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

Elizabeth Chambers

**Introduction**

**Problem**

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 (Donald Trump and Joe Biden) 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.

**Data**

The data utilized was obtained through Kaggle’s open database, and contains 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.

**Methodology**

**Tools**

To carry out this project, a number of Python packages were imported and utilized, such as: *numpy* (to perform mathematical operations on the dataframes), *pandas* (for manipulating and analyzing data with ease), *matplotlib* and *seaborn* (to create straight-forward and useful data visualizations), *TextBlob* (to analyze the sentiment polarity of each tweet), *collections* (to perform word counts) and *nltk* (to more easily perform natural language processing by “cleaning” each candidates tweets through tokenization and the removal of stop-words, and to find common bigrams within the cleaned tweets).

**Methods**

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:

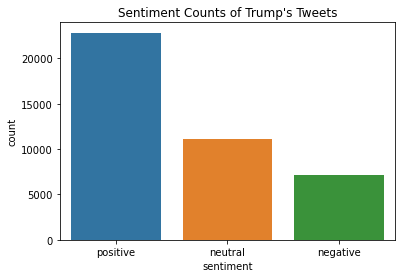


Figure 1

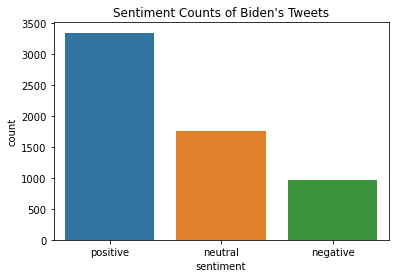


Figure 2

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; a function was written to remove punctuation from the tweets (See appendix B). Next, each tweet in its respective dataframe was tokenized, lowercased, freed of English “stop words” (a collection of generic words such as “the” and “it”, provided by the *nltk.corpus* library) and stored in a list, which was then utilized to find the most frequently occurring words of each user by implementing the *Counter()* function provided by *collections* (See Appendix C). The *matplotlib* package was then utilized to create bar graphs which depict each user’s top ten most commonly used words:

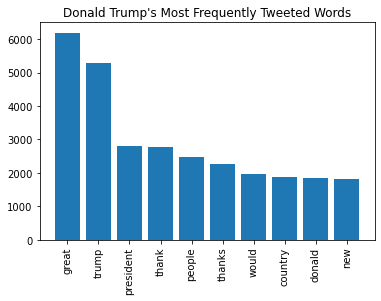


Figure 3

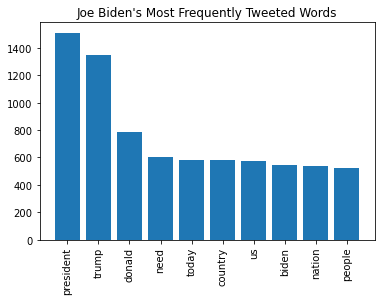


Figure 4

In order to further analyze the implementations of these commonly used words, particularly intriguing words such as “thank” and “need” were further investigated by creating lists of tweets which contained such words and inspecting the tweet content of those lists (See appendices D and E).

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:

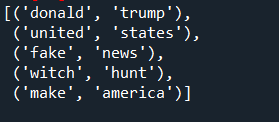


Figure 5, Common bigrams used by Donald Trump

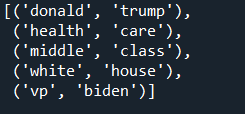


Figure 6, Common bigrams used by Joe Biden

**Results**

**Sentiment Analysis**

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).

**Commonly Used Words**

Figures 3 and 4 reveal some interesting insight into each user’s most commonly used words. Mr. Donald Trump favors the word “great” – which he appears to have tweeted over 6,000 times within the last decade. Mr. Joe Biden’s most favored word of the past ten years speaks to his motives for using the twitter platform – he has tweeted the word “president” about 1,500 times. Both users second most commonly tweeted word is “trump”, and the noun “Donald” also makes an appearance in each user’s top ten words, this speaks to one candidate’s unhealthy narcissism and the other’s tendency to criticize his opponent’s egocentric ways.

It was refreshing to detect words such as “country” and “people” in both user’s lists, as this implies the United States’ presidential candidates do in fact put some thought into the land they will be leading. Biden even went so far as to refer to the United States’ people as “us” and “nation” many times within the past decade, which refreshingly implies a sense of unity and acceptance the US has been missing for about four years. It was a relief to see the former president does indeed have manners and commonly types the words “thank” and “thanks”. Further investigation into Trump’s utilization of the word “thank” reveals no sardonic or satiric connotations.

“New” was another shocking word on Donald Trump’s list, as his entire campaign platform revolves around the past – the phrases “*make America great* ***again”***and *“****keep*** *America great”* have no foot in the future and imply a complete disregard for any *new* ways of thought. In keeping with the future tense, Biden’s tendency to tweet the word “need” implies that he is aware of, and has a plan for the country’s future and its needs. Upon further inspection of Biden’s use of the word “need”, no insight as to his plans for adhering to the country’s needs was revealed. However, his tweets do detail how much the country’s people *need* to vote or what its people *need* to do for their government.

**Collocations**

The collocation finder yields phrases one might expect from the United States’ former president, what with his tendency to play the victim and refer to the investigations into his policies and integrity as a “witch hunt” or his impulse to refer to any sort of allegation against him as “fake news”. At least he has remained consistent in his obsession with himself and his mission to “make America” great, again.

Biden’s bigrams are only slightly revitalizing, at least he has a tendency to speak about the “middle class” and “health care” (two things which have surprisingly not been eradicated over the past four years). Again, the new president appears to have an unhealthy obsession with his opponent “Donald trump”, who again makes a guest appearance in his most common list.

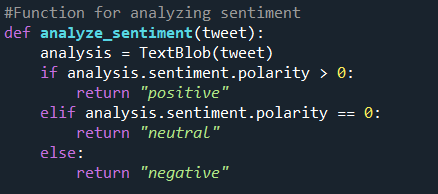
**Conclusion**

The sentiments of the 2020 Presidential Candidates’ tweets have a slight tendency to be positive - though the most commonly used words reveal no particular sophistication or intellect and narcissistic tendencies on both sides. It also appears that politicians often speak of their nation and their plans for its future (a rather comforting discovery). It would be interesting to perform a similar analysis on other texts and count the number of positive and negative connotations of each politician’s most commonly used words and phrases, or delve even further and perform an analysis to determine which political party (or which country) has the most “positive” politicians.

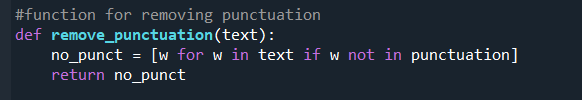
Though one might imagine that the United States’ elite are a faction of silver-tongued, pretentious beings who never say “thanks” and never associate with “us”, this research (whether it be “fake news” or not) demonstrates otherwise. This research reveals that at their core, politicians are people. They fumble over the nuances of the English language and implement entirely unsophisticated words more often than not. They strive to appear positive to the public, even when the world is in turmoil. They are, simply, average joes.

**Appendices**

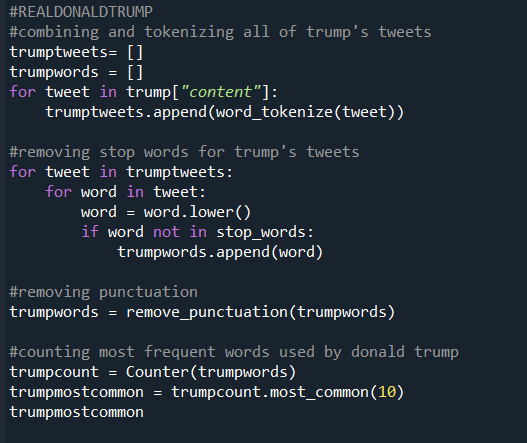
**Appendix A:** Python function for analyzing the sentiment of a tweet based on its polarity determined by TextBlob’s sentiment analysis function.



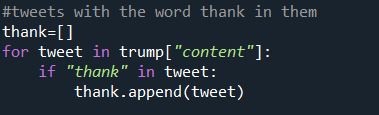
**Appendix B:** Python function for removing punctuation.



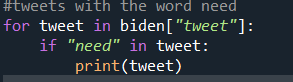
**Appendix C:** Python code for cleaning text and counting most frequently used words (this was conducted for both Donald Trump and Joe Biden).



**Appendix D:** Python code to inspect Trump’s tweets which contain the word “thank”.



**Appendix E:** Python code to inspect Biden’s tweets which contain the word “need”.



**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.