Final Project - Presidential Tweets

Author:
Parin Patel

Professor:
Jillian Lando

March 28, 2020

Contents

1   Introduction

President Trump is changing the political landscape by using his tweets to talk directly to the American people[1] in an attempt to move his policy agenda. By using Twitter, he is changing the discussion for Congress, the Judiciary, the media, and his own administration. This change has implications for policy based on the significant reactions to these tweets by both politicians and Federal employees (Graham, 2019).[7] Sentiment analysis[2] and natural language processing can help assess the potential impact of how these words(tweets) shape policy by his administrators, as well as the impact on his favorability ratings for re-election.

Natural language processing (NLP) is a methodology that allows researchers to analyze the sentence and language-related data using computational techniques to determine specific outcomes. NLP has a wide variety of uses, for example, it is used to predict words and texts on mobile phone messaging apps. It is also used to translate web pages and help users when conducting online searches. Within NLP, we specifically chose to use a tool called sentiment analysis, or opinion mining. It is the automated process of identifying and extracting the subjective information that underlies a text. This can be either an opinion, a judgment, or a feeling about a topic or subject. The most common type of sentiment analysis is called 'polarity detection' and consists of classifying a statement as 'positive', 'negative' or 'neutral' (MonkeyLearn, 2020).[9]

This project will evaluate Trump's tweets and determine trending using current data science techniques including NLP, Word Clouds and Sentiment analysis (IEE Explore, 2018).[8] In addition, we will also use the Trump approval rating, as measured by various polls, as an additional variable for comparison (trump-approval-ratings, 2020).[6]

2   Recommendations

Tweeting at critical times does impact policy and forces administrators to respond to the "tweet crisis" as demonstrated in 3 specific cases. (Border Security, Impeachment, AmericaFirst-TPP) Polling data shows that there are trends in tweeting and retweeting data that affect polling sentiment. This is seen in the data the week after either a high negative or high positive tweet.

For our model we chose the cosine model. This allows us to group similar tweets into categories to see if there is behavior before and after a policy change that influence the policy. Our three examples are: Impeachment, Trans Pacific Partnership and Border Security.

Presidential re-electability (PRE) based on tweets needs a different data set than is currently captured. We define the PRE using the polling data and can see the trends but to predict PRE need to map the policies that are being affected in order to predict PRE. We would need additional polls on economy, tax reform and other polls to truly determine PRE.

We believe that his tweeting (related to the market) does not largely affect his approval or disapproval rating.

Target Audience

- President: If you are the President understand the power of your position. Some tweets have negative effects. Understand the implications of a direct to voter tool like Tweets and how it impacts the voter base.

- Director/Policy Manager: If you are working in the White House you need to understand the implications of what would happen if policies in your department were impacted by a presidential Tweet. Understand the policies and positions of the president and have a plan for policies that are at risk of not moving in the direction of the president. Ensure that your department has a good understanding of the polling related to the policies in your department.

- Voter: As a voter you should understand what the data means and how it affects your views.

3   Specification

Our hypothesis states that Trump's tweets impact voters' perception of him and impact his chances for electability. We will measure this by analyzing data collected concerning his tweeting behavior and compare it with polling data in an attempt to find a correlative relationship. In addition to the primary hypothesis, we will also attempt to find correlation between tweeting behavior and stock market fluctuations.

1. Can we produce a best guess estimate for the re-election of a sitting president based on the Tweet sentiments?

2. Does the positive or negative tone of the tweets have an impact on his favorability based on consecutive poll data?

3. Does a specific word use have an impact on his poll ratings or the stock market?

4. From our models, are we able to compare successful outcomes to unsuccessful outcomes to see if the language between the two was consistent?

5. We can also use the same method in analyzing a drop-in poll's.

6. Is there a pattern of tweet habits? Often times with social media postings, there are specific "best times" to posts for highest visibility. Does President Trump's tweets follow or not follow these recommendations?

4 Observation

4.1 The Data

4.1.1 Tweet Data

The tweet data used in this paper was an open source dataset provided by TrumpTwitterArchive.[5] The dataset consisted of every tweet when from President Trump's official Twitter account form 1/1/2017 to 3/7/2020 (~16,000 rows). Each row represents a single tweet and contained the following features.

Data Features
Datetime Created
Source of Tweet (iphone, andriod, etc)
# of Retweets
# of Likes
Was Retweet

Since the majority of our analysis centers around the sentiment of Trump's tweeting we need a way to score each tweet in this regard. To do this analysis we used both polarity and sentiment analysis. We also used the Cosine model.

4.1.2 Polling Data

The polling data used was provided by the fivethirtyeight[6] project.

4.1.3 Stock Market Data

The stock market data used was provided by Kaggle[10] and Yahoo[11]

5 Analysis

We used 3 analysis techniques. Polarity (sentiment), Cosine and NLP. Polarity analysis is an efficient method that helps to understand how President Trump reaches out to his followers, the community, and the public. This analysis revolutionized the freedom of speech, the feelings and opinions of the individual into a specific topic. The data show that the majority of the people are consistent in their polls, despite the language. Overall the positive sentiment of the tweets has increased over the years.

The Cosine model was used to evaluate the consistency in tone and the language of the tweets to determine what President Trump wanted in policy and if the policy was subsequently implemented. This model calculated angle distance to generate comparable word weights that were used to grade similar tweet language. The model output was based on a single date of significance and generated the top five tweets that had the highest score in comparison to the words of the original tweet.

In the first model output on Border Security, the date was 27 January 2017, a time when Border Security was an issue for Trump. Subsequently, the Homeland Security Director, Kirstjen Michele Nielsen "resigned" after Trump took issue to the general weakness of the border and reflected a critical tone in his tweet on 27 January 2017. The model also generated five comparable tweets on border security and policy intent on 19 July 2017, 3 March 2017, 25 November 2018, and 4 June 2018. The policy direction of these comparable tweets were also respectively implemented as tariffs were officially enacted with China and the US ended participation in the TPP. The tone of his policy is therefore consistently defined in his tweet language and clearly results in direct action within his administration.

In the second example, a date within the Impeachment Trial of 24 Sept 2019 was chosen to further analyze the tweet language compared to the policy outcome. The model in this case returned language that consistently showed just how much the President played to his base and not to the overall public, to include Democrats, in these tweets. Instead he used tweets to encourage pressure from his base on Lawmakers as a resource of direct support in enacting preferable policy. Furthermore, on 7 February 2020, Trump used a tweet against Mitt Romney as a Senator who voted to impeach the President. The language is repetitive on 14 February 2020, along with a re-tweet of a supporter also against Mitt Romney. This language is once again a clear use of Twitter language to influence a desired outcome by encouraging public pressure from supporters.

5.1 Tweeting Patterns

We thought it would be interesting to analyze Trump's tweets temporally to see if any patterns arise in his behavior. Figure 1 plots Trump's daily tweet count over time. It is very easy to see from this plot that Trump has become increasingly more active as well as sporadic with his tweeting volume the further he has gone into his first term as president. The increase in tweets may also be a reflection of the success in generating policy results and favorable outcomes.
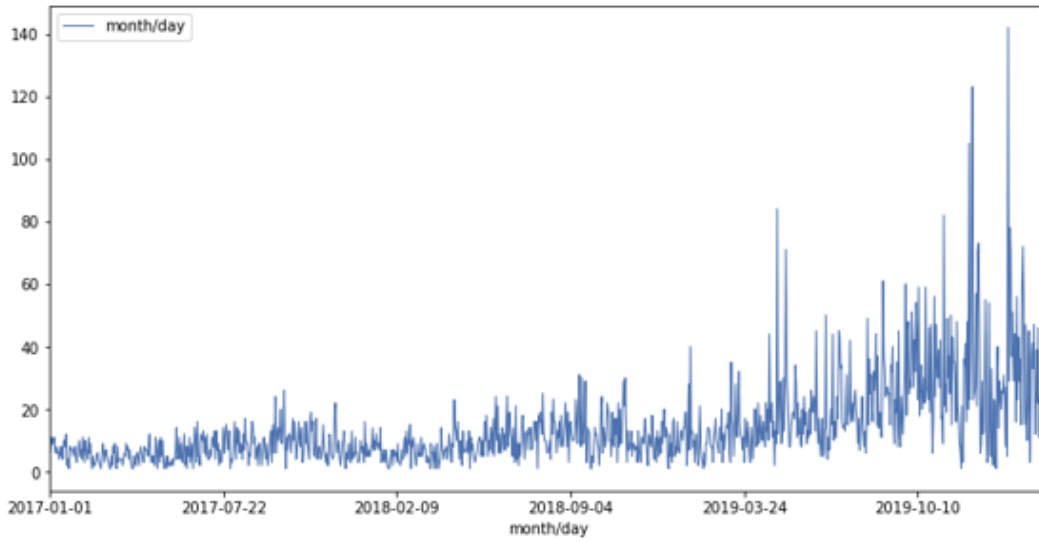
Figure 1: all Trump Tweets

Figure 2 paints a similar picture except it also demonstrates the fact that the majority of his tweets come from his iPhone. We hypothesis that non-iPhone tweets are actually authored by someone other than Trump and this is consistent with the observation that "Trump speaks directly to the people". His tweets often contain grammar and syntax errors as well as misspellings which further support this assertion. The following graphs help to visualize the data set.
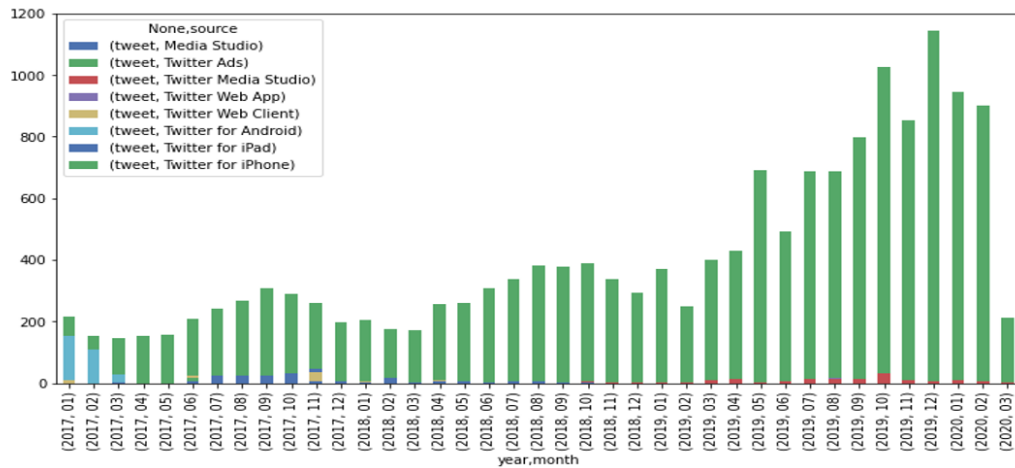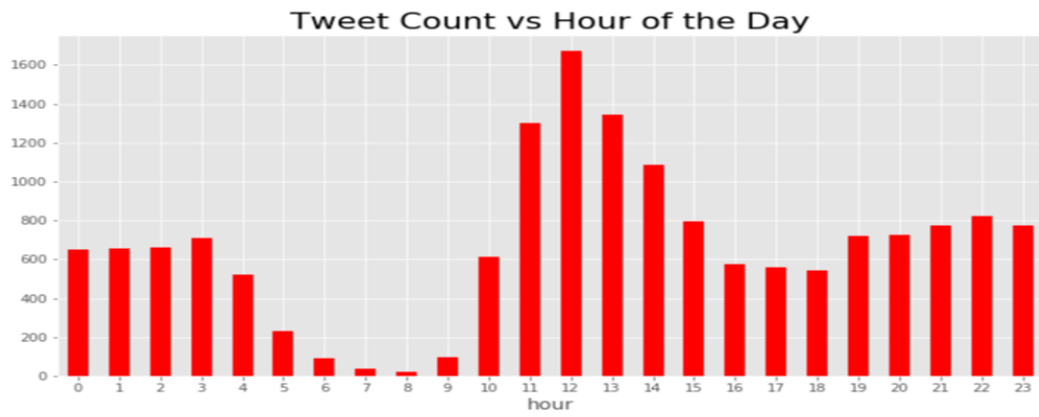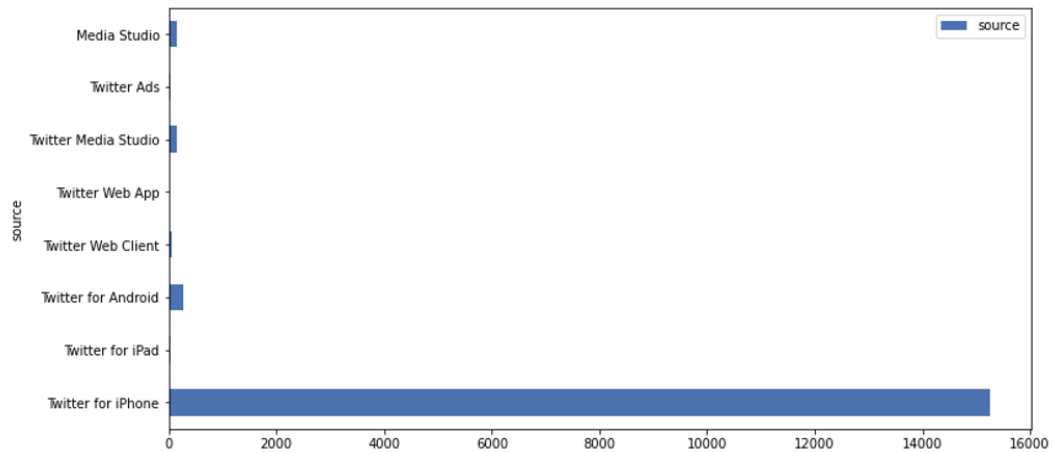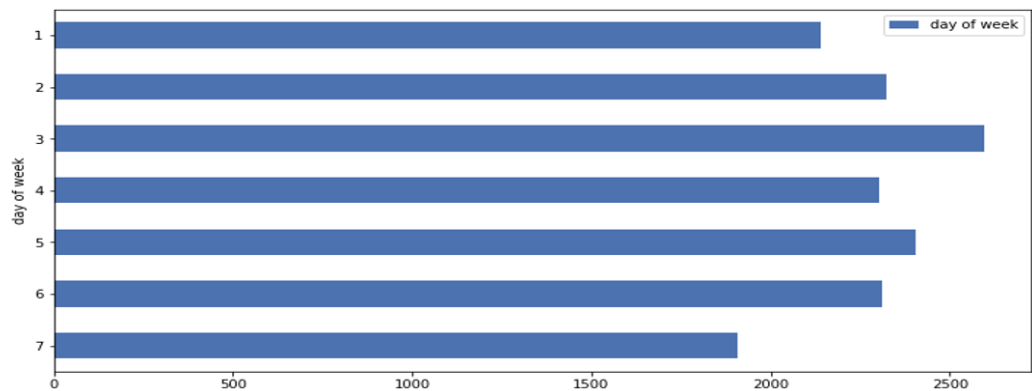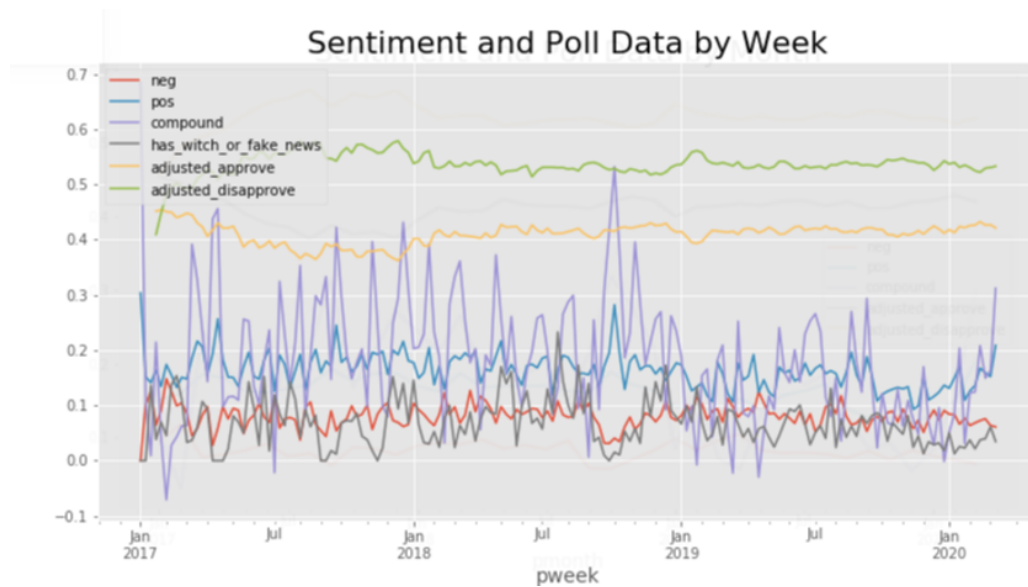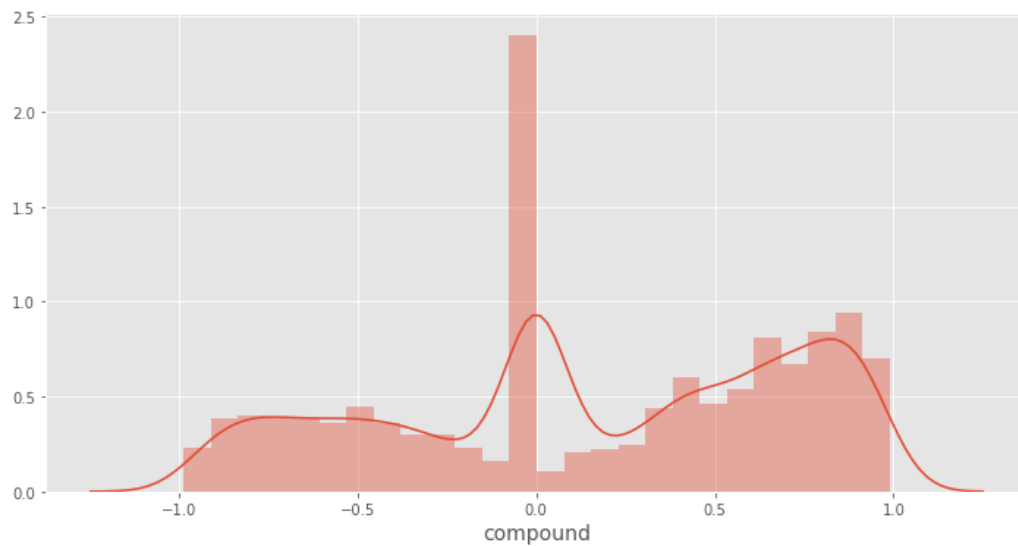

Figure 2: All Tweets by Source

Figure 3: Tweet Count vs Hour of Day

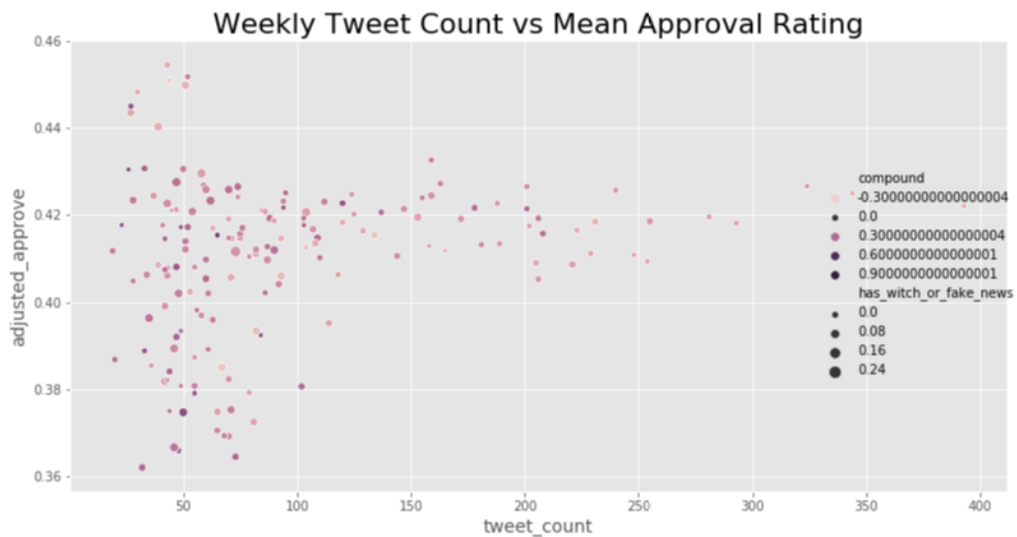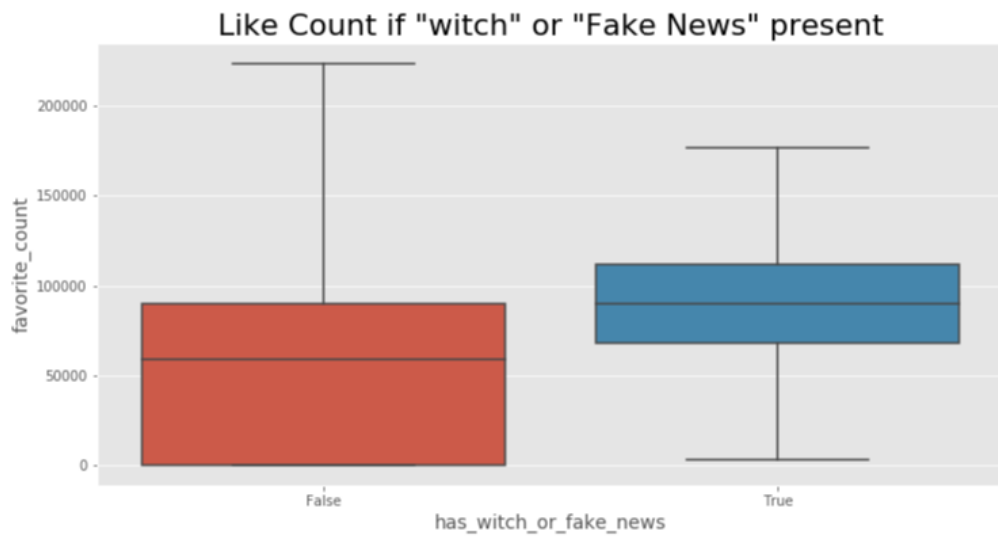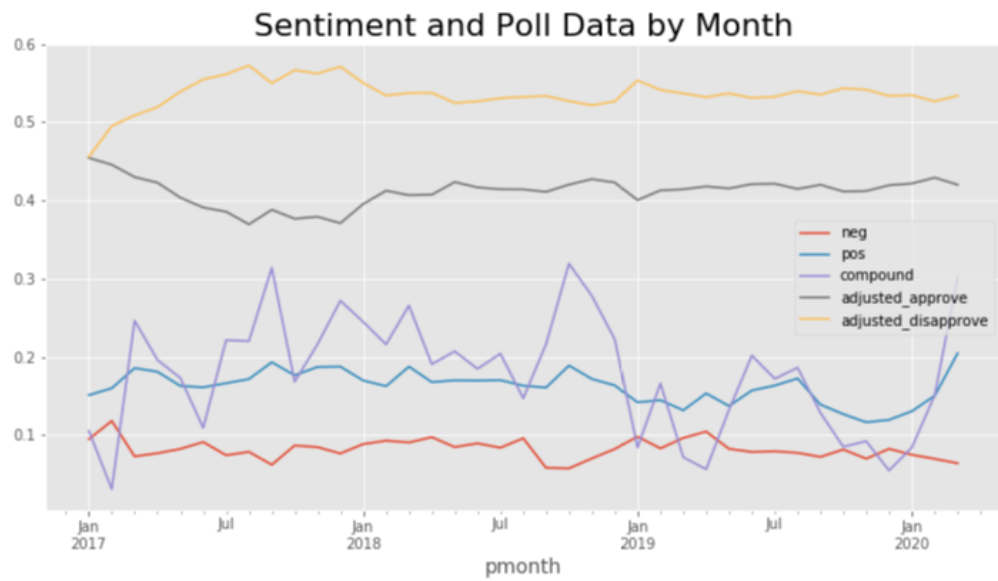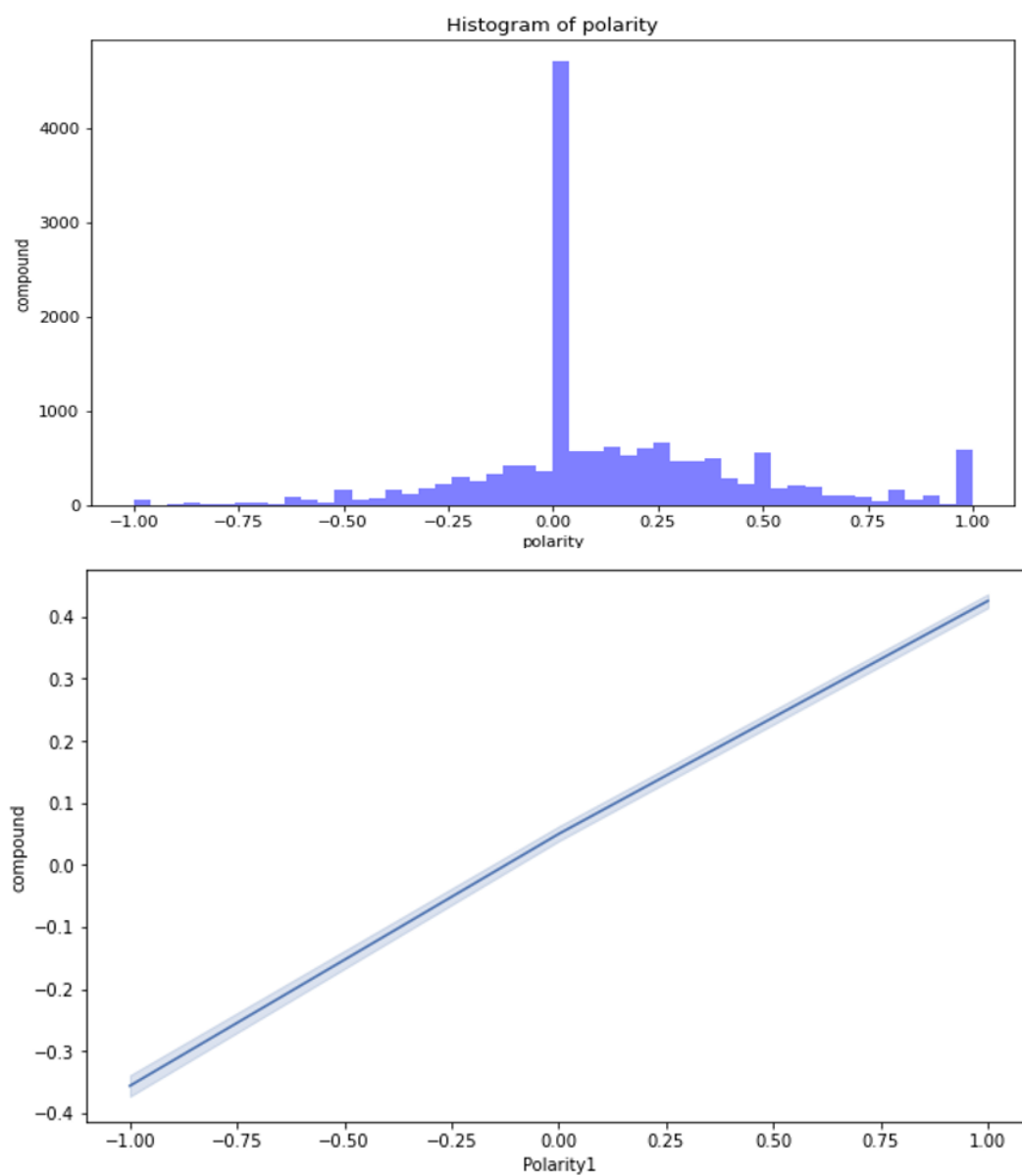## 5.2 Sentiment Analysis

The following 5 graphs help to tell the story about the sentiment analysis. We not only see the polarity but also track the weekly tweet count and compare it to the mean approval rating.
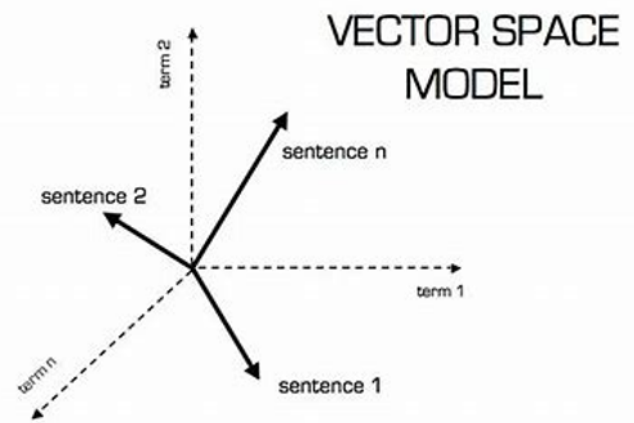

Sentiment and Poll Data by Week

Sentiment and Poll Data by Month

Like Count if "witch" or "Fake News" present

Weekly Tweet Count vs Mean Approval Rating

Histogram of polarity

5.3 The Cosine Model

The cosine model shows the similarities between tweets and we used the policy tracker. This model calculated angle distance to generate comparable word weights that were used to grade similar tweet language. Re-tweets were not included in the model unless directly entered by the President, although these re-tweets had significance in showing unified approval and applying direct pressure from supporters.

	0	1	2	3	4	5	6	7	8	9	...	13122	13123	13124	13125	13126	13127	13128	
0	1.000000	0.059846	0.000000	0.0	0.0	0.358715	0.189661	0.0	0.007911	0.0	...	0.012778	0.000000	0.000000	0.0	0.000000	0.030580	0.000000	0
1	0.059846	1.000000	0.075989	0.0	0.0	0.045831	0.056999	0.0	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.025845	0
2	0.000000	0.075989	1.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0
3	0.000000	0.000000	0.000000	1.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0
4	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
13127	0.030580	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.043258	1.000000	0.000000	0
13128	0.000000	0.025845	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.036039	0.000000	0.008854	0.0	0.000000	0.000000	1.000000	0
13129	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.018366	0.000000	0.020135	0.0	0.000000	0.000000	0.035787	1
13130	0.062243	0.037534	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.011229	0.0	...	0.000000	0.030867	0.000000	0.0	0.000000	0.043404	0.041345	0
13131	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.000000	0.000000	0.020509	0.0	0.000000	0.000000	0.028261	0

13132 rows × 13132 columns

```
Original Tweet from 2017-01-27: Mexico has taken advantage of the U.S. for long enough. Massive trade deficits &amp; little he
lp on the very weak border must change NOW!
-------------------------------------------------------------------

Similar Tweet from 2018-07-19 (Similarity=0.278) I told you so! The European Union just slapped a Five Billion Dollar fine on
one of our great companies Google. They truly have taken advantage of the U.S. but not for long!

Similar Tweet from 2017-03-30 (Similarity=0.274) The meeting next week with China will be a very difficult one in that we can
no longer have massive trade deficits...

Similar Tweet from 2018-11-25 (Similarity=0.234) Europe has to pay their fair share for Military Protection. The European Unio
n for many years has taken advantage of us on Trade and then they don□t live up to their Military commitment through NATO. Th
ings must change fast!

Similar Tweet from 2018-06-04 (Similarity=0.231) Farmers have not been doing well for 15 years. Mexico Canada China and others
have treated them unfairly. By the time I finish trade talks that will change. Big trade barriers against U.S. farmers and oth
```

Original Tweet from 2019-09-24: They (Dems) are scrambling for a theme and narrative. They've gone everywhere from Russian Hoax to Russian Collusion...and now they've come to this...they think they should have won the 2016 election they think in their biz arre brains that they did... https://t.co/xqYFEAzT8D

Similar Tweet from 2018-03-03 (Similarity=0.276) Mainstream Media in U.S. is being mocked all over the world. They□ve gone CRAZ Y! https://t.co/4UGYuJpUA7

Similar Tweet from 2019-06-13 (Similarity=0.270) They□ve been wrong all along! https://t.co/z5tSV2JLEf

Similar Tweet from 2019-10-16 (Similarity=0.228) Do you think they like me? https://t.co/TDmUnJ8HtF

Similar Tweet from 2019-03-09 (Similarity=0.211) RT @dbongino: Is there anything that the Democrats stand for that aligns with American values anymore? Now that they□ve gone all-in on anti□

Original Tweet from 2020-02-07: ....always continue. Every Republican Senator except Romney many highly religious people all v ery smart voted against the Impeachment Hoax. @SenCapito was all in (a great person). I was told by many that Manchin was just a puppet for Schumer & Pelosi. That□s all he is!

Similar Tweet from 2020-02-14 (Similarity=1.000) ....always continue. Every Republican Senator except Romney many highly relig ious people all very smart voted against the Impeachment Hoax. @SenCapito was all in (a great person). I was told by many that Manchin was just a puppet for Schumer & Pelosi. That□s what he is!

Similar Tweet from 2020-02-09 (Similarity=0.130) Romney hurt some very good Republican Senators and he was wrong about the Imp eachment Hoax. No clue! https://t.co/J6rtaZYYxD

Similar Tweet from 2019-07-04 (Similarity=0.122) Tammy is a great person! https://t.co/GRyl1OMsux

Similar Tweet from 2020-02-08 (Similarity=0.116) Can□t say I mind the fact that the great people of West Virginia are furious at their puppet Democrat Senator Joe Manchin. They will never forget his phony vote on the Impeachment Hoax. All he had to do

---

Big League Politics @bigleaguepol · Feb 5

Replying to @KimStrassel

Mitt Romney has stabbed Trump in the back and will join Democrats in their shameful attempt to overturn an election.

Mitt may think it's payback...

BUT Utah is preparing a process to #RECALLMITT!

It would be a shame if this story went viral.

Bye Mitt? Utah Legislature Prepares Senator Recall Process - Big League...
The new measure could be used to remove Mitt Romney from office.
🔗 bigleaguepolitics.com

💬 1.5K   ♻ 10.7K   ♡ 25.1K   ⇪

5.4   Word Clouds

Natural Language Processing allows us to create word clouds, compare sentiments and process the text for comparing Tweets and responses. This tweet was seen as positive.

Word Clouds are a collection of words sorted by count and the more the word appears the bolder it is.

Figure 18 an example of one of President Trumps positive Tweets as the sentiment analysis showed.

WordClouds give us a way to quickly visualize the content of a tweet and show the relationship of the words in the tweet. We can also compare the tweets using the cosine analysis as described above. These visualization tools allow us to describe and tag the words in ways that allow us to see

how the tweets affect policy.

The following word clouds are examples of positive and negative and plaintive tweeting. This is important because it is critical to understand if are undermining your own policies when a negative tweet does not have the effect that you are looking for.
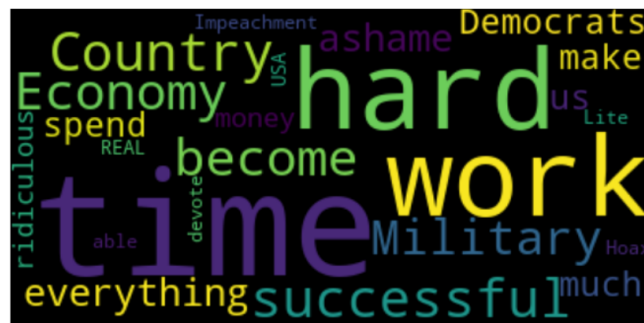
This is an example of a positive tweet.

This is an example of a negative tweet.

Although it is clear that the tweet was intended for a specific audience you can see that it will lose context very quickly when removed from the thread that generated the tweet.

When President Trump wants to know what the American voters to know how he feels he will tweet things that sometimes appear to be complaining.


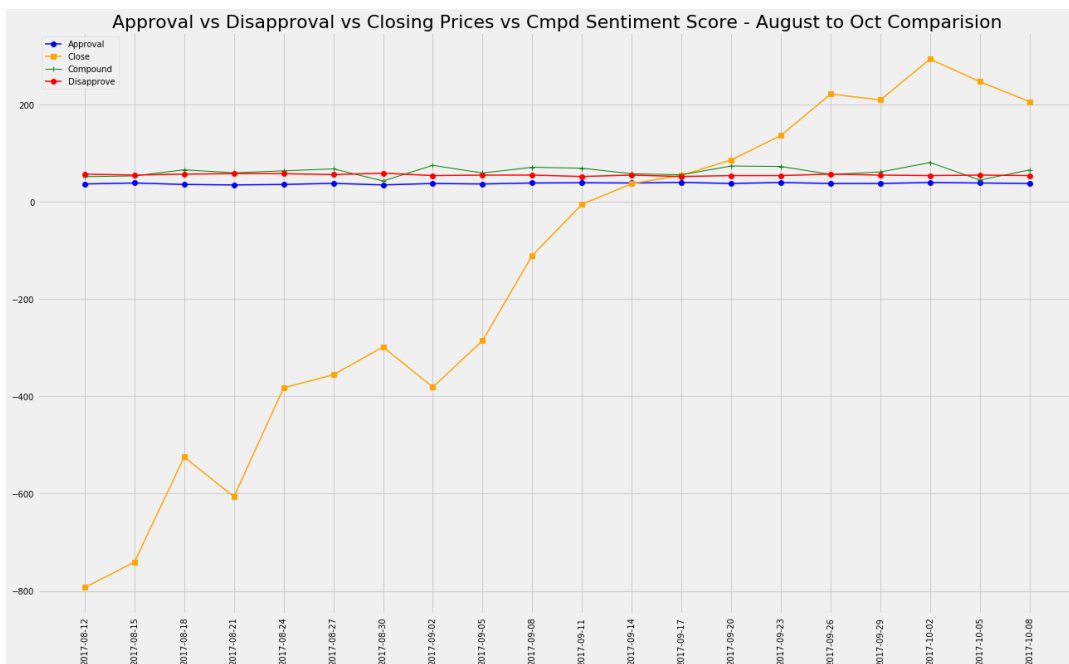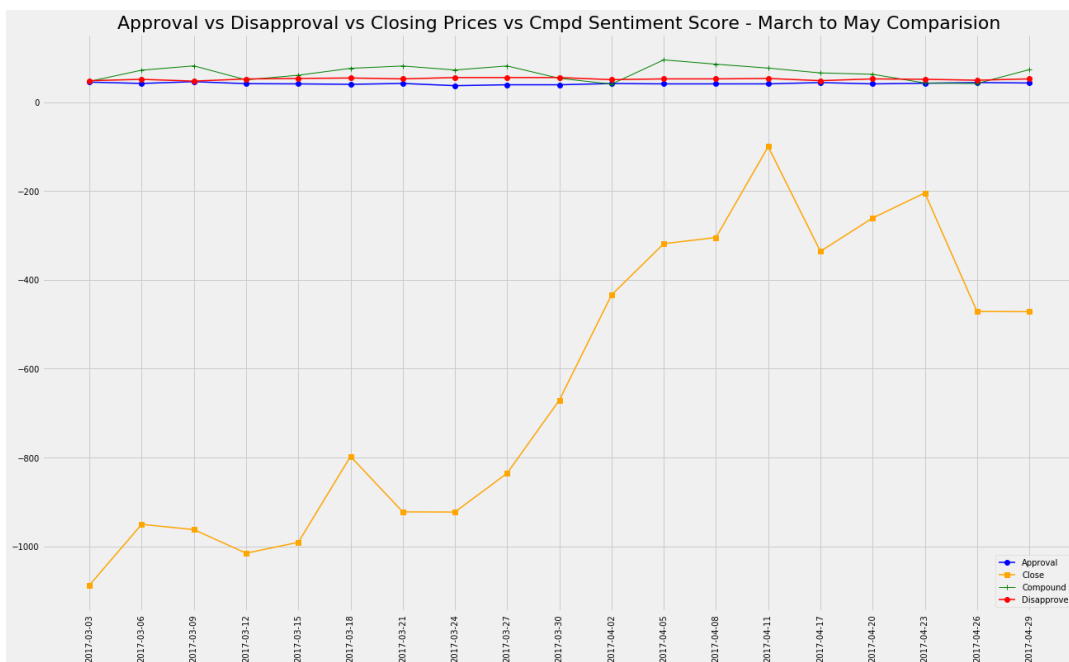WordClouds are a powerful visualization aid that can help with understanding your data.

6 Stock Market

We additionally analyzed President Trumps approval ratings and compounded sentiment rating against the 2017 S&P 500 average daily market data from YahooFinance.[11] The S&P500 is a stock market index that tracks the stocks of 500 large-cap US companies. It represents the stock market's performance by reporting the average market high, low, opening, and closing indices. Our analysis helped to understand the impact of Donald Trump's tweets to his approval rating and the S&P 500.
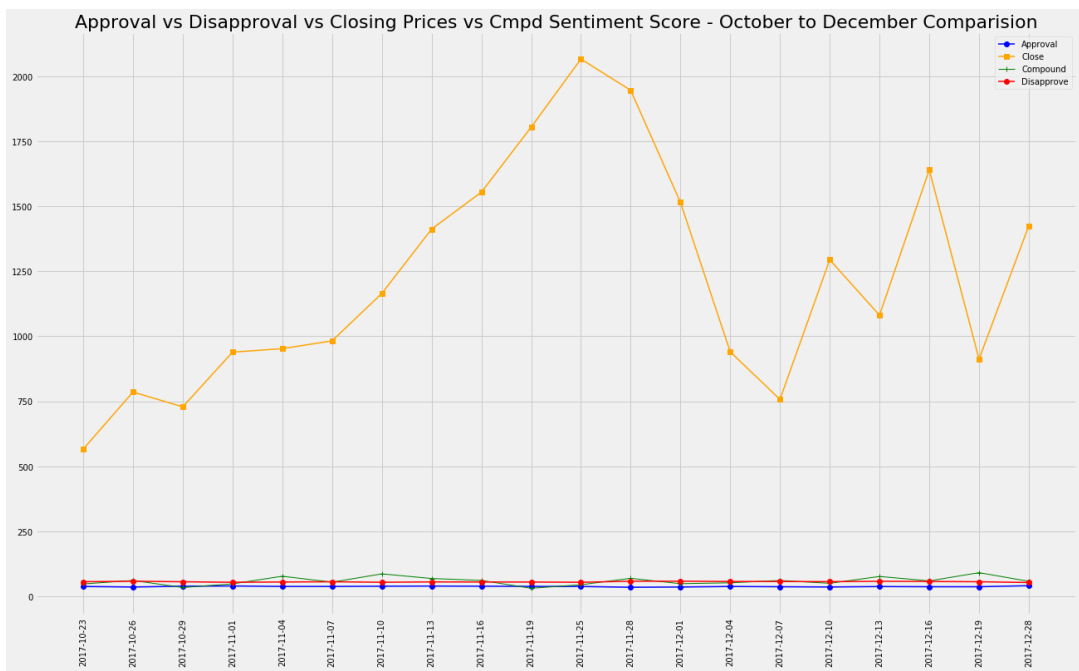
Therefore, after reviewing the 2017 S&P 500 market trend (Figure 20), we saw that 2017 was a overall robust year for industry. In general, the market saw an overall net gain and had few extreme volatile instances.
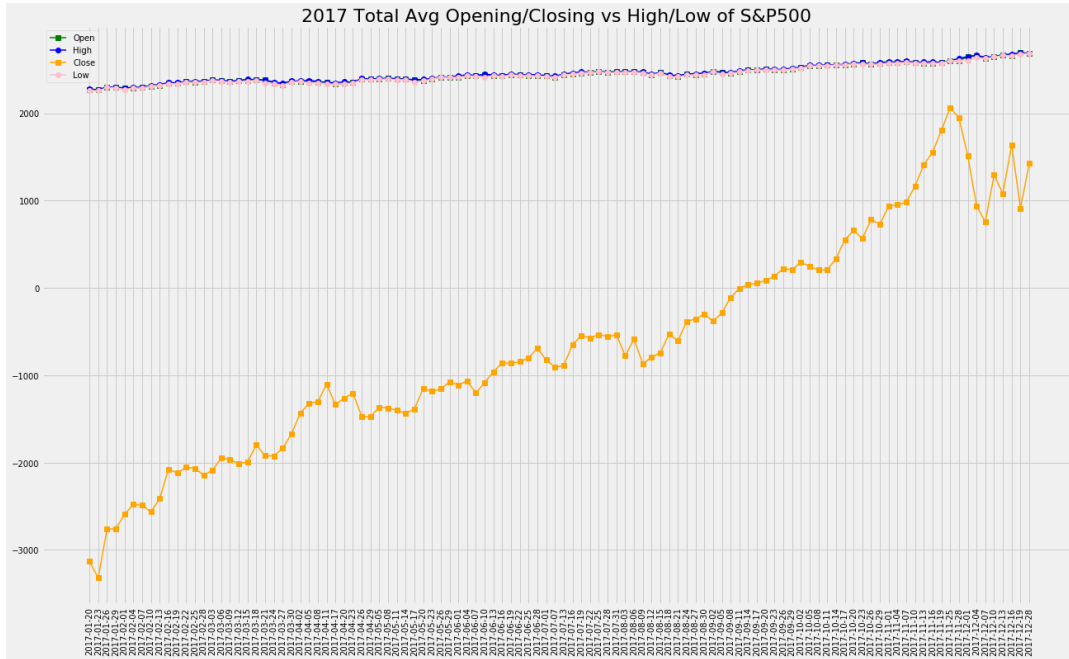
Next, we decided to review the President's approval ratings/sentiment against the stock market by 2-month increments. Figure 21 shows the market trend for March to May, or the spring/summer. Overall, we noticed that despite strong growth in the overall market, President Trumps approval rating did not change much. Additionally, we did see stronger sentiment in his tweets after a large market increase. In Figure 22, we reviewed the markets from August to October, here we initially saw a slight decrease in the markets, which was then followed by a sudden spike. Interestingly, just like in Figure 21, where we saw an increase in sentiment from President Trump after a market increase, we again saw this trend. Additionally, we also saw that despite the change in the market, President Trumps approval rating did not change. Finally, Figure 23, showed the strongest market variability for 2017. However, despite the markets net gain or loss, we saw that President Trump's tweets during this time did not have a high sentiment, nor were his approval ratings affected by the market.

Approval vs Disapproval vs Closing Prices vs Cmpd Sentiment Score - March to May Comparision

Approval vs Disapproval vs Closing Prices vs Cmpd Sentiment Score - August to Oct Comparision

Approval vs Disapproval vs Closing Prices vs Cmpd Sentiment Score - October to December Comparision

7 Recommendation / Conclusion

We analyzed 3 policies to see if Trump's tweets changed the behavior of staff, administrators and policy managers. It is clear that just like President Kennedy changed the communications landscape in the early 60's by using television Trump is advancing his agenda by using Twitter. But like all changes policy around presidential communications must change. Social media must now be included. Leta Lohrmeyer wrote in a paper at the University of Nebraska, "Results support the idea that the definition of public forum needs to be expanded and updated to include social media"[2]

It is clear from the results that this president has redefined the communications methods for the oval office and "Some would say that his use of this platform has reshaped the presidency." (Shear et al.2019)[1]

As politics moves into the social sphere the ability to touch voters directly is producing results that have never been seen before. It is clear that the politicians will have to agree on what the rules are and create a level playing feel for all potential candidates.

Tweeting at critical times does impact policy and forces administrators to respond to the "tweet crisis" as demonstrated in 3 specific cases (Border Security, Impeachment, AmericaFirst-TPP). The Cosine model was used to evaluate the consistency in the language of the tweets and was significant in setting the explicit tone for policy and resulted in direct action within President Trump's Administration, largely based on the pressure of supporters.

Consistent polling data shows after 2017 shows how either the sentiment analysis score or the language has an effect on the opinions of the public. Instead, the polls are clearly divided by party lines.

Presidential re-electability (PRE) based on tweets needs a different data set than is currently captured. We defined PRE based on polling data and see the trends but to predict PRE need to map the policies that are being effected in order to predict PRE. We would need additional polls on economy, tax reform and other polls to truly determine PRE.

We believe that that his tweeting related to the market also does not largely effect his approval or disapproval rating.

References

[1] Rich Harris; Blacki Migliozzi; Matthew Rosenberg; Rachel Shorey, How Trump Reshaped the Presidency in Over 11,000, Tweets. NewYorkTimes, New York, 2 Nov. 2019

[2] Leta Lohrmeyer, Executive Authority and Free Speech:An Analysis on the Restraints of Presidential Power, University of Nebraska, 2019

[3] Rachel Schutt ; Cathy O'Neil, Doing Data Science, O'Reilly Media, Inc., 2014

[4] Miller, Thomas W. , Modeling Techniques in Predictive Analytics with Python and R: a Guide to Data Science., Pearson Education, 2015

[5] Brown, Brendan., Trump Twitter Archive, http://www.trumptwitterarchive.com/, 2020

[6] fivethirtyeight, "https://projects.fivethirtyeight.com/trump-approval-data/approval_polllist.csv", projects.fivethirtyeight.comtrump-approval-ratings, 2020

[7] Graham, D. A., www.theatlantic.com, November 2019

[8] IEE Explore, ieeexplore.ieee.org, https://ieeexplore.ieee.org/document/8748599, 2018

[9] MonkeyLearn, sentiment-analysis-of-twitter, https://monkeylearn.com/blog/sentiment-analysis-of-twitter/' 2020

[10] Kaggle, https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs, 2020

[11] yahoo finance, https://finance.yahoo.com/quote/