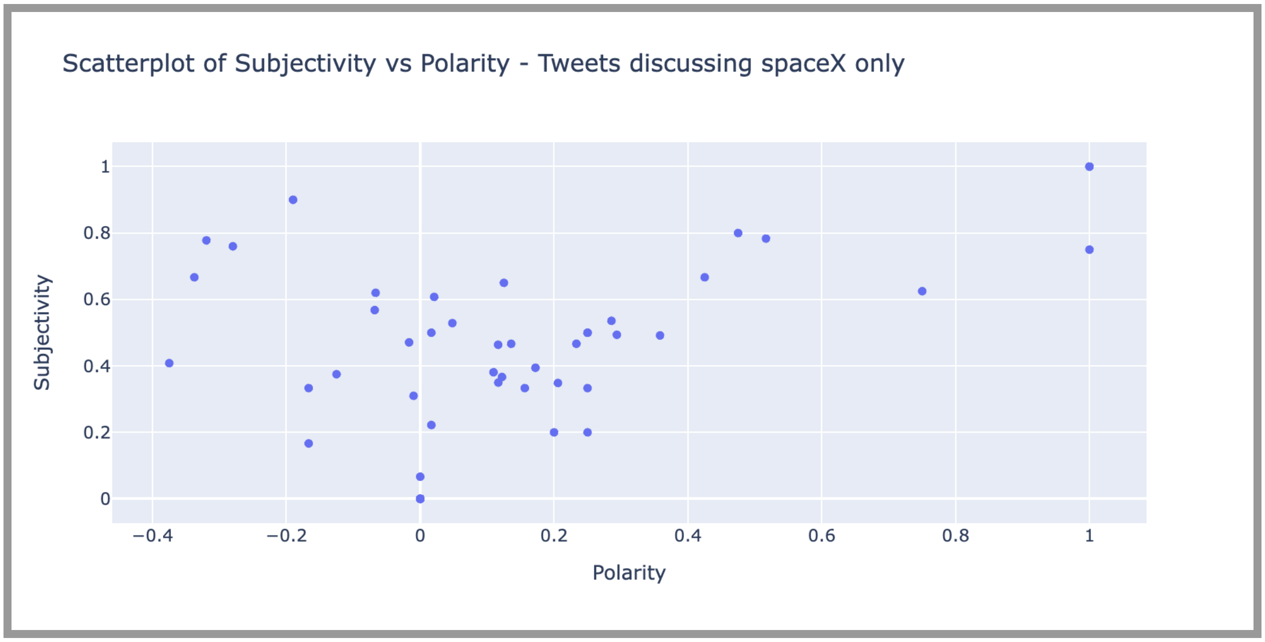

Figure 3

Tesla (Figure 4) also has a unique pattern compared to the other two companies. The polarity of the tweets is more overwhelmingly positive, even though it still clusters more in the 0.0-0.5 range. The subjectivity of the tweets varies greatly across the whole spectrum. This would seem to indicate that Musk has a lot to say about Tesla, both opinionated and objectively, and tweets more positively about Tesla.

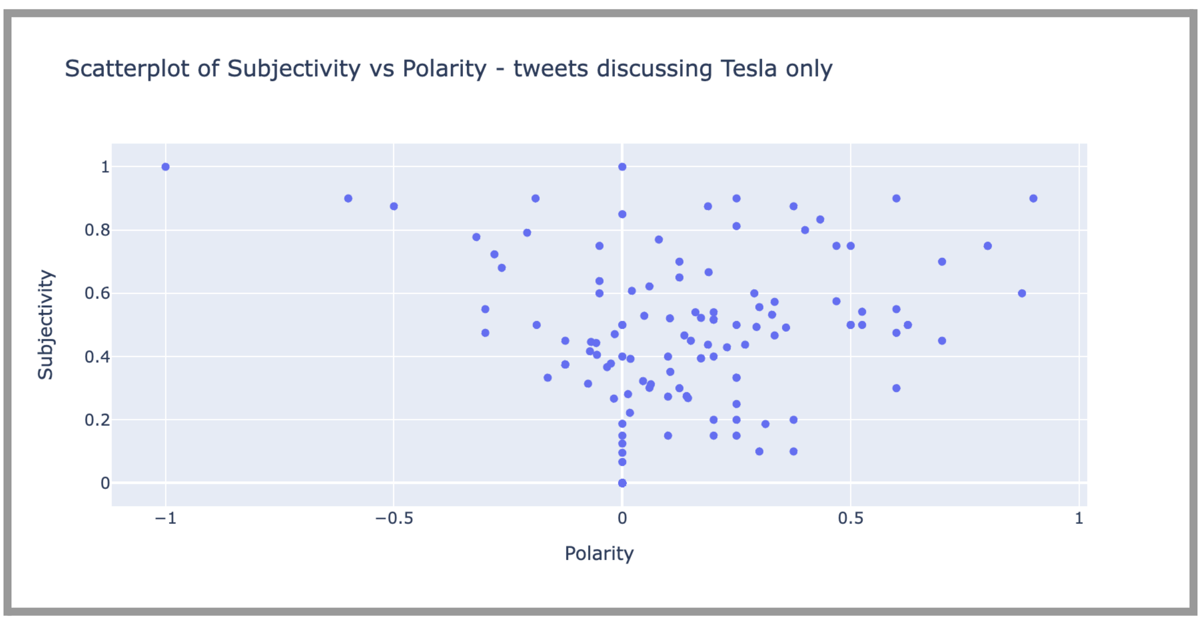

Figure 4

Now that the sentiments of Elon Musk's tweets were examined to demonstrate a variety of polarity and subjectivity values, the next focus was to investigate how the polarity of Musk's tweets may correlate with the stock price of each company. In the case of Twitter (Figure 5), there seems to be little to slight correlation, with the line of best fit of having a slope of 0.243 in a scatter plot of Twitter stock price vs tweet polarity. This would indicate that the sentiment of Musk's tweets centered around Twitter, which as seen before range from neutral to slightly positive, have little positive correlation or affect on Twitter's stock price according to our results.

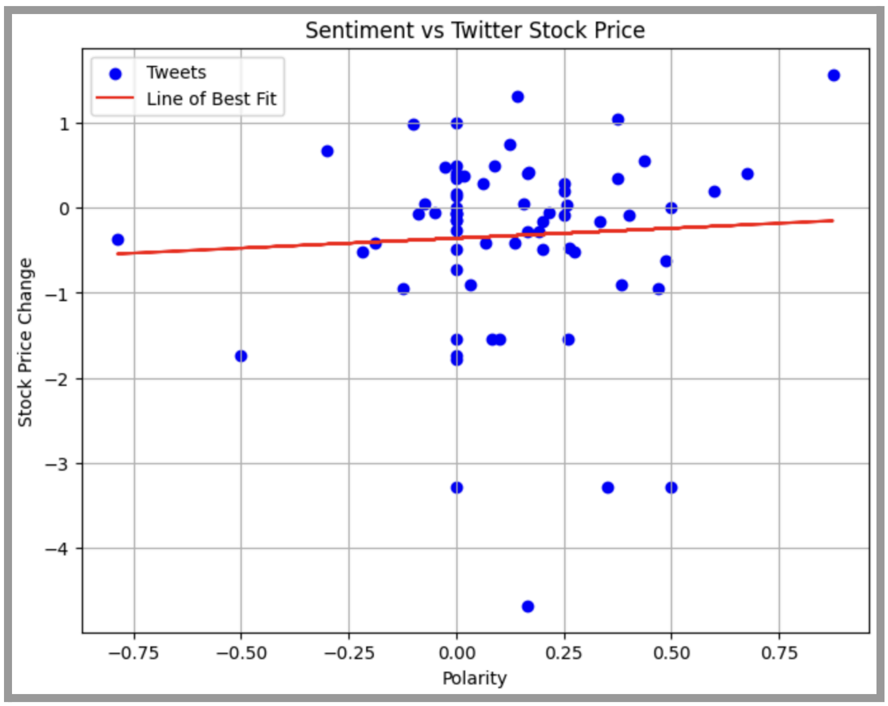

Figure 5:

Slope = 0.243

Tesla's stock price vs tweet polarity produces a very interesting visualization (Figure 6). The line of best fit actually demonstrates a negative correlation, with a slope of -3.146. However, it should also be noted that there is noticeable volatility in Tesla's 2022 stock price, fluctuating beyond even a large $40 range. This indicates the presence of larger factors contributing to stock price that would have a larger effect than tweets from the company's CEO, and as such any correlation between the two may be hidden by these larger effects. For example, as a whole Tesla stock and the stock market in general were on a downward trend in 2022, which is where our dataset is from.

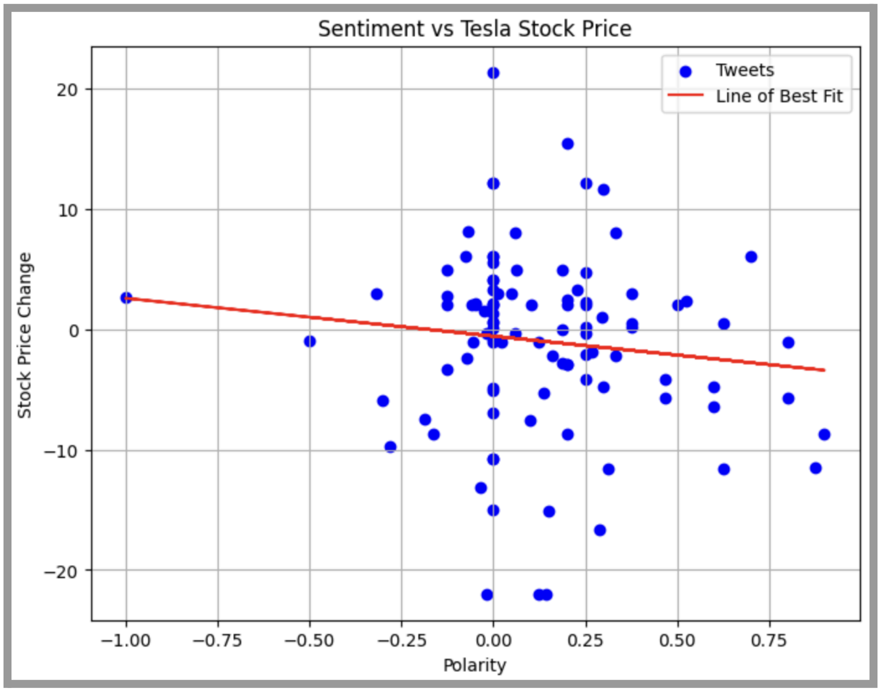

Figure 6

<u>Slope = -3.146</u>

Stock data for SpaceX could not be collected separately, as SpaceX is not a publicly traded company. However, we still felt it was relevant to include as the company is managed and founded by Musk and makes up a good portion of the content of his tweets about his company.

Additionally, it should be noted that our results for each company are naturally potentially inclined to skew negative as Elon Musk is inclined to speak positively to the public about the companies he manages, especially knowing that there could be a potential effect. As such, the weak or negative correlations shown in our results can be explained.

<u>Conclusion</u>

In this project, we looked to examine whether Elon Musk's tweet about his companies—Twitter, SpaceX, and Tesla— had a significant impact on their respective stock prices relative to that day. By analyzing the sentiment, or polarity, of the tweets and comparing it with the stock prices changes using linear regression, we sought to find if there was any consistent correlation or relationship between the world's richest man's statements and his companies market behavior. Specifically, the reason we were drawn to this project was to see if Musk's high social media status, characterized by his frequent and often very influential tweets could directly affect and influence market prices for these companies he tweets about so often.

The results of our project showed that there was a very little to almost no consistent correlation between the sentiment of the tweets and the stock price change of his companies. For Twitter, we concluded a slight positive correlation, but the slope of the line of best fit was not very strong at 0.243 (Figure 5). This suggests that while there may be some effect, it is not great enough to be considered significantly impactful to twitter's stock price. For Tesla, we saw a strong negative correlation of -3.146 between tweet sentiment and stock price (Figure 6). This, again, further suggests that the sentiment of Elon Musk's tweets does not have predictable influence on stock price movements. These weak and even negative correlations highlight the unpredictability of market sentiment which has no correlation to Musk's tweets sentiments. Such mixed and inconsistent results emphasize the complexity of the stock market as a whole and its dynamics. It suggests that other factors besides Musk's tweets play a more important role in the setting and movement of stock sentiment. Thus, while Musk's tweets garner significant attractions from the media and public, their direct effect on stock prices and sentiment appears to be limited when compared to the broader market forces such as investor sentiment, economic and macroeconomics factors.

Limitations in our approach were also prevalent. A large limitation laid in the sentiment analysis tool we used—TextBlob. TextBlob assigns a sentiment score based on pure textual content, without the consideration of contextual factors, such as nuances, sarcasm, or more subtle implications of Musk's tweets. Elon Musk is known for his witty, sometimes cryptic, and provocative language which we noticed in the results that TextBlob would miscategorize. For example, leading a tweet with all the negatives of the current state of twitter followed by a vow to fix it is seen from a human perspective as a positive statement, however, in the eyes of TextBlob that gets overshadowed by the amount of negative sentiment in the tweet. Additionally, our analysis only used tweets from 2022, which limited the scope of analysis and made it difficult to observe long term trends and find correlation. If we had used a larger database covering multiple years, it could have provided a more robust result and correlation. Furthermore, as mentioned before, stock prices are influenced by a number of different factors, which makes it quite difficult to attribute changes in sentiment to a single individual's tweets no matter how influential they may be. These complexities display that even if Musk's tweets do have some influence, it is likely that they are interwoven with other forces.

Future analysis in this matter could be improved by incorporating additional variables that may affect stock price movement, such as volume, market indices, or major news headlines. Including factors such as these would provide a true and more accurate representation of what drives stock prices and help assist in prediction alongside Musk's tweets by essentially "looking at the bigger picture". Moreover, the use of a more advanced sentiment analysis tool or machine learning algorithm could capture more complex linguistics such as sarcasm or contextual embeddings, leading to a more accurate evaluation of a tweet's impact. Additionally, using a database which covers a longer period of time and includes more detailed market data could help uncover stronger correlations of how Musk's tweets create an impact over time.

In the final account, our study did not find a strong or consistent relationship between the sentiment of Elon Musk's tweets and the stock price change of Twitter and Tesla. However, we found valuable insight into the complexities of financial markets and highlighted the insufficient power a single factor holds in a stock's valuation. Our findings suggest that simplistic correlations do not capture the intricacies at play. Furthering our analysis with the use of more sophisticated sentiment analysis, broader datasets, and a deeper emphasis of market factors could help shed light on the way influential figures impact market behavior and uncover this questionable power.

Works Cited

Article Reference:

Hasan, Md, Maliha, Maisha, and M. Arifuzzaman. (2019). Sentiment Analysis with NLP on Twitter Data. 1-4. https://doi.org/10.1109/IC4ME247184.2019.9036670

Software References:

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. Nature, 585, 357-362.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

Loria, S. (2018). TextBlob: Simplified Text Processing. Retrieved from https://textblob.readthedocs.io/en/dev/

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Plotly Technologies Inc. (2015). Collaborative data science. Retrieved from https://plot.ly

Waskom, M. L. (2021). Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021.

Dataset References:

Marta99. (2022). Elon Musk's Tweets Dataset 2022. Retrieved from https://www.kaggle.com/datasets/marta99/elon-musks-tweets-dataset-2022/data

Aspillai. (2022). Tesla Stock Price with Indicators (10 Years). Retrieved from https://www.kaggle.com/datasets/aspillai/tesla-stock-price-with-indicators-10-years/discussion/483377

Maharshipandya. (2022). Twitter Stocks Dataset. Retrieved from
https://www.kaggle.com/datasets/maharshipandya/twitter-stocks-dataset