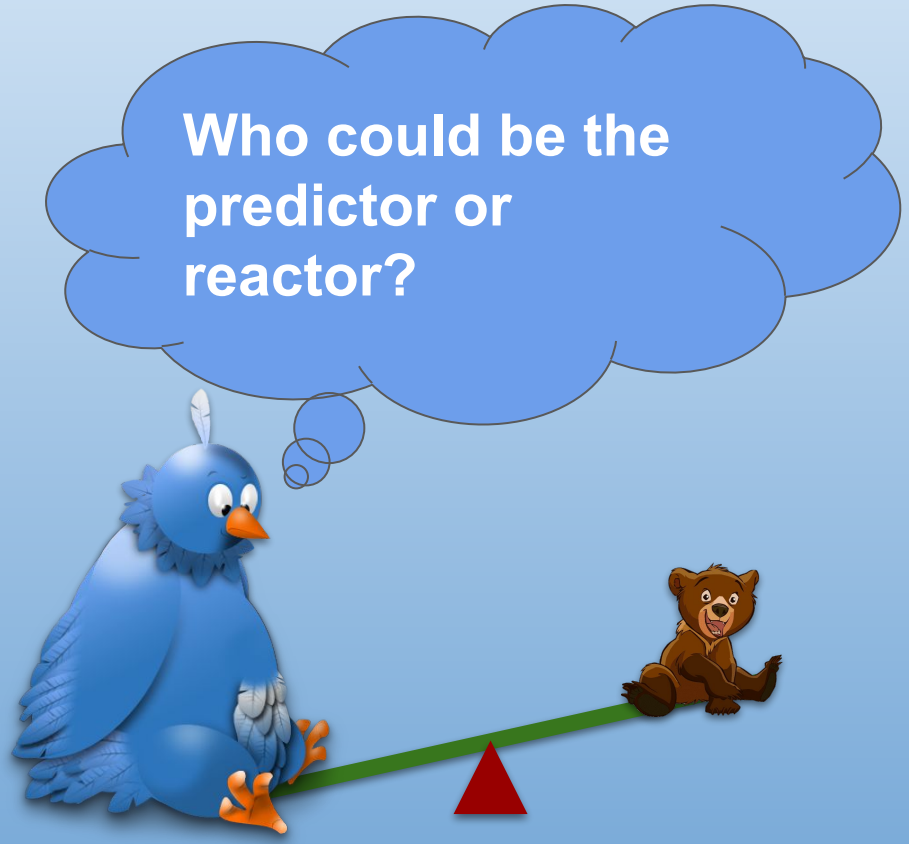
MARKET ANALYSIS

Tommy Steed, Joshua Bui,
Irakli Intskirveli, Ethan Litke

SUSPICION

There is a relationship between Twitter sentiment and stock market performance



Data Collection & Processing

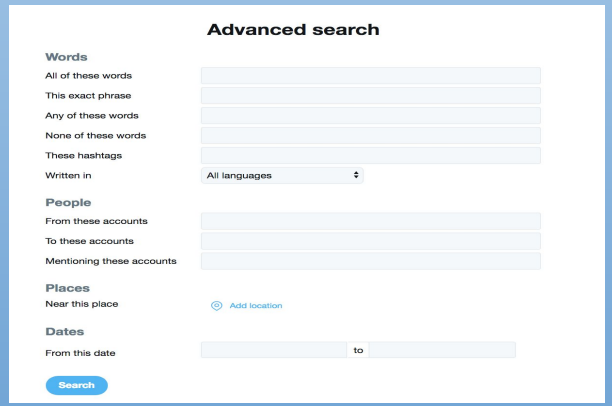
- Twitter Scraper
  - Target Users and Terms
  - CLI output to json file
  - Twitter API has limits
    - Estimated: 400K Tweets
- Pandas Datareader
  - Yahoo Stock Data

For example, from the URL `https://twitter.com/search?l=&q=Trump%20near%3A%22Seattle%2C%20WA%22%20within%3A15mi%20since%3A2017-05-02%20until%3A2017-05-05&src=typd&lang=en`

you need to copy the following part: `Trump%20near%3A%22Seattle%2C%20WA%22%20within%3A15mi%20since%3A2017-05-02%20until%3A2017-05-05`

You can use the CLI with the advanced query, the same way as a simple query:

- based on a daterange: `twitterscraper Trump%20since%3A2017-01-03%20until%3A2017-01-04 -o tweets.json`
- based on a daterange and location: `twitterscraper Trump%20near%3A"Seattle%2C%20WA"%20within%3A15mi%20since%3A2017-05-02%20until%3A2017-05-05 -o tweets.json`
- based on a specific author: `twitterscraper Trump%20from%3AAlWest13 -o tweets.json`




```
#tickers = ['^IXIC', '^DJI', '^GSPC', '^RUT', 'BTC-USD','^N225','^GDAXI', '^XAX']
tickers = ['^IXIC', '^DJI', '^GSPC', '^RUT', '^N225','^GDAXI', '^XAX']

# Define which online source one should use
data_source = 'yahoo'

# We would like all available data from 01/01/2000 until 12/31/2016.
start_date = '2017-11-14'
end_date = '2017-12-15'

# User pandas_reader.data.DataReader to load the desired data. As simple as that.
panel_data = data.DataReader(tickers, data_source, start_date, end_date)
```

Data Analysis

- Python Modules/Libraries
  - Matplotlib
  - NumPy
  - Pandas
  - Seaborn
  - Statsmodels.api
  - vaderSentiment
- Excel
  - Create some graphs
  - Used for pre-work

for Tweets & Stocks:

Target Terms:

- Stock Market
- Dow Jones
- Nasdaq
- NYSE

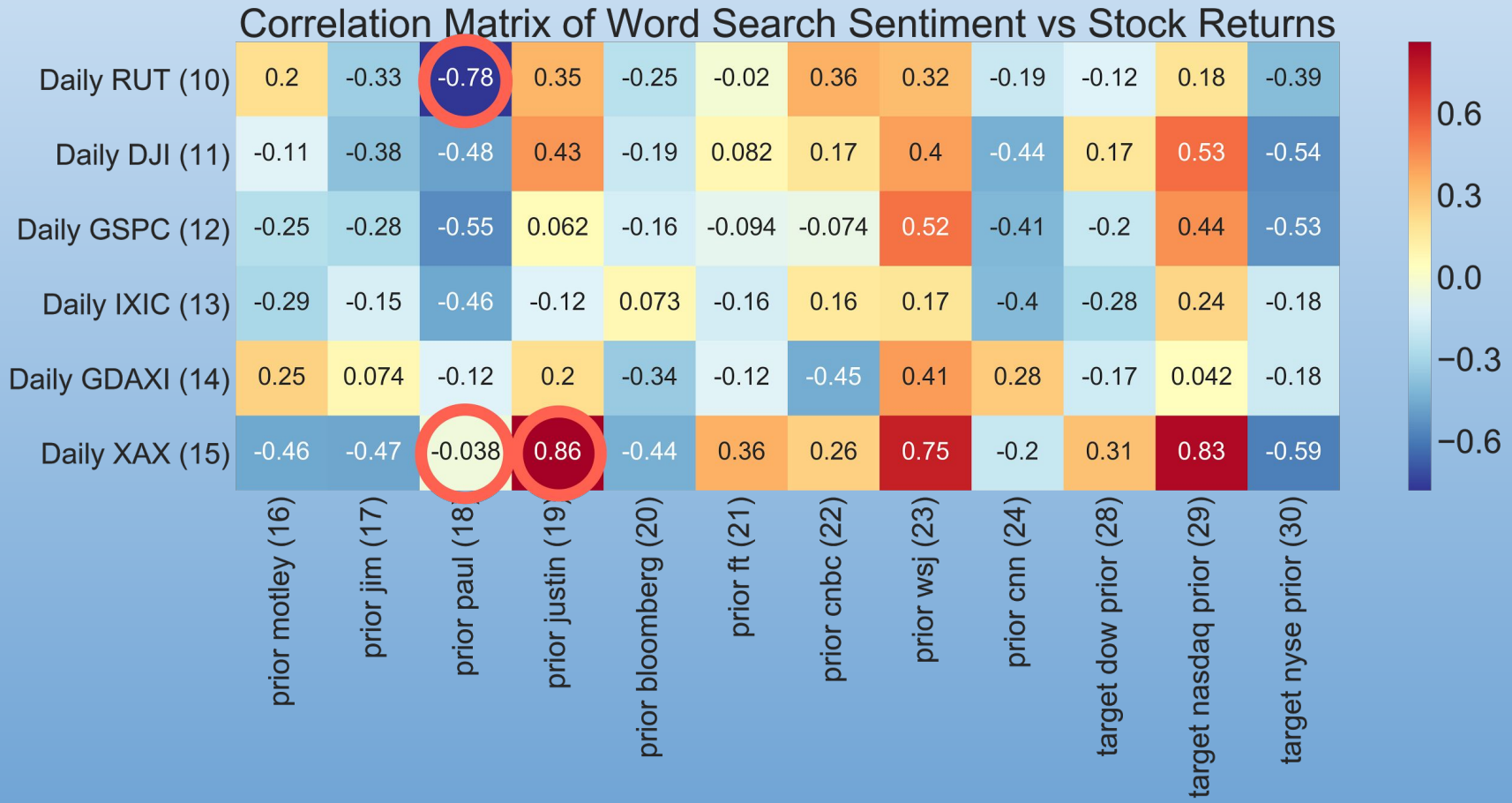
Stocks/Index:

- Russell 2000 (RUT)
- NYSE (XAX)
- DOW (DJI)
- S&P500 (GSPC)
- NASDAQ (IXIC)
- DAX (GDAXI)

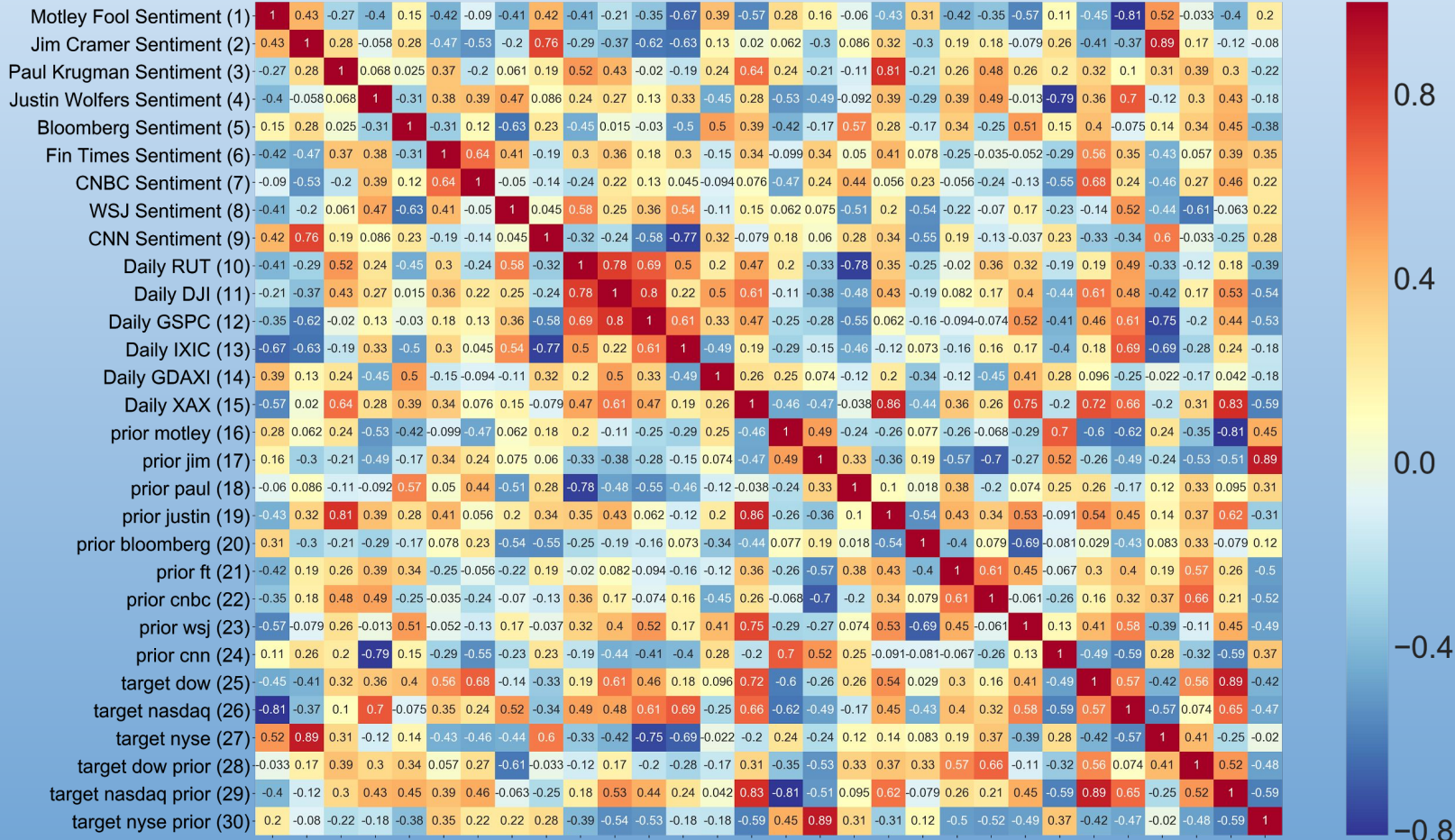
Target Users:

- wsjmarkets
- cnbc
- themotleyfool
- cnnmoney
- markets
- financialtimes
- nytimesbusiness
- paulkrugman
- JustinWolfers
- jimcramer

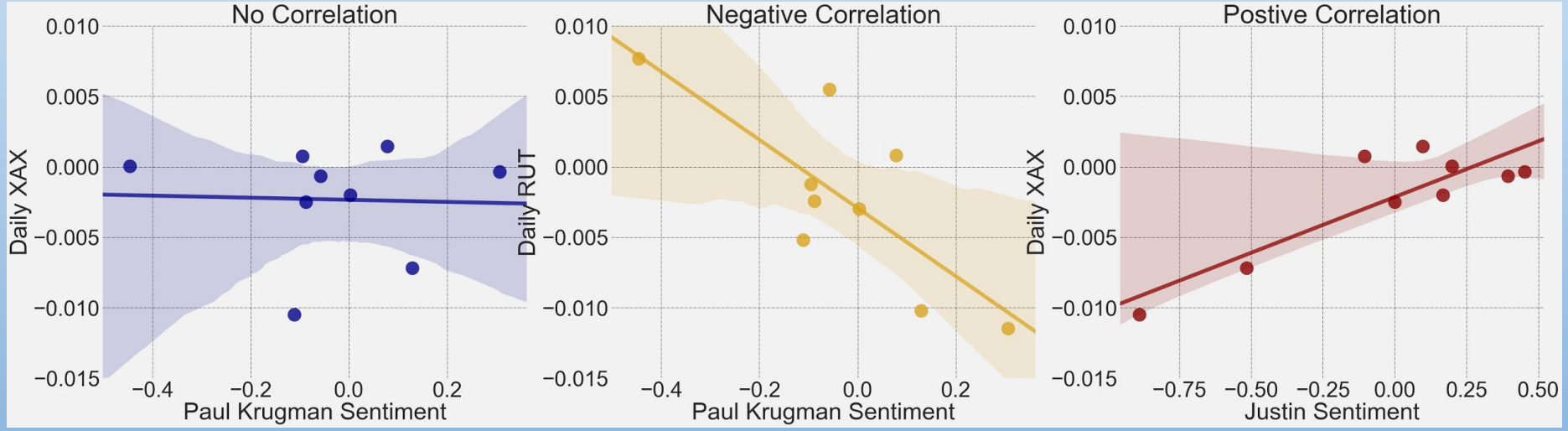
Correlation of market performance and prior day's twitter sentiments:



Correlation Matrix Word Search Sentiment vs Stock Returns



CORRELATIONS



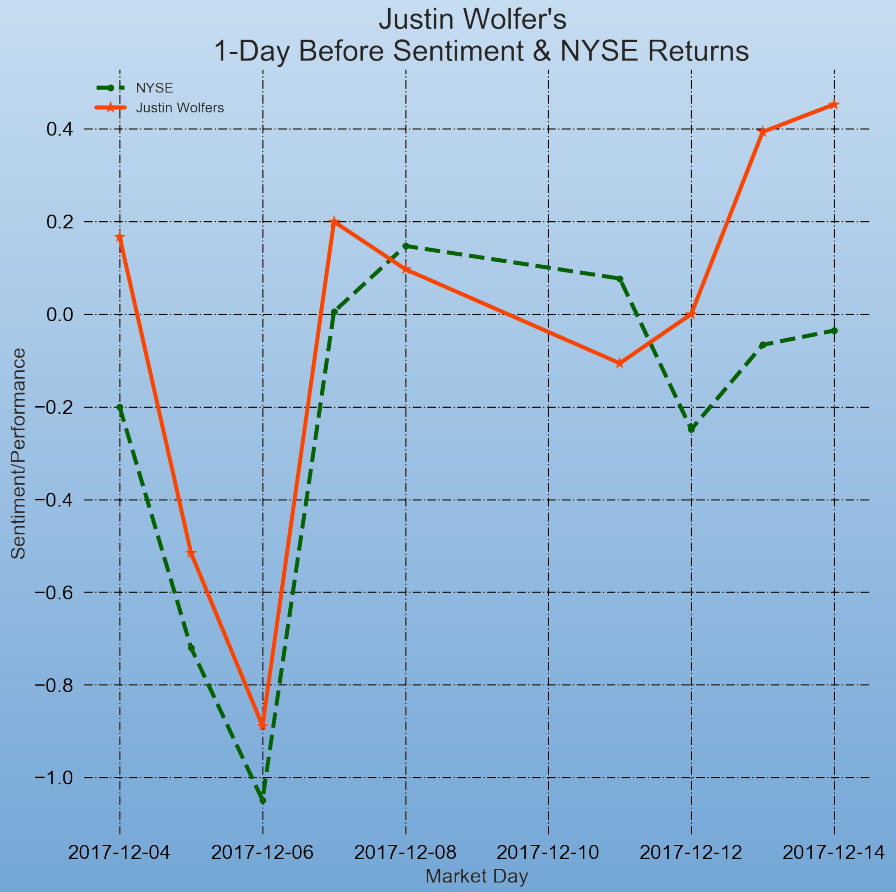
Correlation: -0.01

Correlation: -0.78

Correlation: 0.92

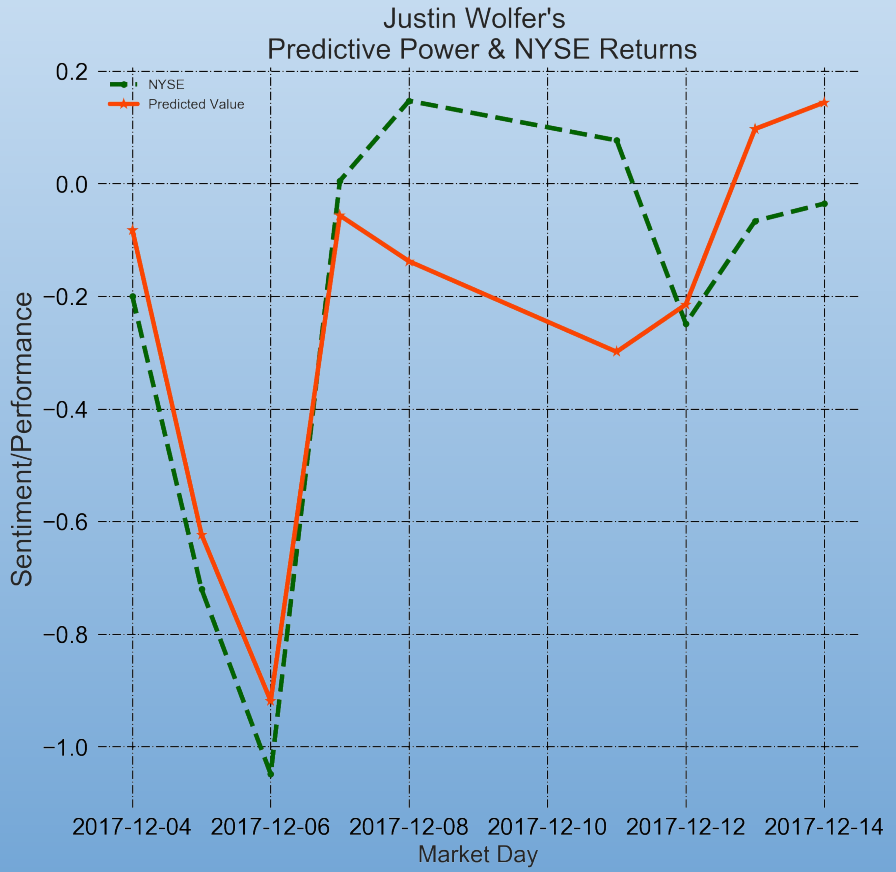
All within a 95% confidence level

HISTORICAL PERFORMANCE



- Justin Wolfer's Sentiment is from the prior day.
- Movements are in the same direction -- in line with the 92% correlation.

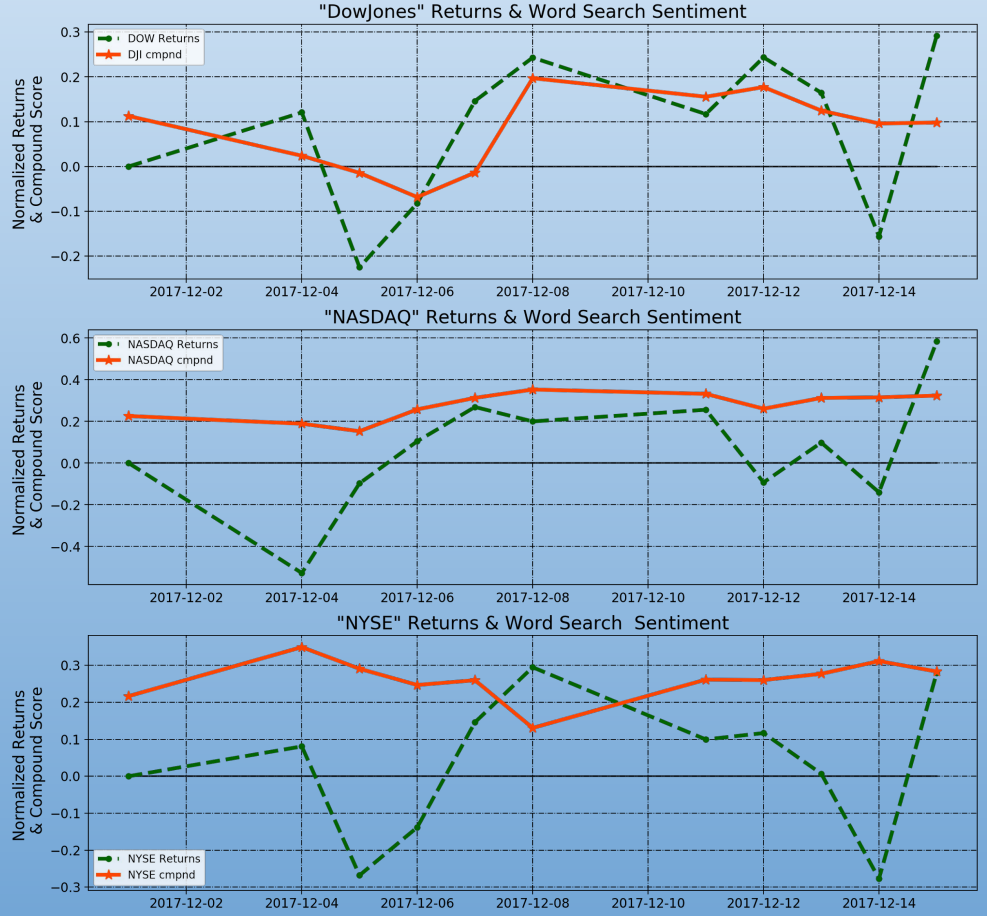
PREDICTED VALUES

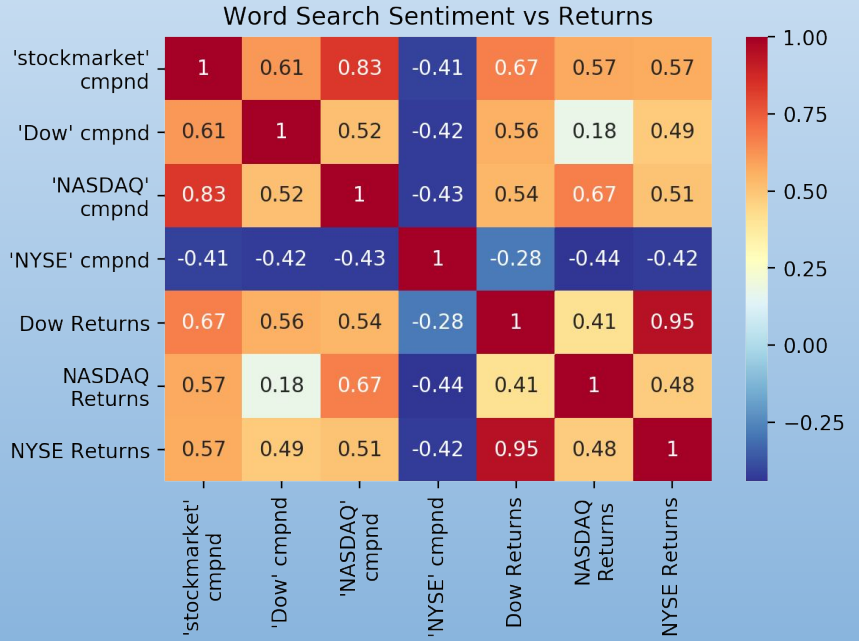


Daily NYSE % Return(t) = 0.81 * Wolfer's Sentiment (t-1) - 0.25

- This is a simple model - only one predictor is used.
- The historical timeframe is small, and does not cover a wide range of market conditions.
- Due to data constraints - using the same dataset to develop and test predictions. Would split testing and validation data in a more robust approach.

SAME DAY SENTIMENT & STOCK PERFORMANCE





Reactor:
Tweets on the "Dow Jones"

INSIGHTS

- Limited dataset does show some relationships (high correlation).
- Analyzed sentiment values for select target users compared to next day's market value.
- Highest pairwise correlation was 92% between yesterday's Justin Wolfers sentiment and today's NYSE performance.
- Further analysis on target term suggested today's twitter sentiment is influenced by yesterday's market performance.

RECOMMENDATIONS

- Initial results suggest potential relationships.
- In order to provide actionable results - analysis of a more robust dataset would be required.
- More extensive data generation should be investigated (consider average of users or trailing daily averages for example).
- More robust models should also be investigated (consider multiple factor models).

Q & A