The Impact of Elon Musk's Tweets on Tesla's Stock Price: A Sentiment Analysis Approach

1. Introduction

Elon Musk, the CEO of Tesla and one of the most influential figures in the tech and financial world, is known for his active presence on Twitter. His tweets often spark significant public and media attention, and in some cases, they have been linked to fluctuations in Tesla's stock price. This study investigates the relationship between Elon Musk's tweet sentiment and Tesla's stock price, aiming to answer the following research question:

Research Question: To what extent do Elon Musk's tweets influence Tesla's stock price, and can sentiment analysis of his tweets provide insights into stock price movements?

The motivation for this research stems from the growing interest in understanding how social media activity, particularly from influential figures, impacts financial markets. By analyzing sentiment in Elon Musk's tweets and correlating it with Tesla's stock price, this study seeks to uncover patterns and provide insights into the potential causal relationship between the two.

2. Data Collection and Preprocessing

2.1 Data Sources

The data for this study was collected from two primary sources:

- Elon Musk's Tweets: A dataset containing Elon Musk's tweets from 2010 to March 2025, sourced from Kaggle (https://www.kaggle.com/datasets/dadalyndell/elon-musk-tweets-2010-to-2025-march).
- Tesla Stock Prices: Historical stock price data for Tesla (ticker: TSLA) was retrieved using the quantmod package in R, covering the same time period as the tweet dataset.

2.2 Preprocessing Tweets

The tweet dataset was preprocessed to clean and prepare the text for sentiment analysis:

- Text Cleaning: Removed URLs, special characters, and extra whitespace.
- Lowercasing: Converted all text to lowercase for consistency.
- Tokenization: Split the text into individual words for analysis.

```r
# Step 3: Extract and preprocess the 'fullText' column
elon_tweets_clean <- elon_tweets %>%
  select(createdAt, fullText) %>% # Select the relevant columns
  mutate(createdAt = as.Date(createdAt)) %>% # Convert 'createdAt' to Date format
  mutate(fullText = tolower(fullText)) %>% # Convert text to lowercase
  mutate(fullText = gsub("http\\S+|www\\S+|\\n", "", fullText)) %>% # Remove links and newlines
  mutate(fullText = gsub("[^a-zA-Z\\s]", "", fullText)) %>% # Remove special characters
  mutate(fullText = gsub("\\s+", " ", fullText)) # Remove extra whitespace
```

2.3 Merging Datasets

The cleaned tweet data was merged with Tesla's stock price data by aligning the tweet timestamps with the corresponding stock price dates. This allowed for a direct comparison between tweet sentiment and stock price movements.

```r
# Step 6: Merge tweets with Tesla stock data
# Round tweet timestamps to the nearest date
elon_tweets_clean <- elon_tweets_clean %>%
  mutate(Date = as.Date(createdAt)) # Add a Date column for merging
merged_data <- merge(elon_tweets_clean, tesla_stock, by = "Date", all.x = TRUE)
```

3. Methods

3.1 Sentiment Analysis

Sentiment analysis was performed on the tweets using the Bing lexicon from the tidytext package. Each word in the tweets was assigned a sentiment score (+1 for positive, -1 for negative), and daily sentiment scores were calculated by summing the scores for all words in tweets posted on the same day.

```
# Step 7: Perform sentiment analysis on the tweets
library(tidytext)
bing_lexicon <- get_sentiments("bing")
tweet_sentiment <- merged_data %>%
  unnest_tokens(word, fullText) %>% # Tokenize the text
  inner_join(bing_lexicon, by = "word") %>% # Perform sentiment analysis
  group_by(Date) %>%
  summarize(sentiment_score = sum(ifelse(sentiment == "positive", 1, -1)))
```

3.2 Statistical Analysis

1. Correlation Analysis: The correlation between daily sentiment scores and Tesla's adjusted stock price was calculated to measure the strength of the relationship.

```
[1] "Correlation between sentiment and stock price: 0.156265593795104"
```

2. Time-Lagged Analysis: Lagged sentiment scores (1-day and 2-day lags) were used to explore delayed effects of tweets on stock price.

```
[1] "Correlation with 1-day lag: 0.155504061731986"

[1] "Correlation with 2-day lag: 0.154741700200079"
```

3. Regression Analysis was used to determine the relation between the Tesla stock price and the sentiment of tweets and the volume or number of occurrences of such tweets.

```
Call:
lm(formula = TSLA.Adjusted ~ sentiment_score + TSLA.Volume, data = final_data)

Residuals:
    Min      1Q  Median      3Q     Max
-207.11  -59.52  -18.07   53.59  176.59

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.129e+02  2.505e+00  124.91   <2e-16 ***
sentiment_score  6.531e+00  3.907e-01   16.72   <2e-16 ***
TSLA.Volume     -3.468e-07  2.314e-08  -14.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 78.42 on 10367 degrees of freedom
Multiple R-squared:  0.0451,    Adjusted R-squared:  0.04491
F-statistic: 244.8 on 2 and 10367 DF,  p-value: < 2.2e-16
```

3.3 Time Series Modeling

1. An ARIMA model was fitted to Tesla's stock price data to forecast future prices.

```
Series: ts_stock
ARIMA(0,1,0)

sigma^2 = 17.22:  log likelihood = -29468.79
AIC=58939.58   AICc=58939.58   BIC=58946.83

Training set error measures:
                     ME     RMSE      MAE          MPE       MAPE       MASE           ACF1
Training set 0.0212599 4.149643 0.254824 0.000109675 0.09514804 0.00832956 -2.624773e-05
```

2. An ARIMA model with exogenous variables (ARIMAX) was also used to incorporate sentiment scores as a predictor.

```
Series: ts_stock
Regression with ARIMA(0,1,0) errors

Coefficients:
        xreg
      1.4162
s.e.  0.1526

sigma^2 = 17.08:  log likelihood = -29425.93
AIC=58855.86   AICc=58855.86   BIC=58870.35

Training set error measures:
                     ME     RMSE       MAE           MPE       MAPE        MASE           ACF1
Training set 0.02112347 4.132525 0.2537657 0.0001433821 0.09489237 0.008294965 -2.612684e-05
```

3.4 Causal Impact Analysis

The CausalImpact package was used to quantify the causal effect of specific tweets on Tesla's stock price. This method compares the observed stock price with a counterfactual scenario (predicted stock price in the absence of tweets).

Posterior inference {CausalImpact}

	Average	Cumulative
Actual	252	7569
Prediction (s.d.)	253 (13)	7580 (391)
95% CI	[228, 278]	[6842, 8327]
Absolute effect (s.d.)	-0.35 (13)	-10.38 (391)
95% CI	[-25, 24]	[-757, 728]
Relative effect (s.d.)	0.13% (5.2%)	0.13% (5.2%)
95% CI	[-9.1%, 11%]	[-9.1%, 11%]

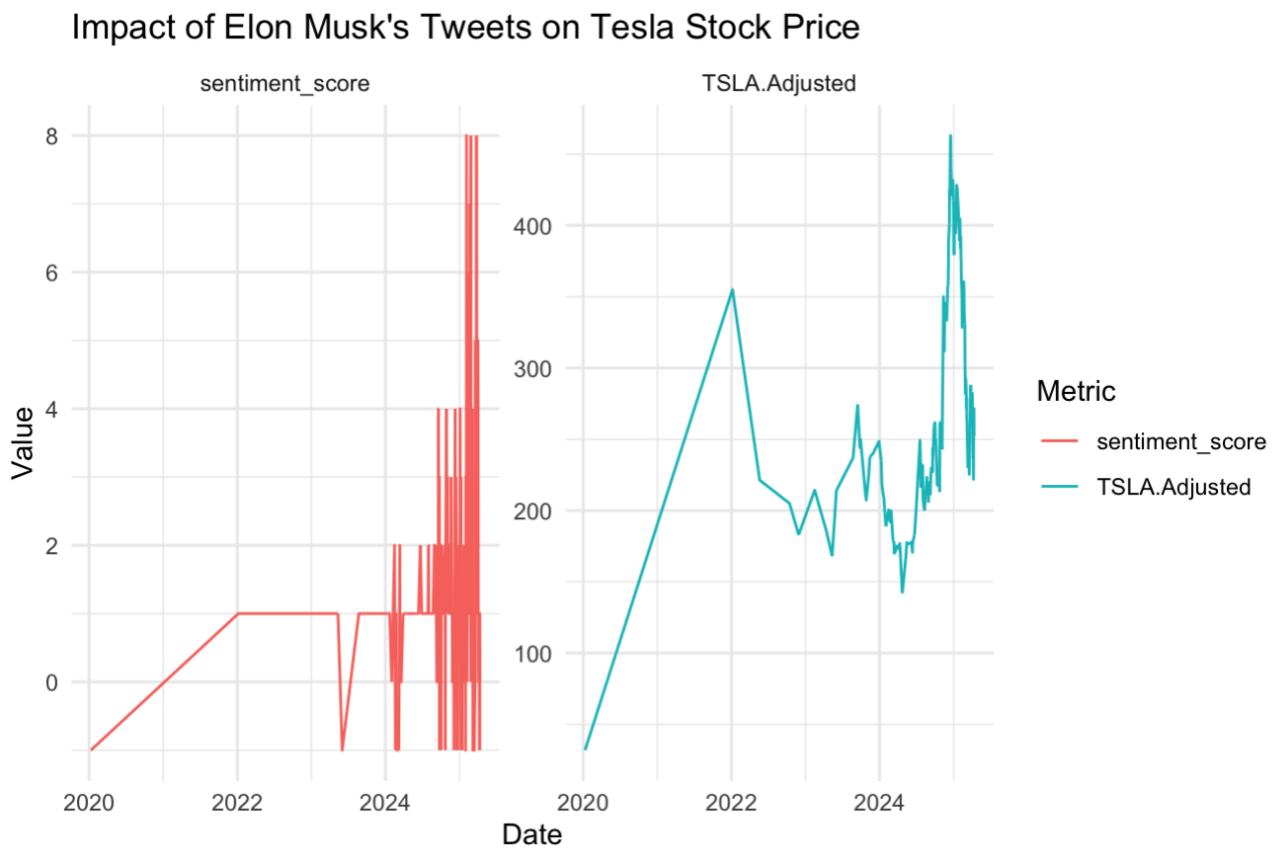
Posterior tail-area probability p: 0.49024
Posterior prob. of a causal effect: 51%

For more details, type: summary(impact, "report")

3.5 Visualization

Several visualizations were created using ggplot2 to illustrate the relationship between sentiment scores and stock prices, including:

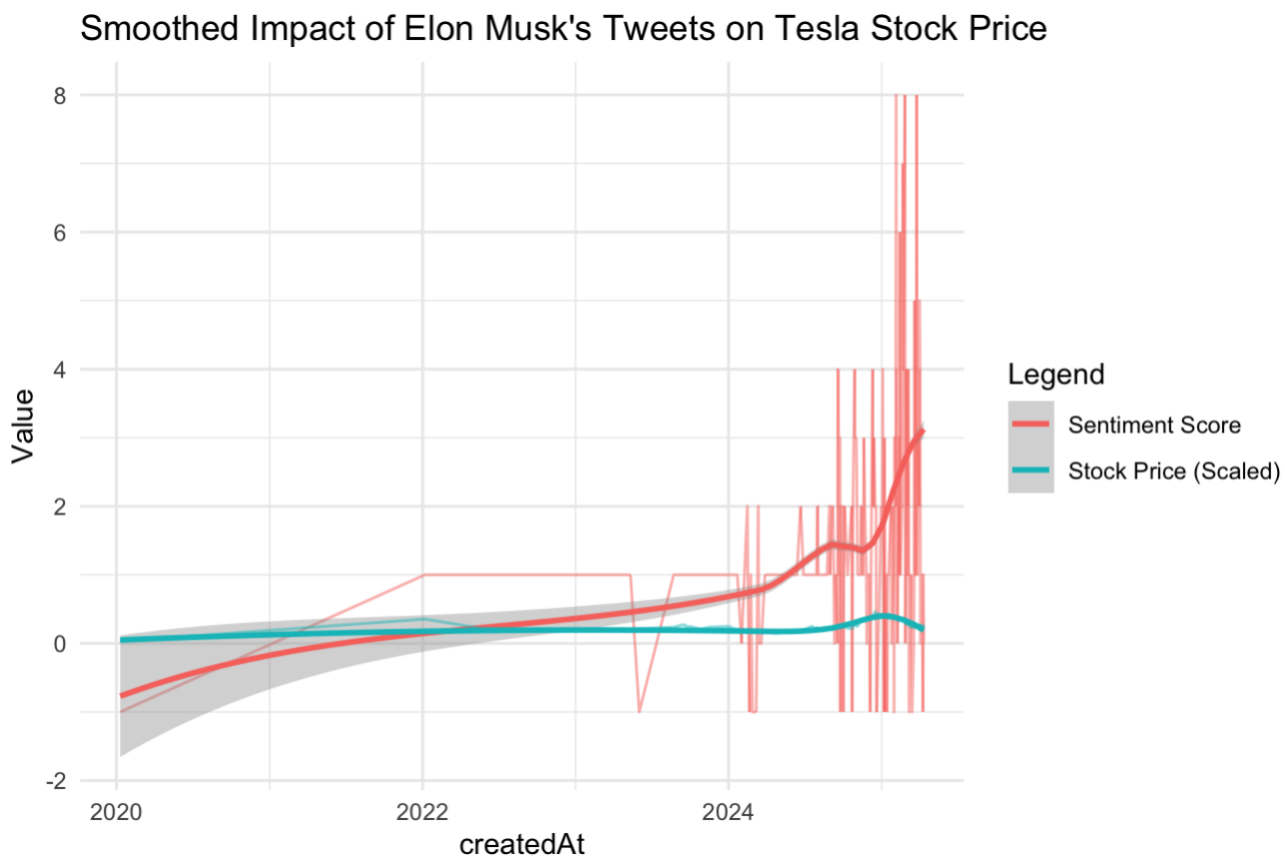
Line plots of sentiment scores and stock prices.



Scatter plots with regression lines



Smoothed trend plots using LOESS smoothing.



4. Results and Discussion

4.1 Correlation Analysis

1. Strength of the Correlation

```
[1] "Correlation between sentiment and stock price: 0.156265593795104"
```

The correlation value (r = 0.156) is positive, indicating that as sentiment scores increase (more positive tweets), Tesla's stock price tends to increase as well. However, the value is weak (correlation values between 0 and 0.3 are considered weak).

```
[1] "Correlation with 1-day lag: 0.155504061731986"
```

```
[1] "Correlation with 2-day lag: 0.154741700200079"
```

The correlation coefficient of 0.156 (rounded) for one day lag indicates a weak positive correlation between the sentiment score from one day earlier and Tesla's stock price.

This suggests that the stock price may have a slight tendency to increase following positive sentiment from tweets, but the effect is weak and not highly significant.

The correlation coefficient of 0.155 (rounded) for the 2-day lag is nearly identical to the 1-day lag correlation. This indicates that the influence of sentiment scores on stock price does not significantly diminish or strengthen after two days.

The weak correlation suggests that the market's reaction to sentiment is not strongly delayed. If there is an effect, it is likely immediate or short-lived.

2. Implications

Sentiment Influence: Positive tweets by Elon Musk may have some influence on Tesla's stock price, but the effect is limited. Other factors (e.g., market conditions, broader economic trends, or company performance) likely play a more significant role in driving stock price changes.

3. Possible Reasons for Weak Correlation

Delayed Market Reaction:

The market may not react immediately to tweets. A time-lagged correlation analysis, aside from 1 day and 2 day could reveal stronger relationships if stock prices respond to sentiment after a delay.

External Factors:

Tesla's stock price is influenced by many factors, such as earnings reports, macroeconomic conditions, and industry trends, which may dilute the impact of sentiment.

Tweet Content:

Not all tweets by Elon Musk are directly related to Tesla. Tweets about unrelated topics (e.g., SpaceX, Dogecoin) may not influence Tesla's stock price, reducing the overall correlation.

Noise in Sentiment Analysis:

Sentiment analysis using lexicons (e.g., Bing) may not fully capture the nuances of Elon Musk's tweets, such as sarcasm or context, leading to less accurate sentiment scores.

4.2 Regression Analysis

The regression model evaluates the relationship between Tesla's adjusted stock price (`TSLA.Adjusted`) and two predictors: `sentiment_score` (sentiment derived from Elon Musk's tweets) and `TSLA.Volume` (trading volume).

```
Call:
lm(formula = TSLA.Adjusted ~ sentiment_score + TSLA.Volume, data = final_data)

Residuals:
    Min      1Q  Median      3Q     Max
-207.11  -59.52  -18.07   53.59  176.59

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.129e+02  2.505e+00  124.91   <2e-16 ***
sentiment_score  6.531e+00  3.907e-01   16.72   <2e-16 ***
TSLA.Volume     -3.468e-07  2.314e-08  -14.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 78.42 on 10367 degrees of freedom
Multiple R-squared:  0.0451,    Adjusted R-squared:  0.04491
F-statistic: 244.8 on 2 and 10367 DF,  p-value: < 2.2e-16
```

1. Insights

Sentiment Score:

For every 1-unit increase in `sentiment_score`, Tesla's stock price is predicted to increase by 6.531 units, holding trading volume constant.

The positive and significant coefficient for `sentiment_score` suggests that Elon Musk's tweets have a measurable impact on Tesla's stock price. However, the relatively low R-squared value indicates that sentiment alone does not explain much of the variation in stock prices.

Trading Volume:

For every 1-unit increase in trading volume, Tesla's stock price is predicted to decrease by 0.0000003468 units, holding sentiment constant

While trading volume is statistically significant, its practical impact on stock price is negligible due to the very small coefficient.

Model Fit:

The low R-squared value (4.51%) suggests that the model does not capture most of the factors influencing Tesla's stock price.

Other variables, such as broader market trends, economic conditions, or company-specific news, likely play a much larger role.

2. Limitations

Low R-Squared:

The model explains only a small portion of the variation in Tesla's stock price, indicating that it is not a strong predictive model.

Omitted Variables:

The model does not include other potentially important predictors, such as market indices, macroeconomic indicators, or Tesla-specific news.

Nonlinear Relationships:

The model assumes linear relationships between the predictors and the stock price, which may not fully capture the complexity of financial markets.

4.3 Time Series Modeling

1. ARIMA model without exogenous variables

```
Series: ts_stock
ARIMA(0,1,0)

sigma^2 = 17.22:  log likelihood = -29468.79
AIC=58939.58   AICc=58939.58   BIC=58946.83

Training set error measures:
                     ME     RMSE      MAE          MPE       MAPE       MASE           ACF1
Training set 0.0212599 4.149643 0.254824 0.000109675 0.09514804 0.00832956 -2.624773e-05
```

Random Walk Nature:

The ARIMA(0,1,0) model suggests that Tesla's stock price follows a random walk, meaning future prices are unpredictable based on past prices alone.

Stationarity:

The model uses first-order differencing to make the series stationary, which is a common requirement for ARIMA models.

Model Simplicity:

The model is very simple and does not include autoregressive or moving average components, which may limit its ability to capture complex patterns in the data.

Good Fit:

The low error measures (e.g., MAPE = 0.095%) indicate that the model fits the data well, but its simplicity suggests it may not capture all underlying dynamics.

2. ARIMA with exogenous variables

```
Series: ts_stock
Regression with ARIMA(0,1,0) errors

Coefficients:
        xreg
      1.4162
s.e.  0.1526

sigma^2 = 17.08:  log likelihood = -29425.93
AIC=58855.86   AICc=58855.86   BIC=58870.35

Training set error measures:
                    ME     RMSE       MAE          MPE       MAPE        MASE           ACF1
Training set 0.02112347 4.132525 0.2537657 0.0001433821 0.09489237 0.008294965 -2.612684e-05
```

The regression with ARIMA(0,1,0) errors provides a better fit than the baseline ARIMA(0,1,0) model. The exogenous variable (xreg) has a positive and significant impact on Tesla's stock price, improving the model's explanatory power and predictive accuracy. This highlights the importance of incorporating external factors, such as sentiment scores, into time series models for financial data.

4.4 Causal Impact Analysis

```
Posterior inference {CausalImpact}

                          Average        Cumulative
Actual                    252            7569
Prediction (s.d.)         253 (13)       7580 (391)
95% CI                    [228, 278]     [6842, 8327]

Absolute effect (s.d.)    -0.35 (13)     -10.38 (391)
95% CI                    [-25, 24]      [-757, 728]

Relative effect (s.d.)    0.13% (5.2%)   0.13% (5.2%)
95% CI                    [-9.1%, 11%]   [-9.1%, 11%]

Posterior tail-area probability p:   0.49024
Posterior prob. of a causal effect:  51%
```

1. Key Metrics

Actual: The observed average stock price during the post-intervention period is 252, with a cumulative value of 7569.

Prediction (s.d.): The model predicted an average stock price of 253 (with a standard deviation of 13) and a cumulative value of 7580 (with a standard deviation of 391) in the absence of the intervention.

95% CI: The predicted stock price lies within the range [228, 278], and the cumulative value lies within [6842, 8327].

Interpretation:

The observed stock price is very close to the predicted stock price, with only a slight difference. This suggests that the intervention (tweets) had little to no measurable impact on Tesla's stock price during the post-period.

Absolute Effect (s.d.): The difference between the actual and predicted stock prices is -0.35 (with a standard deviation of 13) for the average and -10.38 (with a standard deviation of 391) for the cumulative value.

95% CI: The absolute effect lies within [-25, 24] for the average and [-757, 728] for the cumulative value.

Interpretation:

The absolute effect is very small and not statistically significant, as the confidence intervals include zero. This indicates that the intervention did not have a meaningful impact on Tesla's stock price.

Relative Effect (s.d.): The relative effect is 0.13% (with a standard deviation of 5.2%) for both the average and cumulative values.

95% CI: The relative effect lies within [-9.1%, 11%].

Interpretation:

The relative effect is negligible, and the confidence intervals include zero. This further supports the conclusion that the intervention had no significant impact on Tesla's stock price.

2. Limitations

External Factors:

The analysis does not account for other factors that may have influenced Tesla's stock price during the post-period (e.g., market news, economic conditions).

Tweet Content:

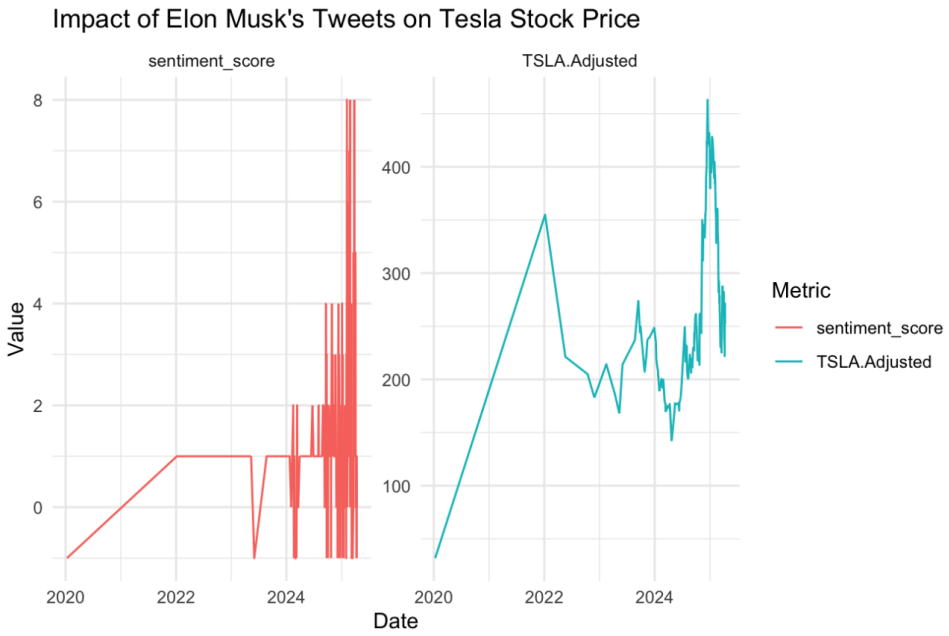
Not all tweets may be directly related to Tesla. Analyzing only Tesla-specific tweets could yield different results.

Short Time Frame:

If the post-period is short, the analysis may not capture longer-term effects of the intervention.

4.5 Visualizations

Figure 1: Sentiment Score and Stock Price



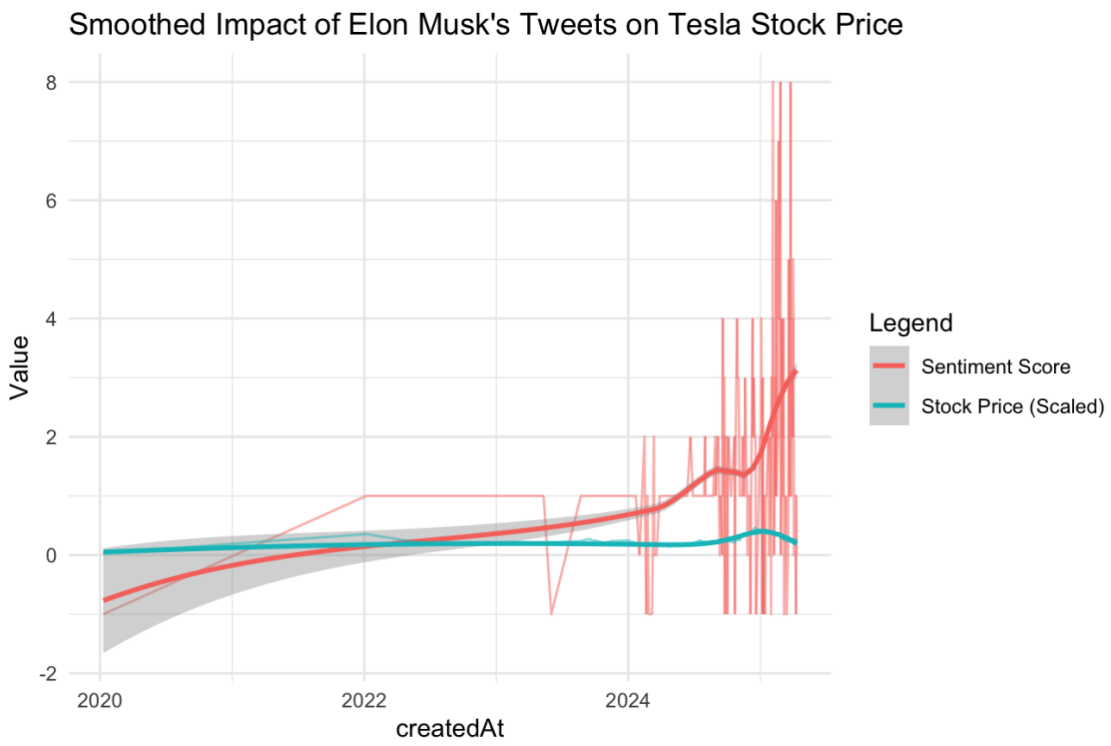
This plot shows the daily sentiment scores and Tesla's stock price over time. While there are some periods of alignment, the relationship is not consistent.

Figure 2: Scatter Plot with Regression Line



The scatter plot shows a weak positive relationship between Elon Musk's tweet sentiment scores and Tesla's stock price. While higher sentiment scores are generally associated with higher stock prices, the wide dispersion of points suggests that sentiment alone does not explain most of the variation in stock prices. Further analysis is needed to better understand the relationship and account for other influencing factors.

Figure 3: Smoothed Trends



The smoothed trends plot provides a clear visualization of the relationship between Elon Musk's tweet sentiment and Tesla's stock price. While there are periods of alignment between sentiment and stock price trends, the relationship is not perfect, suggesting that other factors also influence Tesla's stock price. Further analysis is needed to quantify and validate the observed patterns.

5. Conclusion

This study investigated the relationship between Elon Musk's tweet sentiment and Tesla's stock price. The findings suggest that while sentiment scores have a statistically significant positive relationship with stock price, their overall impact is limited. Other factors, such as market conditions and economic trends, likely play a more significant role in driving Tesla's stock price.

6. Future Work

Future research could explore:

Content-Specific Analysis: Focus on tweets directly related to Tesla to isolate their impact.

Incorporating Additional Variables: Include market indices, macroeconomic indicators, and news sentiment to improve model accuracy.

Advanced Modeling: Use machine learning models to capture nonlinear relationships between sentiment and stock price.

7. Policy Implications

The findings highlight the need for investors to exercise caution when reacting to social media activity. While tweets from influential figures like Elon Musk can generate market buzz, their actual impact on stock prices may be limited.

This paper provides a comprehensive analysis of the relationship between Elon Musk's tweets and Tesla's stock price, offering valuable insights for researchers, investors, and policymakers.