

# Social Media Sentiment and Tesla Stock Price

MADS: mbeards1, amukhe12, fozturk1, sbelin

## Hypothesis

There is a significant relationship between the sentiment of social media mentions of Tesla stock on Twitter and Reddit and Tesla stock price on a given day.

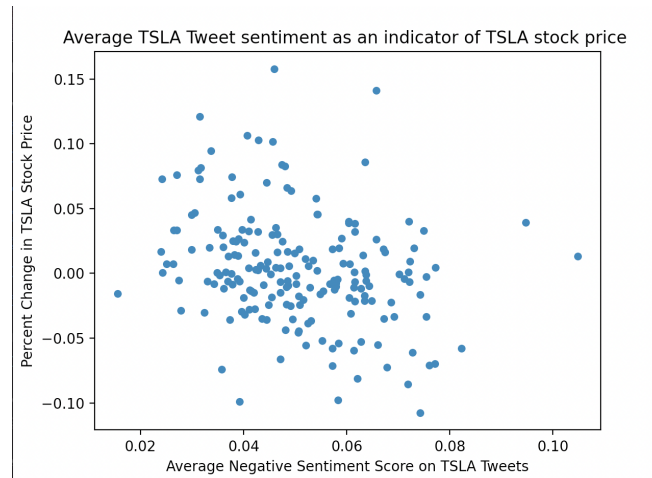
## Data

A few key datasets are used as the basis for the analysis. Tweets mentioning Tesla are scraped from Twitter using Selenium and cleaned, and Reddit posts mentioning Tesla are accessed via the PushShift API. Tesla stock data was downloaded from <https://www.nasdaq.com/market-activity/stocks/tsla/historical> and Russell stock data was downloaded from <https://finance.yahoo.com/quote/%5ERUA/history/>. We found that the social media data was noisy and difficult to accurately assign sentiment using natural language processing APIs.

## Findings

**Claim #1:** The sentiment of social media mentions of Tesla stock on Twitter can be used as an indicator of Tesla's same-day stock performance.

**Support for Claim #1:** The p-value for the multiple regression test on Twitter data was  $1.29 \times 10^{-5}$ , and the p-value for the linear regression test on Twitter data was 0.008. The data used in the multiple regression test is shown in the below scatterplot, where a downward trend between negative sentiment and percentage change in stock price is apparent.



**Claim #2:** Results vary widely based on the natural language processing API used, the social media platform analyzed, and the statistical test performed, as measured by the statistics produced by the regression models.

**Support for Claim #2:** Shown in the below table are the three statistical methods used and the resulting statistics. As shown in the table, the p-values and r-squared values vary significantly.

Method 1: Reddit Linear Regression	Method 2: Twitter Linear Regression	Method 3: Twitter Multiple Regression
P-value: 0.12 R-squared: 0.07(without filter)	P-value: 0.008 R-squared: 0.04	P-value: $1.29 \times 10^{-5}$ R-squared: 0.215

**Claim #3:** There is no evidence to suggest that there is a significant relationship between the magnitude of activity on TSLA-related tweets and the share price of TSLA for a given day.

**Support for Claim #3:** If we accept the likes and comments on a Tweet to represent the magnitude of activity surrounding a Tweet, we can explore the relationship between the average number of likes and comments on Tweets of a given day and the change in share price on that day. Below are the Pearson correlation constants between two significant independent variables and percent change in share price vs those of “likes” and “comments”. “Likes” represents the average number of likes on Tweets of a given day and “comments” represents the average number of comments on Tweets of a given day. Both have correlation constants close to 0, and therefore do not have a significant relationship with a change in share price.

Independent Variable	Pearson's r with respect to % change in TSLA share price
Average Positive Sentiment Score across TSLA Tweets	0.190
Average Negative Sentiment Score across TSLA Tweets	-0.270
“likes”	0.055
“comments”	0.061