Social Media Mentions and Stock Performance

Stock Scrapers: cbetancu, jbyers3, ndepiero, hadrian

Hypothesis

Social media has largely proved itself to be an impactful component in users' decision-making. Whether it influences the clothes they wear or the music they listen to, it has also proven to shape the way certain individuals invest their money. We wanted to observe the relationship between the amount of social media mentions a stock receives on any given day and the following performance of that stock. While we devised three hypotheses, we generally tested the following statement: increased social media mentions results in increased stock price volatility for respective mentioned companies.

Data

We collected data about mentions of stocks from Twitter and Reddit online datasets along with the respective prices of the referenced stocks from the Yahoo Finance API. Our Tweets were from 4/9/2020 to 7/16/2020 and Reddit posts from 1/28/2021 and 8/16/2021. Using the Yahoo Finance API, we collected the stock price and volatility for each stock during the respective time periods. We created four data tables: two tables listing the number of mentions for each stock ticker on each day (one for Twitter and one for Reddit), and two tables listing the price volatility for each stock ticker on each day (one for Twitter and one for Reddit). We had to clean the data by removing duplicates and reformatting stock ticker names to match across datasets. Overall, the data was mostly skewed as a result of larger tickers, or stock companies, that received many more mentions on both social media platforms.

Findings

Claim #1: There is no sufficient evidence suggesting an existing relationship between social media mentions and price volatility.

Support for Claim #1:

We ran a two sample t-test that observed price volatility the day after high social media mentions compared to a day after low social media mentions which did not return statistically significant results. Thus, failing to support a statistical difference between price volatility after high versus low social media mentions. Moreover, our linear regression graphed an almost horizontal line, further suggesting social media mentions have little to no effect on stock price volatility. The test return values are displayed below.

Test	**Two Sample T-Test (Price Volatility After High Mentions vs Low Mentions)**
T-Statistic	1.371

P-Value	0.171

Claim #2: There is no evidence suggesting a difference between the impact of Twitter social media mentions on price volatility and that of Reddit social media mentions.
Support for Claim #2: We ran a Two Sample T-Test comparing the price volatility of a stock one day after high social media mentions on Twitter versus one day after high social media mentions on Reddit. Our results were not significant, thus failing to support the claim that Twitter social media mentions impact stock price volatility differently than Reddit social media mentions. The test return values are displayed below.

Test	**Two Sample T-Test (Price Volatility After High Mentions on Twitter vs High Mentions on Reddit)**
T-Statistic	0.448
P-Value	0.657

Claim #3: There is sufficient evidence suggesting an association between a company's market capitalization and daily trading volume.
Support for Claim #3:
We ran a k-means algorithm in hopes of finding relationships within our data set. Observing our k-means distribution graphically, we noticed it resembled a grouping by stock market capitalization (value of a company). Clusters from both graphs were distributed along the volume axis suggesting a relationship between market cap and daily trading volume. Our graphical data is displayed below.