Overview

The stock market is a complex machine, made up of various indices and thousands of individual securities. Firms, individuals, and even computer applications are constantly buying and selling which results in price fluctuations. Many factors influence stock market movements, such as company earning reports, geopolitics, and statements from influential individuals and organizations.

Twitter is a social media platform that allows individuals to type their thoughts for other users to see in real time. The combination of Twitter's chronological newsfeed and 140 character limit make it an excellent place to find succinct and up-to-date news.

When influential individuals use Twitter, they can reach a very large audience. We noticed that when presidential candidates, for example, tweeted about a certain company or industry, that security could fluctuate much more than usual. We decided to investigate this phenomenon further.

Research Questions/Purpose

There were several particular questions we were interested in answering this project. These questions were used to guide our research, analysis, and implementation.

1. What is the relationship between sector fund returns, volume, and tweet data?
2. Can powerful or famous individuals influence a sector funds' price and if so, to what degree?
3. How long do these tweets affect the price movement of a security?

Related Work

The first study we encountered, "Stock market one-day ahead movement prediction using disparate data sources" (Megahed, et. al), analyzes the next-day price movements of stocks on a number of factors. A few of these factors include Google News articles and Wikipedia page visits for specific topics. This study helped us decide that we wanted to focus on social media and short term price movements. It also helped us locate many other useful studies. Link: http://www.sciencedirect.com/science/article/pii/S0957417417301331#bib0039

"Sentiment analysis on social media for stock movement prediction" (Nguyen, et. al) is a fascinating study that uses Yahoo! Finance message boards to predict stock price movement. It touches on some of the challenges of studying social media sentiment with stock prices and the successes they had. Link: http://www.sciencedirect.com/science/article/pii/S0957417415005126

This study, titled "Twitter mood predicts the stock market", details how the authors collected Twitter data, classified the mood or sentiment of the tweet, and analyzed daily sentiment versus the DJIA returns. While their type of Twitter data focuses on many users and ours focuses on few users, their statistical methods could be helpful in creating a predictive model.

Link: http://www.sciencedirect.com/science/article/pii/S187775031100007X

Potential Twitter Accounts

Donald Trump @realDonaldTrump
Hillary Clinton @HillaryClinton
Steve Burns @SJosephBurns
Scott Redler @RedDogT3
MarketWatch @MarketWatch
TheStreet @TheStreet

Exploratory Data Analysis

We began our data analysis by downloading the most recent tweets from the above account. After opening the Excel file for each account, we combined the tables into one Excel file (3832 Tweets). We then copied and pasted this complete table into another worksheet (Titled Tweets_Adjusted) and deleted the irrelevant columns. These columns either contained all of the same values or were of no interest to our analysis. In this worksheet we also adjusted the date and time of the tweets, which involved converting the original time (column “created”) from GMT to EST. Next, we uploaded the Excel sheet into Microsoft Access and called the project “Project_1_Database”. We used the filter options in Access to understand how many levels each variable had and ensured there were no null values. Then we ran a query called “Keywords” to select any tweet that contained any terms relative to the stock market or individual stocks. From here, we saved the query results as the table “Keyword_Tweets”. After doing this, we wrote queries to select positive tweets and negative tweets from the table that contained any keywords. We then copied these positive and negative tweets into Excel and coded them accordingly. Also, we ran a query in Access to identify tweets that were selected by both the positive and negative queries. These were coded as Neutral in Excel (0,0) for the columns (Positive, Negative). We then analyzed the tweet and assigned it to the appropriate State Street SPDR ETF, which tracks sector performance. Unfortunately, there is no easier way than doing this last step manually. Then, we created a column that returns only the date of the tweet. From here, we created data ranges in Excel of the tweets for each particular SPDR fund. Then, we connected the Excel files in Tableau on the date-only values.

Queries

Keywords:

```
SELECT *
FROM Tweets_Adjusted
WHERE (Text LIKE "*price*") OR (Text LIKE "*long*") OR (Text LIKE "*short*") OR (Text LIKE "*upgrade*") OR (Text LIKE "*$*") OR (Text LIKE "*downgrade*") OR (Text LIKE "*equity*") OR (Text LIKE "*upside*")
OR (Text LIKE "*downside*") OR (Text LIKE "*buy*") OR (Text LIKE "*sell*") OR (Text LIKE "*gain*") OR (Text LIKE "*loss*");
```

Positive Tweets:

```
SELECT *
FROM Keyword_Tweets
WHERE (Text LIKE "*buy*") OR (Text LIKE "*gain*") OR (Text LIKE "*upside*") OR (Text LIKE
"*long*") OR (Text LIKE "*bull*") OR (Text LIKE "*bullish*");
```

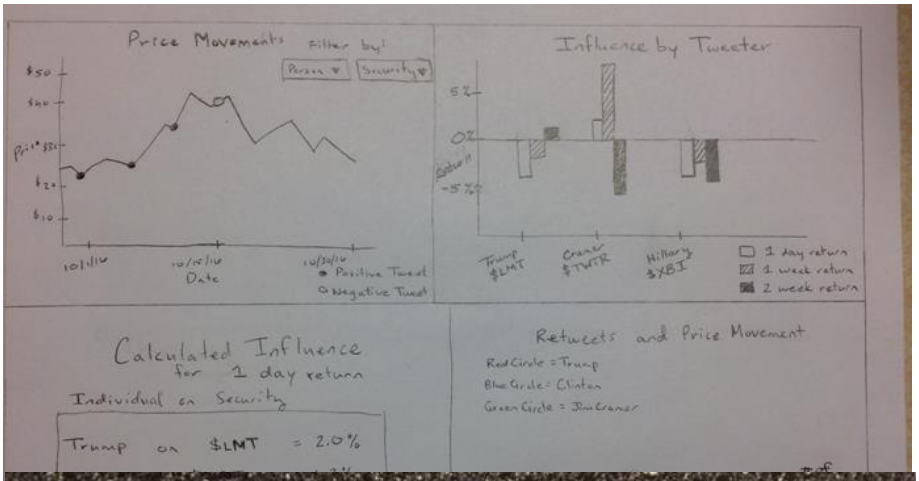
Negative Tweets:

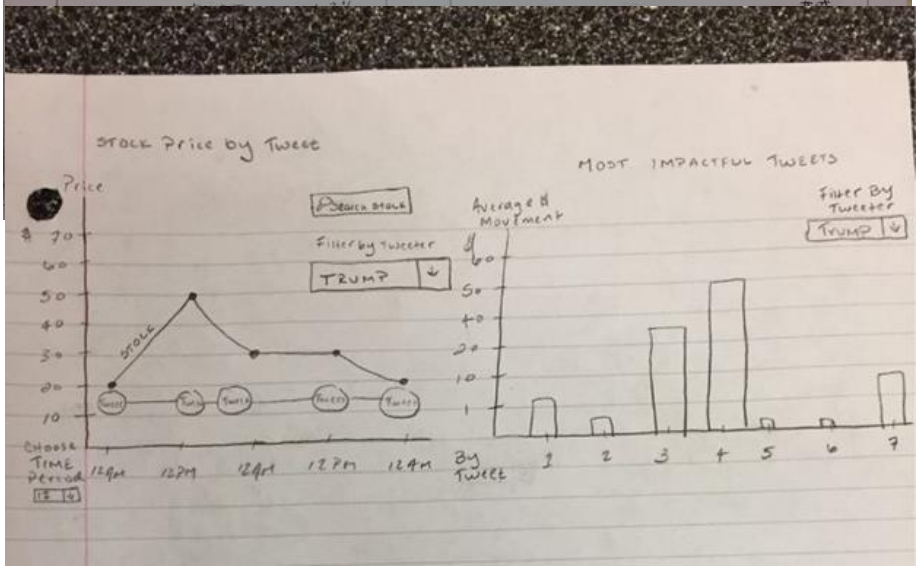
```
SELECT *
FROM Keyword_Tweets
WHERE (Text LIKE "*sell*") OR (Text LIKE "*loss*") OR (Text LIKE "*downside*") OR (Text
LIKE "*short*") OR (Text LIKE "*bear*") OR (Text LIKE "*bearish*");
```

Neutral Tweets:

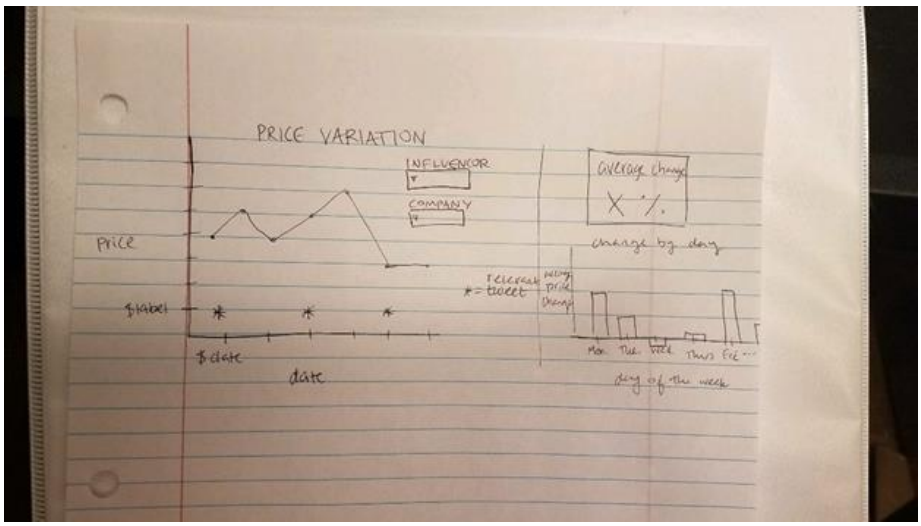
```
SELECT *
FROM Positive_Tweets
INNER JOIN Negative_Tweets ON Positive_Tweets.ID = Negative_Tweets.ID;
```

Sketches/Images of Visualizations

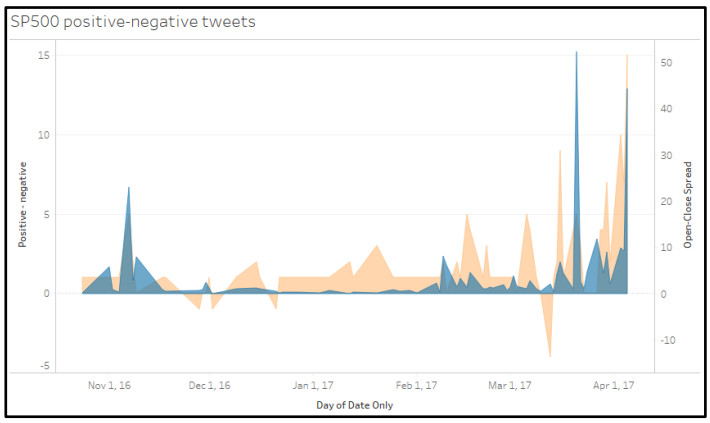


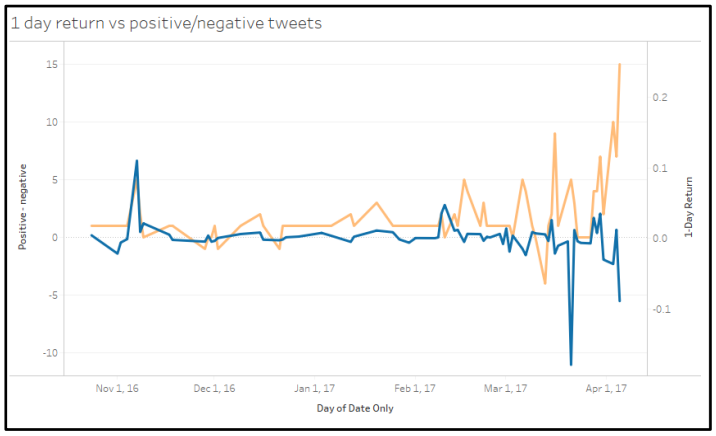


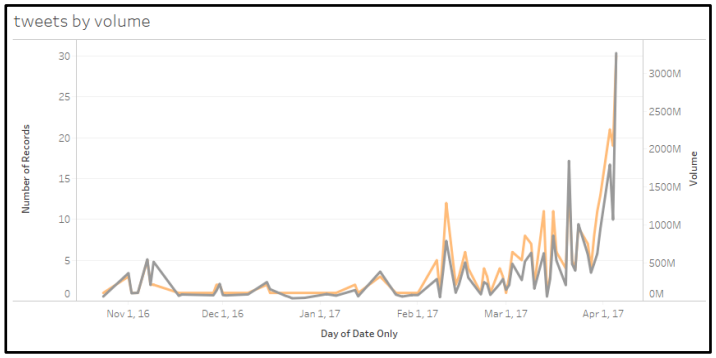
Shown below are some initial visualization ideas that we presented.

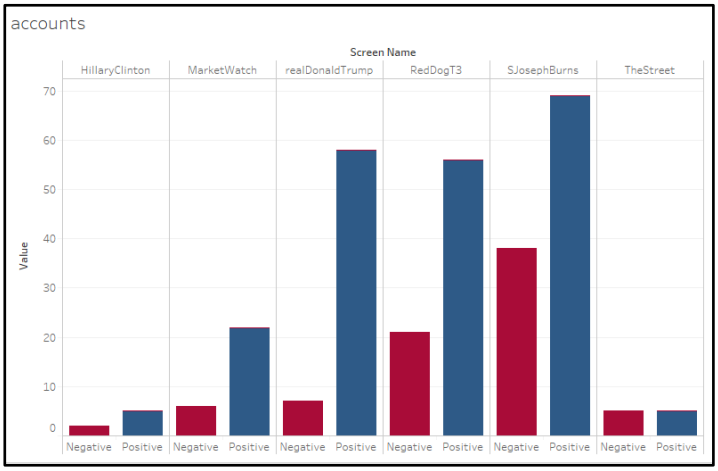


Dashboard Visualizations









Implementation

R

We used the TwitteR package within R mainly to retrieve data from Twitter. We then saved the data we retrieved from individual influencers in csv files before exporting them to Microsoft Excel and Microsoft Access.

Microsoft Access

Within Microsoft Access, we cleaned our data and got rid of unnecessary columns. Furthermore, we ran multiple queries to further analyze the phrasing and timing of various tweets. We then saved these queries as new tables to export.

Tableau

Within Tableau, we created our visualizations and eventually our dashboard.

Conclusion

To sum it all up, the activity of these Twitter accounts and their sentiment on the markets may be an indicator of market volatility, activity, and returns. Negative tweets and a trend towards negative tweet sentiment can influence the S&P 500's volatility and volume in the short term. We see less correlation between the number of tweets and actual returns in the S&P 500. The number of tweets, does however, look to be a leading indicator of volume. While we cannot prove causation, it appears that these influential Tweets can influence volume and volatility in the S&P 500. These effects appear to be short term, 1 to 2 weeks at the most. It is more difficult to measure how Tweeters affect the returns of the S&P 500. We believe this is partially due to the reason that market returns are an inherently noisy and complex variable to measure.

What does our research mean for investors?

For those with a long term investment horizon (1+ year), tweets should not be a concerning factor. For those investors that are aiming to buy and hold securities for less than a year, this information could be helpful. Risk-adverse investors may want to avoid investing on days that directly follow large changes in tweet sentiment and activity. The increased risk that comes with higher volatility and volume is not desirable for this investor. Risk-favoring investors can find buying opportunities that follow these tweet indicators by betting on volatility and large market swings. Short term volatility securities ($VIXY for example) may be of interest.