**Analysis of Joe Biden and Donald Trump’s Tweets**

Elizabeth Chambers

**Introduction**

The 2020 Presidential election led to much turmoil and unrest throughout the United States, is this entropy due to particularly garish candidates, or are politicians simply an unsavory genotype? In this project, I will utilize sentiment analysis to gain insight on how negative, positive, or neutral the two most prominent 2020 Presidential candidates come across on the popular social networking website known as *Twitter*.

The strut and demeanor of many politicians begs the question: do politicians possess a particularly sophisticated vocabulary? This paper will also explore the most commonly tweeted words and collocations of the United States’ top two 2020 Presidential candidates.

**Methodology**

**Data**

The data utilized was obtained through Kaggle’s open database, and contain tweets from both Joe Biden and Donald Trump for the years 2007-2020 and 2009-2020, respectively. Both datasets were originally obtained through use of Twitter’s API service. The two datasets consist of the variables: tweet content, timestamp, tweet id, tweet url, number of retweets and number of favorites.

**Procedure**

In order to analyze the most prominent sentiments of Joe Biden and Donald Trump on *Twitter*, datasets containing both candidates’ tweets from the years 2009-2020 were loaded into the python environment. Next, a function which takes in a tweet and returns its sentiment, based on *TextBlob’s* analysis of sentiment polarity – a “positive” sentiment having > 0 polarity, a “neutral” sentiment having 0 polarity and a “negative” sentiment having < 0 polarity. (See Appendix A). Utilizing the aforementioned function, a column was added to each candidate’s respective dataframe, detailing what “sentiment” is associated with each tweet. Count plots for each user’s sentiments were then plotted by use of the *Seaborn* package:

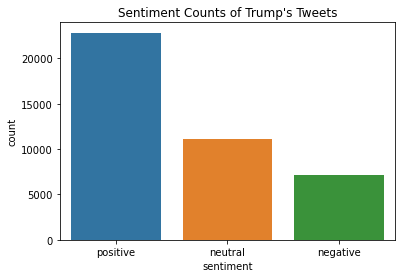


Figure 1

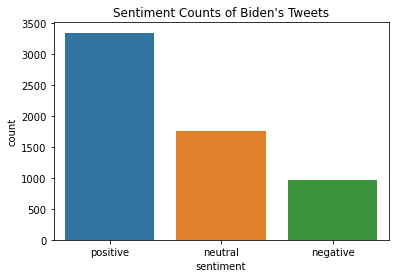
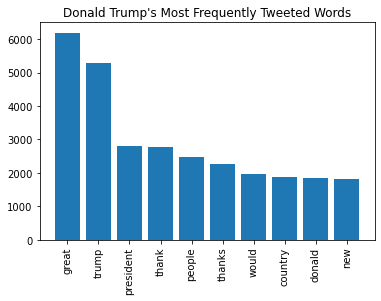


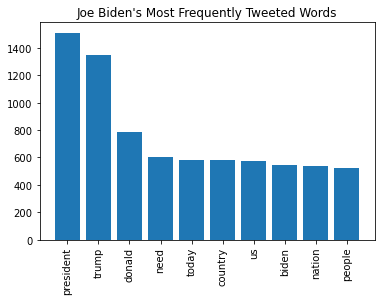
Figure 2

Next, the proportions of each sentiment for both users were calculated by utilizing *NumPy* to calculate the sum of each sentiment, and dividing the sum by the total number of observations in each dataset (found using the *shape()* function provided by the *Pandas* package).

In order to analyze which words and expressions each user implements most frequently, functions were written to remove punctuation as well as html’s from the tweets (See appendices B and C). Next, each tweet in its respective dataframe was tokenized, lowercased, freed of “stop words” and stored in a list, which was then utilized to find the most frequently occurring words used by each user by implementing the *Counter()* function provided by *collections.* (See Appendix D). *MatPlotLib* was then utilized to create bar graphs which depict each user’s top ten most commonly used words:



Figure



Figure

Lastly, to perform collocation and identify words that commonly co-occur in each user’s tweets, the *BigramCollocationFinder* from *nltk.metrics* was used on each candidate’s list of words tweeted. This generated some entertaining common bigrams.

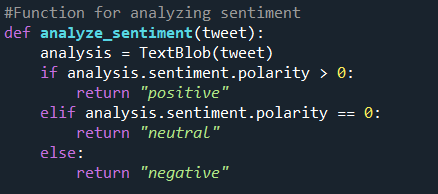
**Results**

When comparing figures 1 and 2, while it is clear which user has composed more tweets from 2009-2020 (Trump tweeted greater than 40,000 times as opposed to Biden’s approximately 6,000 tweets), it is more difficult to discern which user has more tweets with a “positive”, “negative” or “neutral” sentiment. Looking at the graphs, both users appear to mostly have tweets with positive connotations. However, when comparing the proportions of positive tweets to total tweets of each user, Joe Biden has a slightly higher ratio (0.55097) as opposed to Trump’s 0.55007. Inversely, Donald Trump has a slightly higher proportion of negative tweets (0.17497), while Joe Biden’s negative proportion comes in at 0.15952. Keeping with a very close tweet-proportion-race, the candidates come in with neutral proportions of 0.27003 (Trump) and 0.28951 (Biden).

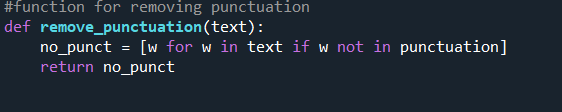
**Conclusion**

**Appendices**

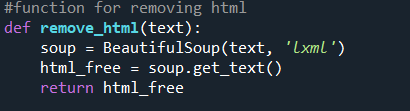
**Appendix A:** Python function for analyzing the sentiment of a tweet based on its polarity determined by TextBlob’s sentiment analysis function.



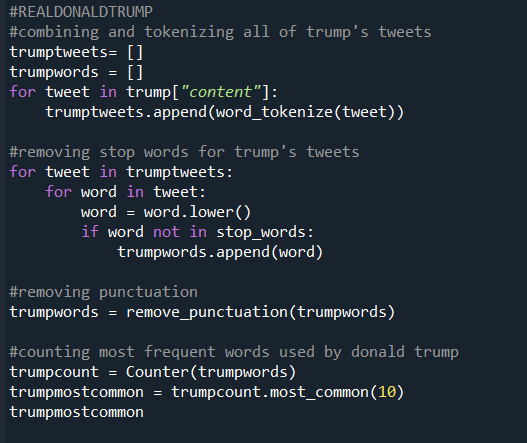
**Appendix B:** Python function for removing punctuation.



**Appendix C:** Python function for removing html’s.



**Appendix D:** Python code for cleaning text and counting most frequently used words (this was conducted for both Donald Trump and Joe Biden).



**Citations**

Reese, A. (2020, June). Trump Tweets. https://www.kaggle.com/austinreese/trump-tweets.

Vopani. (2020, October 31). Joe Biden Tweets (2007-2010). https://www.kaggle.com/rohanrao/joe-biden-tweets.