SI 330 Final Presentation Notebook

Pfizer Tweet Sentiment Analysis Compared to Pfizer Stock Performance

Simona Hendl

Problem Statement

The COVID-19 pandemic surged to new heights during fall of 2020. Pfizer-BioNTech (among other companies) created a vaccine to help combat the virus. Due to this development, Pfizer's stock performance has fluctuated, and one factor, among many, that has some influence in stock performance (fluctuation) is company image often reflected by public opinion.

The goal for this project is to figure out how strong (if any) is the correlation between public sentiment (using Pfizer-BioNTech tweets from Twitter) and Pfizer stock performance.

To answer this question three datasets will be used: Pfizer-BioNTech Twitter Data, Pfizer (PFE) Stock Data, and SPY Stock Data. Each dataset was found via Kaggle and downloaded as a CSV.

Pfizer-BioNTech Twitter Data: 4543533 bytes. (11020 rows and 16 columns)

Pfizer (PFE) Stock Data: 837713 bytes. (12474 rows and 7 columns)

SPY Stock Data: 61670 bytes. (1258 rows and 6 columns)

Data Cleaning and Manipulation

To address my central question, the following data cleaning steps were necessary:

1) Subsetting the data to have the same time frame for each dataset.

2) Converting all emojis to text in the Twitter data.

3) Removing all mentions, urls, and special characters from Twitter data.

4) Converting the dates to datetime objects for the final merges and visualizations.

Manipulation

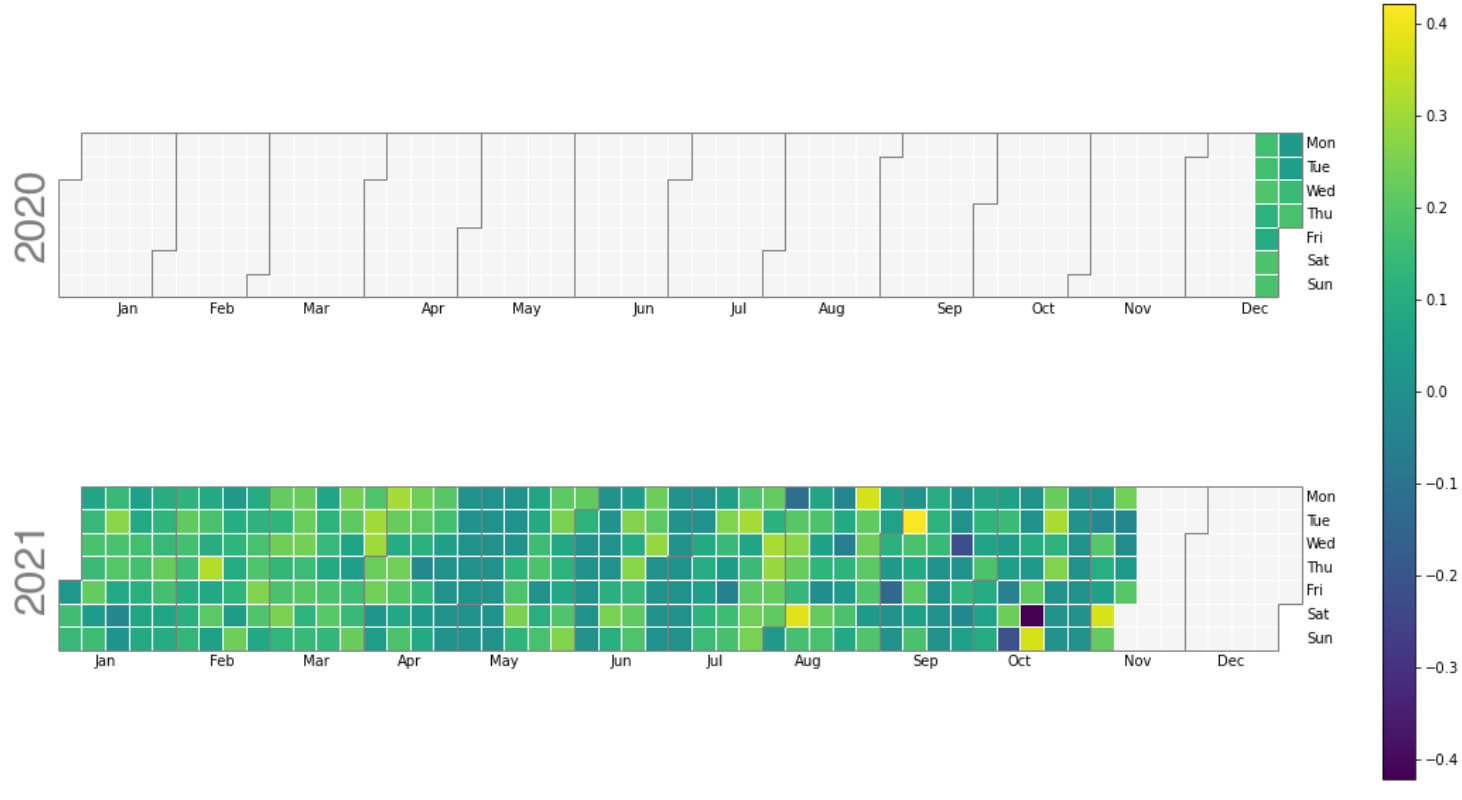
1) Use the Vader Sentiment Analyzer to calculate each tweet's polarity score, and then calcualte the average compound polarity score for each day

2) Calculate the daily percent change ((Close - Open)/Open) for the PFE and SPY data.

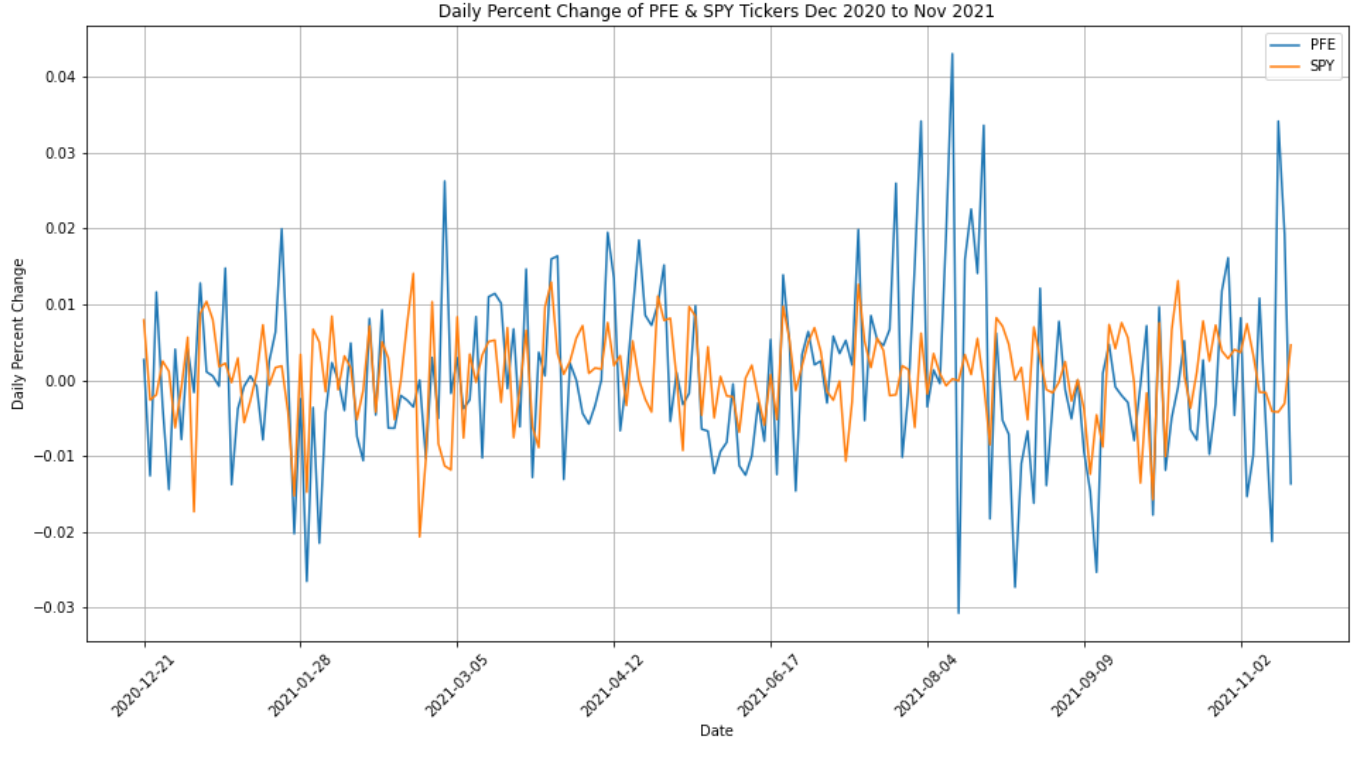
3) Merge the stock and tweet data together.

Analysis

The graph below shows the daily average compound polarity data for the Pfizer-BioNTech Tweets



The graph below shows the comparison between the Pfizer Stock data and the SPY Index Data, comparing the Pfizer stock performance in comparison to the overall stock market.



Slide Type

```
#Correlation coefficient and p-value between daily average compound score and daily percent change for PFE ticker
pearsonr(comp_score_stocks_merged["compound_score"],comp_score_stocks_merged["daily_percent_change_PFE"])
```

```
(-0.11711888385544025, 0.11335193268292013)
```

Ultimately, there was only a slight negative correlation between the daily average compound polarity score (found from the Pfizer BioNTech tweets) and Pfizer Stock Data. The correlation coefficient was -0.07986, and the p-value was 0.2812, so while there is a slight negative correlation, the p-value is above the common threshold of statistical signifiance (5%); therefore, the correlation coefficient is not statistically signficant.

Slide Type

```
#Correlation coefficient and p-value between daily average compound score and daily percent change for SPY ticker
pearsonr(comp_score_stocks_merged["daily_percent_change_PFE"],comp_score_stocks_merged["daily_percent_change_spy"])
```

```
(0.23273821034357922, 0.0014764452687194805)
```

To check that any change in Pfizer stock behavior was not in fact as a result of overall market performance rather than sentiment, the correlation between daily percent change for PFE and daily percent change for SPY (used to track the S&P 500's index) was also found.

The correlation coefficient of 0.2812 is slightly larger meaning that there is stronger correlation between Pfizer's stock performance and the overall market rather than sentiment; however, the correlation coefficient is still small, meaning that the relationship is still relatively weak.

The p-value for this correlation coefficent is 0.001476, which is lower than the necessary threshold hold for statisical signifance. Thus, while the correlation itself is relatively weak between PFE and SPY stock/index performance, the value is statistically significant.

Takeaways

With that said, public sentiment, in this scenario, was less correlated with the Pfizer Stock data relative to the overall stock market performance. While there is some significance, there are other factors that can be tested that may have more influence on Pfizer's Stock data than public sentiment via Twitter.

References

Calplot video: https://www.youtube.com/watch?v=cKMEL9xgq2I