Josh Edelmann

COMP 4447 – Data Science Tools I

Professor Daniel Parada

21 August 2020

Twitter Analysis: A Sentiment Analysis of President Trump, Vice President Biden, and Public Tweets

Section I: Research Question

How does the sentiment from tweets differ between President Trump, Vice President Biden, and general Twitter users?

Section II: Introduction

Twitter has become a platform for individuals to express opinion, share their ideas, or communicate with strangers about certain topics. Twitter is also a place where people get their news, watch sports highlights, or share political opinions; generally, it is a place with varied information and data. Twitter contains many factual and fallacious tweets along with positive and negative sentiment tweets. A challenge is how to make sense of all this data.

In just 3 months, voters will head to the polls to cast a vote for the next President of the United States. With all the noise on Twitter, there is an abundant amount of data out there. In this paper, I attempt to make sense of the data. Using python and the library "Tweepy," I use Twitter API to scrape tweets from various users. Given there is a Presidential election this year I also collect tweets from President Trump, Vice President Biden, and general Twitter users and will build a model to classify their tweets into positive, negative, and neutral sentiment. I will also build a model to classify their tweets based on fact or opinion. First, the tweets will be cleaned by removing punctuation, symbols, hashtags, the hyperlink, and words that do not serve a purpose in the analysis portion. Next, I put the cleaned tweets into a data frame and create new columns containing the sentiment analysis, which will be "polarity" and "subjectivity." From

there, I produce a word cloud for all three of the users I'm scraping tweets from. I then produce charts displaying the sentiment analysis summary from the users and will compare and contrast the final results. The graphics will display the subjectivity and polarity analysis, along with the aggregate of each.

Section III: Literature Review

For the President of the United States, Twitter is a place to abruptly change policy, attack people that share opposing ideas, or praise himself for a job well done. In the New York Times article, *"How Trump Reshaped the Presidency in Over 11,000 Tweets"*, the authors argue that in more than half of his tweets Mr. Trump attacks his enemies; and has praised himself in over 2,000 of his tweets. Mr. Trump's targets of his attack posts include the Russia investigation, entire cities, previous administrations, and the list goes on[1].

Recently, Twitter began restricting false and misleading tweets from appearing on user's feed. Twitter froze a tweet tweeted from the Trump campaign until the tweet was removed. A spokesperson for Twitter stated that the Trump account was "in violation of the Twitter Rules on COVID-19 misinformation" and "The account owner will be required to remove the Tweet before they can Tweet again." Censoring misinformation on Twitter, and other social media platforms, is the result of the spread of misinformation that can lead to deaths, or foreign countries interfering in U.S. elections[2].

This paper expands on these articles by comparing tweets from multiple sources. While the New York Times article analyzed solely Trump tweets, and BBC article looked at Twitter as

[1] Shear, M., et al. (2019, November 02). How Trump Reshaped the Presidency in Over 11,000 Tweets. Retrieved August 21, 2020, from https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter-presidency.html

[2] Facebook and Twitter restrict Trump accounts over 'harmful' virus claim. (2020, August 06). Retrieved August 21, 2020, from https://www.bbc.com/news/election-us-2020-53673797

a whole, this paper will compare Trump, Biden, and general Twitter users. During the comparison, we will visualize the words that appeared most in the tweets along with the sentiment analysis, which contains subjectivity and polarity ratings. We will start by pulling in tweets and cleaning the data.

Section IV: The Data

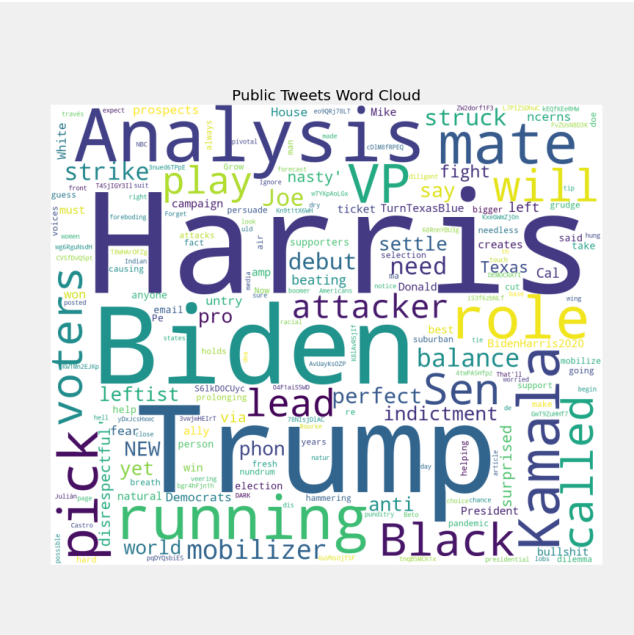
To get tweets from Twitter, I set up a Twitter API with proper credentials. This will allow me to collect a total of 3000 tweets to analyze. When importing public tweets, I set the search terms to 'politics', 'election', '2020', 'biden', 'trump', 'harris', and 'vp.' I set the language to English to collect tweets only in English. Then I import the lastest Trump and Biden tweets. For the process to work, you first have to set up Twitter developer credentials to receive keys and tokens. Once that process is complete, it is then possible to pull tweets using proper syntax.

Section V: Data Clean Up

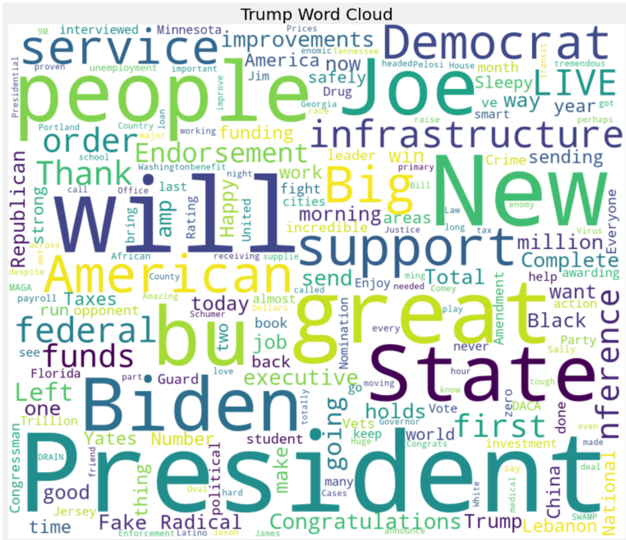
Here, I wrote various functions to clean the tweet data. I removed "@mentions", "#", "RT", the hyperlink, and other characters that will not be useful when doing the sentiment analysis. I also removed stop words, such as "a", "the", "is", "at", and "on" that do not provide any additional insight. Next, I created three functions to get the subjectivity, polarity, and analysis from each tweet. Subjectivity is on a scale from 0 to 1. A rating of 0 would indicate a fact and a rating of 1 would indicate an opinion. Python subjectivity feature rates words based on how subjective each word is and then computes a final score for the text. Polarity is on a scale from -1 to 1. A rating of -1 would indicate strong negative sentiment, a rating of 0 is neutral, and a rating of 1 is a strong positive sentiment. The polarity feature will give a rating to each word based on how negative or positive it is and provides a final score for each tweet.

Section VI: Analysis

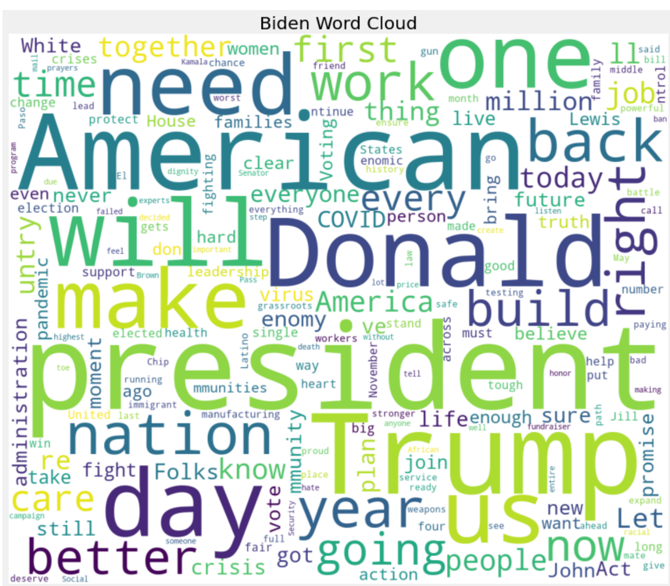
First, let's discuss the word cloud results for public, Trump, and Biden tweets.



Above, are the tweets from general Twitter users using the keywords described previously. I pulled these tweets the day Senator Kamala Harris was announced as Joe Biden's running mate. We can see that *Harris, Biden,* and *Trump* are the most tweeted words by Twitter users. Looking at the other words in the cloud, we see *voters, Black, mate,* and *pick* have also been tweeted some. This may suggest people are tweeting about how voters feel about Harris as Vice President. Also, how Black people feel about the election or Harris.

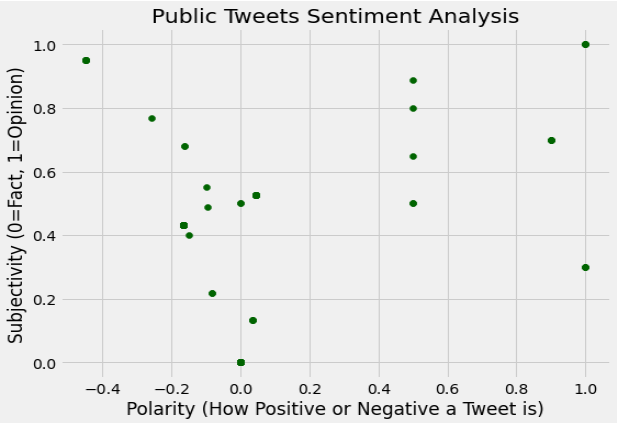


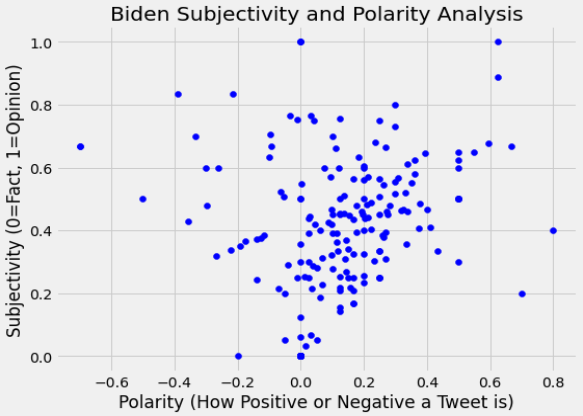
Above, is the word cloud generated from Trump tweets. We can see that *President, Biden, state, new*, and *great* are the most tweeted words by Trump. It appears Trump is tweeting about Biden and the presidency. Looking at the smaller words in the cloud, we see *Fake, radical,* and *taxes* were also tweeted about. Trump might be referring to "fake" news as he often does, and the radical democrats and how they might raise taxes.

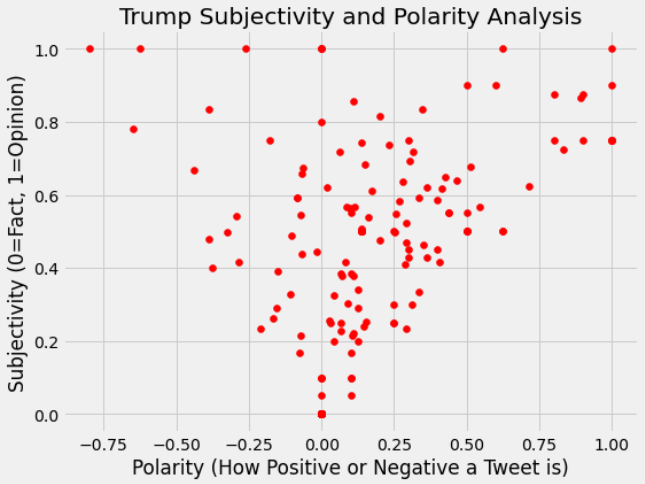


Finally, we have the word cloud generated from Biden tweets. We can see that Biden is tweeting about *American, Donald, president, Trump,* and *day* the most. The smaller words in the cloud such as *covid, crisis,* and *virus* show that the coronavirus is a topic he has been tweeting about.

Now, we will compare scatterplot results for public, Trump, and Biden tweets.



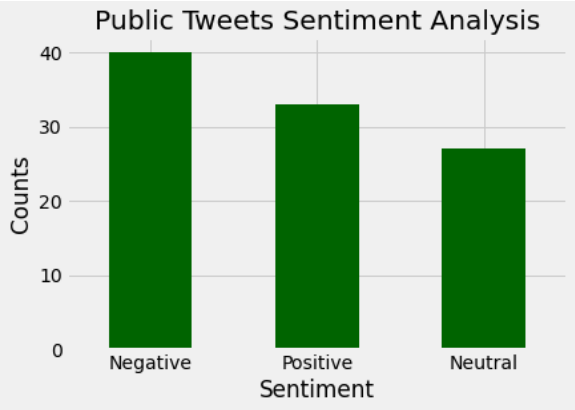


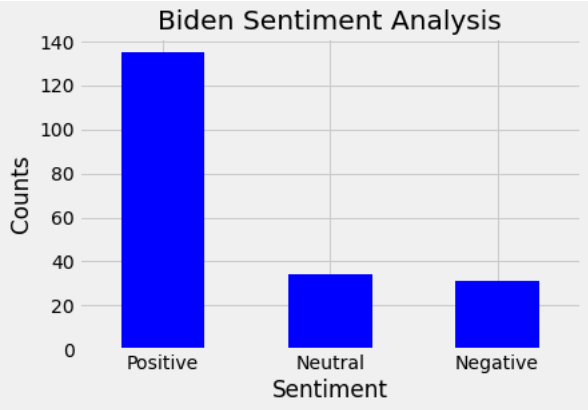


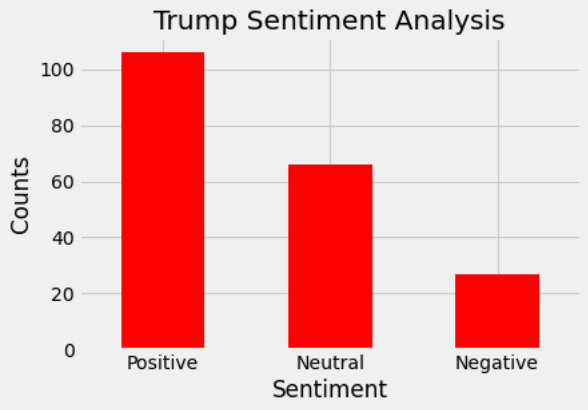
Above, we have the scatterplot results of polarity and subjectivity scores for each user. I want to point out that there appears to be few tweets for the public tweet's scatterplot. I believe this may have to do with Twitters rules with importing tweets. For the tweets we do have, we can see that majority of public tweets fall around neutral to negative rating. Trump's tweets are more

scattered than Biden's. Biden's tweets are clustered between 0.0 and 0.4 on the polarity axis and between 0.2 and 0.6 on the subjectivity axis. This means majority of Biden's tweets that are neutral to positive tend to be more factual. While Trump's tweets tend to more neutral when fact based but then fluctuate between positive and negative as the tweets become more opinionated.

Next, let's compare the bar plot results of the polarity sentiment analysis.

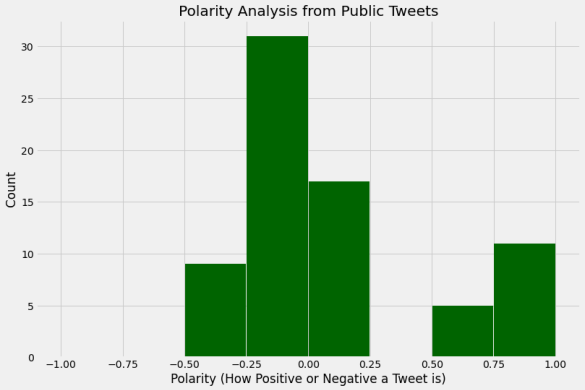


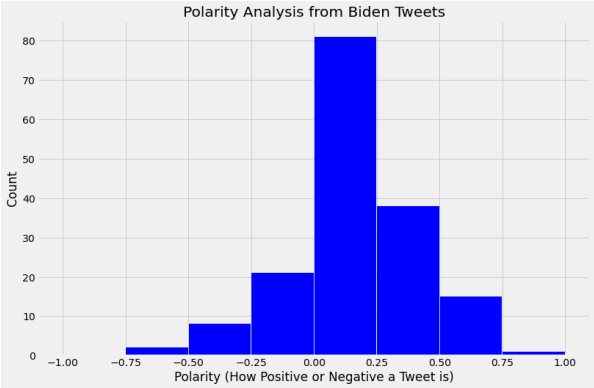


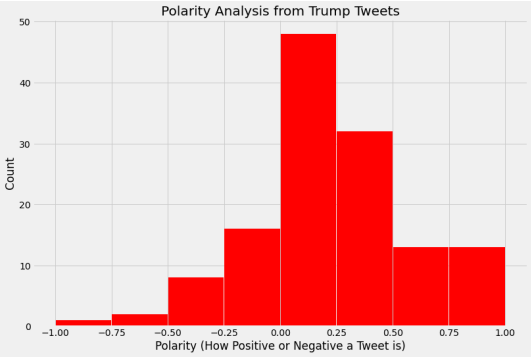


We can see that the majority of public tweets are negative while majority of Biden and Trump's tweets are positive. This may indicate that twitter users hold a negative view of politics or the candidates running for President. Biden has more positive tweets than Trump, but not by much. This could be an indication that Biden and Trump are appealing to voters by using positive statements.

Now, we will compare the distribution of the polarity score for each user.

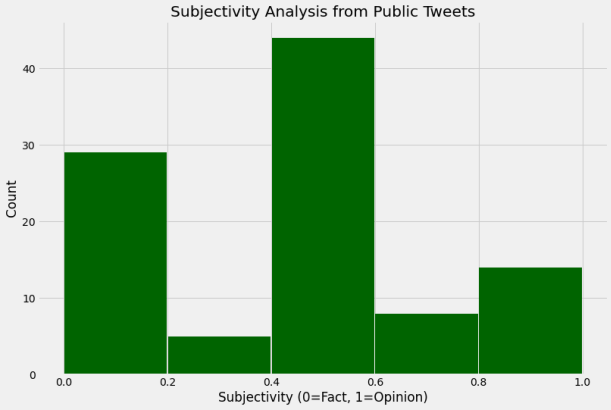


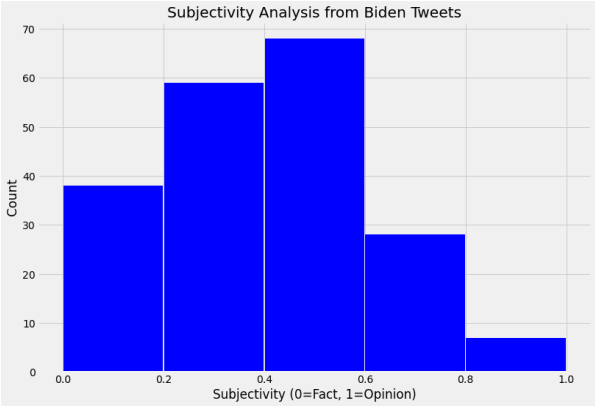


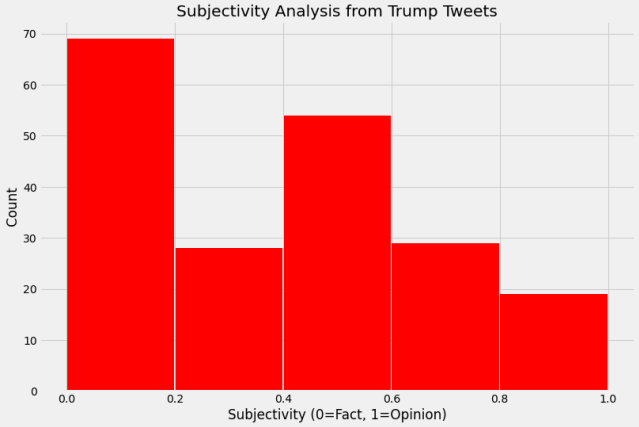


Above, we see the distribution of the polarity score for each user. I excluded any rating of 0.00 because I am more interested in positive and negative tweets than neutral tweets. This confirms the previous graph. Public tweets are more negative than Biden or Trump tweets. It appears that Biden and Trump are evenly split between negative and positive tweets. My assessment still stands that public users may hold a negative view towards politics.

Finally, let's compare the distribution of the subjectivity score for each user.







Pubic tweets are evenly dispersed. Majority of tweets are around 0.5, but there is also a spike on the wings. Majority of Trump and Biden tweets are at or below 0.5, so majority are rated as truthful or somewhat truthful.

Section VII: Conclusion

While this analysis is not complete, it does give us a good overview of how twitter users are feeling about politics. This also shows how Trump and Biden tweets compare to each other and to public tweets. Based on this analysis, public tweets are more negative than Trump or Biden while Biden has more positive tweets than Trump, although not by much. All the three users are split on subjectivity. There is not a clear pattern for public tweets, but for Trump and Biden their tweets are more factual than opinion based. There were no outliers or missing values, so this was not an issue. There were few values for public tweets, which I contribute to Twitter's policy concerning API and importing tweets. For future research, pulling tweets and classifying the user as republican and democrat based on each user's likes and follows would be an interesting topic. A good resource for that would be New York University Center for Social Media and Politics Lab[3].

[3] https://csmapnyu.org/