



# MARKET ANALYSIS

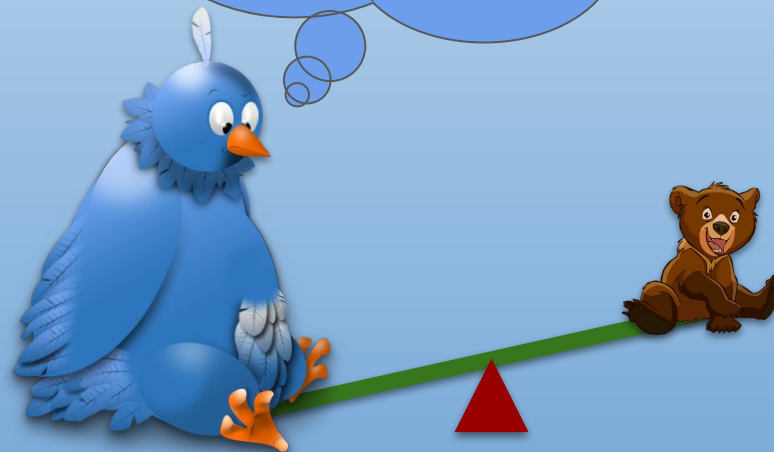
Tommy Steed, Joshua Bui,  
Irakli Intskirveli, Ethan Litke

# SUSPICION

There is a relationship  
between Twitter sentiment and  
stock market performance



Who could be the  
predictor or  
reactor?



# 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>

1=&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 datarange: `twitterscraper Trump%20since%3A2017-01-03%20until%3A2017-01-04 -o tweets.json`
- based on a datarange 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%3AA1West13 -o tweets.json`

**Advanced search**

**Words**

All of these words

This exact phrase

Any of these words

None of these words

These hashtags

Written in  All languages

**People**

From these accounts

To these accounts

Mentioning these accounts

**Places**

Near this place  [Add location](#)

**Dates**

From this date  to

[Search](#)

```
#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:

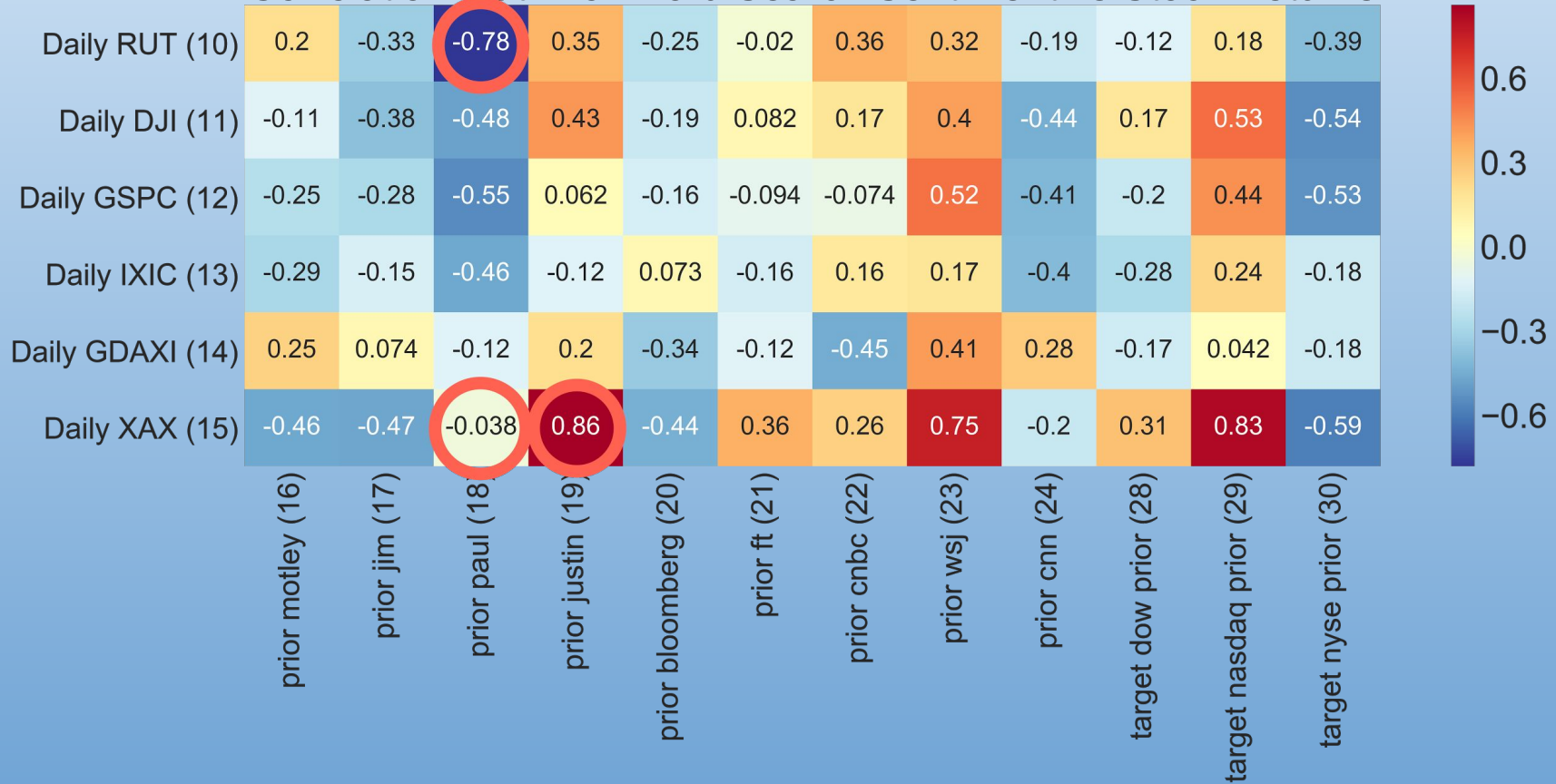
- wsjmarkets
- cnbc
- themotleyfool
- cnnmoney
- markets
- financialtimes
- nytimesbusiness
- paulkrugman
- JustinWolfers
- jimcramer





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

## Correlation Matrix of Word Search Sentiment vs Stock Returns





## Correlation Matrix Word Search Sentiment vs Stock Returns

Motley Fool Sentiment (1)	1	0.43	-0.27	-0.4	0.15	-0.42	-0.09	-0.41	0.42	-0.41	-0.21	-0.35	-0.67	0.39	-0.57	0.28	0.16	-0.06	-0.43	0.31	-0.42	-0.35	-0.57	0.11	-0.45	-0.81	0.52	-0.033	-0.4	0.2
Jim Cramer Sentiment (2)	0.43	1	0.28	-0.058	0.28	-0.47	-0.53	-0.2	0.76	-0.29	-0.37	-0.62	-0.63	0.13	0.02	0.062	-0.3	0.086	0.32	-0.3	0.19	0.18	-0.079	0.26	-0.41	-0.37	0.89	0.17	-0.12	-0.08
Paul Krugman Sentiment (3)	-0.27	0.28	1	0.068	0.025	0.37	-0.2	0.061	0.19	0.52	0.43	-0.02	-0.19	0.24	0.64	0.24	-0.21	-0.11	0.81	-0.21	0.26	0.48	0.26	0.2	0.32	0.1	0.31	0.39	0.3	-0.22
Justin Wolfers Sentiment (4)	-0.4	-0.058	0.068	1	-0.31	0.38	0.39	0.47	0.086	0.24	0.27	0.13	0.33	-0.45	0.28	-0.53	-0.49	-0.092	0.39	-0.29	0.39	0.49	-0.013	-0.79	0.36	0.7	-0.12	0.3	0.43	-0.18
Bloomberg Sentiment (5)	0.15	0.28	0.025	-0.31	1	-0.31	0.12	-0.63	0.23	-0.45	0.015	-0.03	-0.5	0.5	0.39	-0.42	-0.17	0.57	0.28	-0.17	0.34	-0.25	0.51	0.15	0.4	-0.075	0.14	0.34	0.45	-0.38
Fin Times Sentiment (6)	-0.42	-0.47	0.37	0.38	-0.31	1	0.64	0.41	-0.19	0.3	0.36	0.18	0.3	-0.15	0.34	-0.099	0.34	0.05	0.41	0.078	-0.25	-0.035	-0.052	-0.29	0.56	0.35	-0.43	0.057	0.39	0.35
CNBC Sentiment (7)	-0.09	-0.53	-0.2	0.39	0.12	0.64	1	-0.05	-0.14	-0.24	0.22	0.13	0.045	-0.094	0.076	-0.47	0.24	0.44	0.056	0.23	-0.056	-0.24	-0.13	-0.55	0.68	0.24	-0.46	0.27	0.46	0.22
WSJ Sentiment (8)	-0.41	-0.2	0.061	0.47	-0.63	0.41	-0.05	1	0.045	0.58	0.25	0.36	0.54	-0.11	0.15	0.062	0.075	-0.51	0.2	-0.54	-0.22	-0.07	0.17	-0.23	-0.14	0.52	-0.44	-0.61	-0.063	0.22
CNN Sentiment (9)	0.42	0.76	0.19	0.086	0.23	-0.19	-0.14	0.045	1	-0.32	-0.24	-0.58	-0.77	0.32	-0.079	0.18	0.06	0.28	0.34	-0.55	0.19	-0.13	-0.037	0.23	-0.33	-0.34	0.6	-0.033	-0.25	0.28
Daily RUT (10)	-0.41	-0.29	0.52	0.24	-0.45	0.3	-0.24	0.58	-0.32	1	0.78	0.69	0.5	0.2	0.47	0.2	-0.33	-0.78	0.35	-0.25	-0.02	0.36	0.32	-0.19	0.19	0.49	-0.33	-0.12	0.18	-0.39
Daily DJI (11)	-0.21	-0.37	0.43	0.27	0.015	0.36	0.22	0.25	-0.24	0.78	1	0.8	0.22	0.5	0.61	-0.11	-0.38	-0.48	0.43	-0.19	0.082	0.17	0.4	-0.44	0.61	0.48	-0.42	0.17	0.53	-0.54
Daily GSPC (12)	-0.35	-0.62	-0.02	0.13	-0.03	0.18	0.13	0.36	-0.58	0.69	0.8	1	0.61	0.33	0.47	-0.25	-0.28	-0.55	0.062	-0.16	-0.094	-0.074	0.52	-0.41	0.46	0.61	-0.75	-0.2	0.44	-0.53
Daily IXIC (13)	-0.67	-0.63	-0.19	0.33	-0.5	0.3	0.045	0.54	-0.77	0.5	0.22	0.61	1	-0.49	0.19	-0.29	-0.15	-0.46	-0.12	0.073	-0.16	0.16	0.17	-0.4	0.18	0.69	-0.69	-0.28	0.24	-0.18
Daily GDAXI (14)	0.39	0.13	0.24	-0.45	0.5	-0.15	-0.094	-0.11	0.32	0.2	0.5	0.33	-0.49	1	0.26	0.25	0.074	-0.12	0.2	-0.34	-0.12	-0.45	0.41	0.28	0.096	-0.25	-0.022	-0.17	0.042	-0.18
Daily XAX (15)	-0.57	0.02	0.64	0.28	0.39	0.34	0.076	0.15	-0.079	0.47	0.61	0.47	0.19	0.26	1	-0.46	-0.47	-0.038	0.86	-0.44	0.36	0.26	0.75	-0.2	0.72	0.66	-0.2	0.31	0.83	-0.59
prior motley (16)	0.28	0.062	0.24	-0.53	-0.42	-0.098	-0.47	0.062	0.18	0.2	-0.11	-0.25	-0.29	0.25	-0.46	1	0.49	-0.24	-0.26	0.077	-0.26	-0.068	-0.29	0.7	-0.6	-0.62	0.24	-0.35	-0.81	0.45
prior jim (17)	0.16	-0.3	-0.21	-0.49	-0.17	0.34	0.24	0.075	0.06	-0.33	-0.38	-0.28	-0.15	0.074	-0.47	0.49	1	0.33	-0.36	0.19	-0.57	-0.7	0.52	-0.26	-0.49	-0.24	-0.53	-0.51	0.89	
prior paul (18)	-0.06	0.086	-0.11	-0.092	0.57	0.05	0.44	-0.51	0.28	-0.78	-0.48	-0.55	-0.46	-0.12	-0.038	-0.24	0.33	1	0.1	0.018	0.38	-0.2	0.074	0.25	0.26	-0.17	0.12	0.33	0.095	0.31
prior justin (19)	-0.43	0.32	0.81	0.39	0.28	0.41	0.056	0.2	0.34	0.35	0.43	0.062	-0.12	0.2	0.86	-0.26	-0.36	0.1	1	-0.54	0.43	0.34	0.53	-0.091	0.54	0.45	0.14	0.37	0.62	-0.31
prior bloomberg (20)	0.31	-0.3	-0.21	-0.29	-0.17	0.078	0.23	-0.54	-0.55	-0.25	-0.19	-0.16	0.073	-0.34	-0.44	0.077	0.19	0.018	-0.54	1	-0.4	0.079	-0.69	-0.081	0.029	-0.43	0.083	0.33	-0.079	0.12
prior ft (21)	-0.42	0.19	0.26	0.39	0.34	-0.25	-0.056	-0.22	0.19	-0.02	0.082	-0.094	-0.16	-0.12	0.36	-0.26	-0.57	0.38	0.43	-0.4	1	0.61	0.45	-0.067	0.3	0.4	0.19	0.57	0.26	-0.5
prior cnbc (22)	-0.35	0.18	0.48	0.49	-0.25	-0.035	-0.24	-0.07	-0.13	0.36	0.17	-0.074	0.16	-0.45	0.26	-0.068	-0.7	-0.2	0.34	0.079	0.61	1	-0.061	-0.26	0.16	0.32	0.37	0.66	0.21	-0.52
prior wsj (23)	-0.57	-0.079	0.26	-0.013	0.51	-0.052	-0.13	0.17	-0.037	0.32	0.4	0.52	0.17	0.41	0.75	-0.29	-0.27	0.074	0.53	-0.69	0.45	-0.061	1	0.13	0.41	0.58	-0.39	-0.11	0.45	-0.49
prior cnn (24)	0.11	0.26	0.2	-0.79	0.15	-0.29	-0.55	-0.23	0.23	-0.19	-0.44	-0.41	-0.4	0.28	-0.2	0.7	0.52	0.25	-0.091	-0.081	-0.067	-0.26	0.13	1	-0.49	-0.59	0.28	-0.32	-0.59	0.37
target dow (25)	-0.45	-0.41	0.32	0.36	0.4	0.56	0.68	-0.14	-0.33	0.19	0.61	0.46	0.18	0.096	0.72	-0.6	-0.26	0.26	0.54	0.029	0.3	0.16	0.41	-0.49	1	0.57	-0.42	0.56	0.89	-0.42
target nasdaq (26)	-0.81	-0.37	0.1	0.7	-0.075	0.35	0.24	0.52	-0.34	0.49	0.48	0.61	0.69	-0.25	0.66	-0.62	-0.49	-0.17	0.45	-0.43	0.4	0.32	0.58	-0.59	0.57	1	-0.57	0.074	0.65	-0.47
target nyse (27)	0.52	0.89	0.31	-0.12	0.14	-0.43	-0.46	-0.44	0.6	-0.33	-0.42	-0.75	-0.69	-0.022	-0.2	0.24	-0.24	0.12	0.14	0.083	0.19	0.37	-0.39	0.28	-0.42	-0.57	1	0.41	-0.25	-0.02
target dow prior (28)	-0.033	0.17	0.39	0.3	0.34	0.057	0.27	-0.61	-0.033	-0.12	0.17	-0.2	-0.28	-0.17	0.31	-0.35	-0.53	0.33	0.37	0.33	0.57	0.66	-0.11	-0.32	0.56	0.074	0.41	1	0.52	-0.48
target nasdaq prior (29)	-0.4	-0.12	0.3	0.43	0.45	0.39	0.46	-0.063	-0.25	0.18	0.53	0.44	0.24	0.042	0.83	-0.81	-0.51	0.095	0.62	-0.079	0.26	0.21	0.45	-0.59	0.89	0.65	-0.25	0.52	1	-0.59
target nyse prior (30)	0.2	-0.08	-0.22	-0.18	-0.38	0.35	0.22	0.22	0.28	-0.39	-0.54	-0.53	-0.18	-0.18	-0.59	0.45	0.89	0.31	-0.31	0.12	-0.5	-0.52	-0.49	0.37	-0.42	-0.47	-0.02	-0.48	-0.59	1
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

0.8

0.4

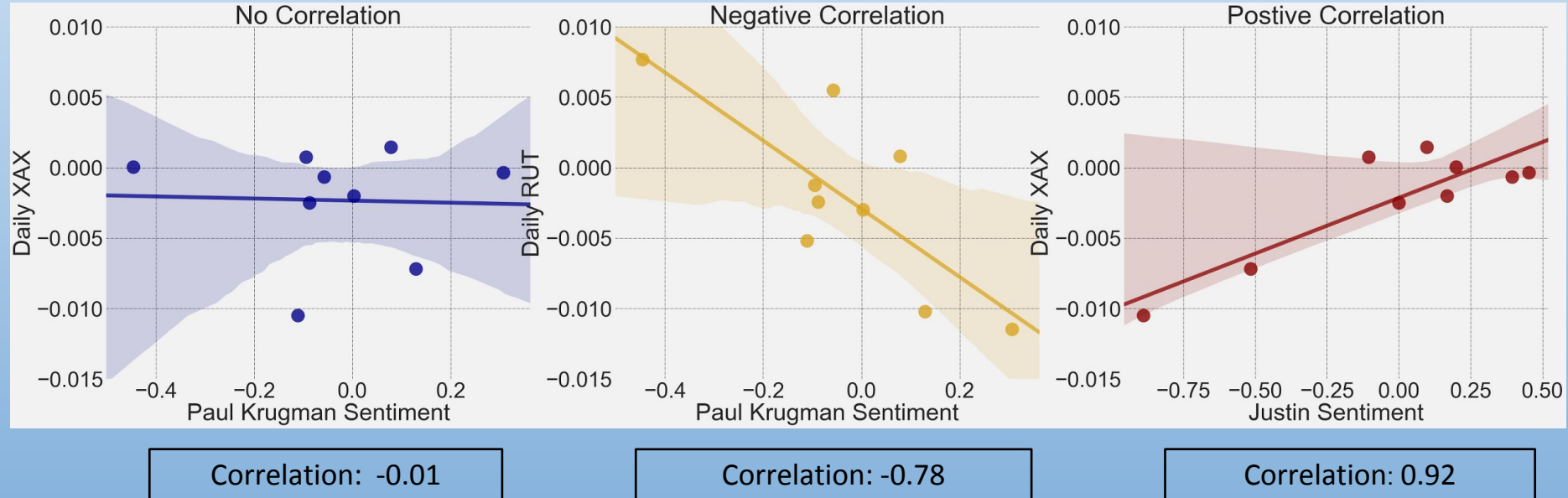
0.0

-0.4

-0.8



# CORRELATIONS



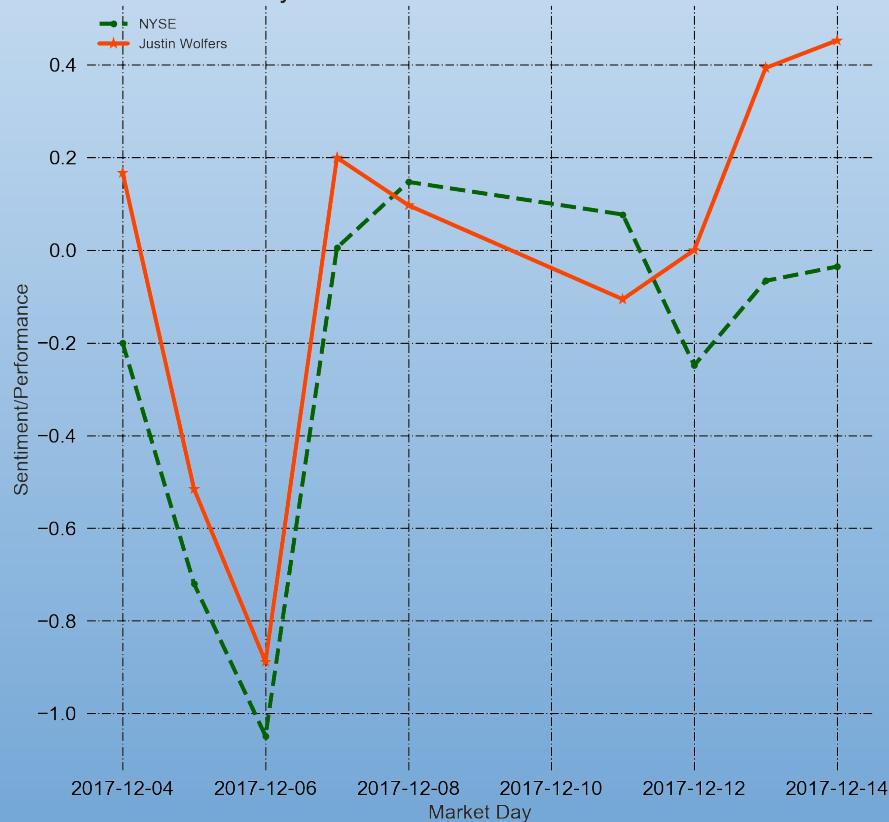
All within a 95% confidence level





# HISTORICAL PERFORMANCE

Justin Wolfer's  
1-Day Before Sentiment & NYSE Returns

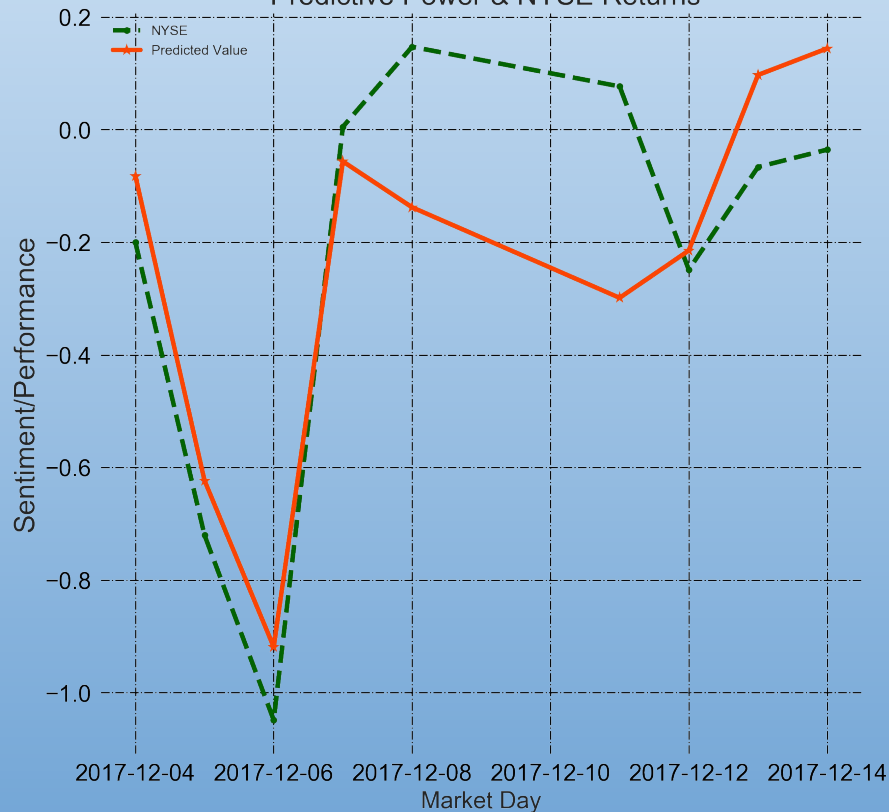


- Justin Wolfer's Sentiment is from the prior day.
- Movements are in the same direction -- in line with the 92% correlation.



# PREDICTED VALUES

Justin Wolfer's  
Predictive Power & NYSE Returns



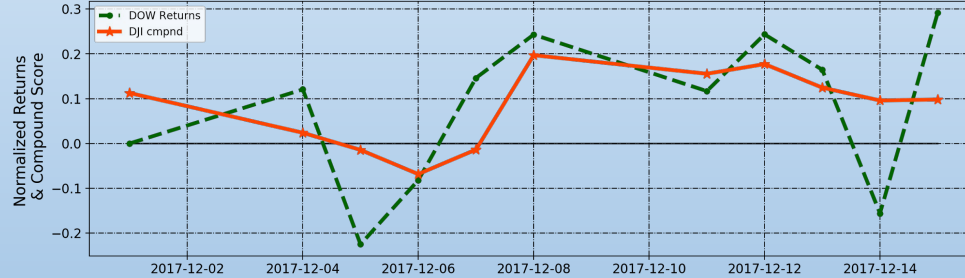
$$\text{Daily NYSE \% Return}(t) = 0.81 * \text{Wolfer's Sentiment}(t-1) - 0.25$$

- 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.

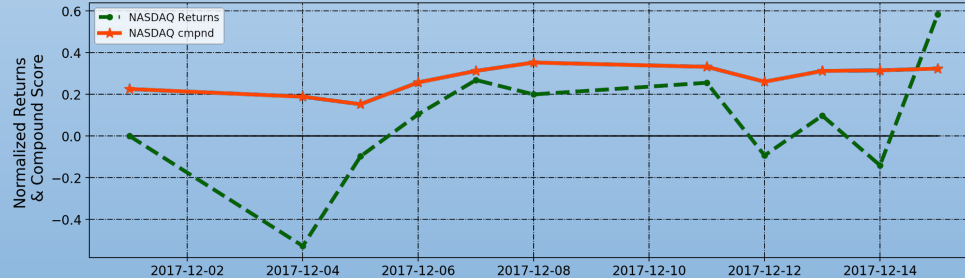


# SAME DAY SENTIMENT & STOCK PERFORMANCE

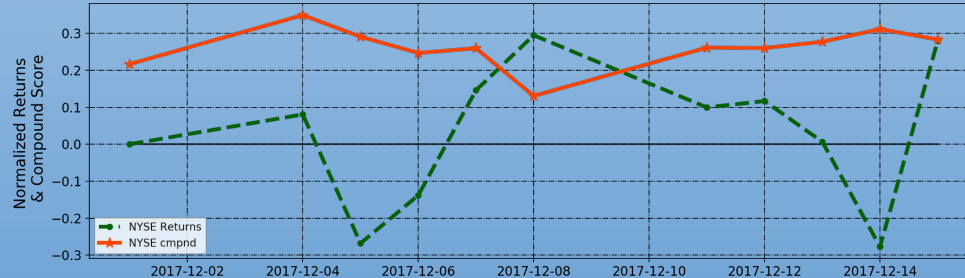
"DowJones" Returns & Word Search Sentiment



"NASDAQ" Returns & Word Search Sentiment



"NYSE" Returns & Word Search Sentiment



Word Search Sentiment vs Returns

	'stockmarket' cmpnd	'Dow' cmpnd	'NASDAQ' cmpnd	'NYSE' cmpnd	Dow Returns	NASDAQ Returns	NYSE Returns
'stockmarket' cmpnd	1	0.61	0.83	-0.41	0.67	0.57	0.57
'Dow' cmpnd	0.61	1	0.52	-0.42	0.56	0.18	0.49
'NASDAQ' cmpnd	0.83	0.52	1	-0.43	0.54	0.67	0.51
'NYSE' cmpnd	-0.41	-0.42	-0.43	1	-0.28	-0.44	-0.42
Dow Returns	0.67	0.56	0.54	-0.28	1	0.41	0.95
NASDAQ Returns	0.57	0.18	0.67	-0.44	0.41	1	0.48
NYSE Returns	0.57	0.49	0.51	-0.42	0.95	0.48	1

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