**Project Overview:**

This project gathered data from twitter. We gathered president Trump’s tweet, but the architecture of the program allows anybody’s tweet to be gathered as long as we know their username. The tweets are then analyzed using a second program, which does basic text processing and sentiment analysis using several modules.

**Implementation:**

There are two programs: twitterscraper.py and twitteranalyzer.py

Twitterscraper.py gathers data from twitter using the Tweepy module. All of the relevant functions are built into the TwitterClient class. There are two main functions, one for getting all of the data related to the tweet, and one for just the text of the tweet. The main reason for this was because the Twitter API limits the amount of tweets one is allowed to get, so in case we want to do further analysis, saving all of the data may come in handy in the future. Because the twitter API only allows 200 tweets to be retrieved at once, we create a loop to gather multiple batches of tweets. We use the id of the tweet to ensure that previously downloaded tweets are not downloaded again, as the Twitter API retrieves the newest tweets first.

Twitteranalyzer.py uses several functions we have used in our course as well as new ones. It uses nltk’s built in tokenize function as well as a simple regex function to parse the text, with both cleaning the text differently. The more interesting function performs a sentiment analysis using TextBlob, which was picked because of it’s easy to use syntax and formatting of information. The sentiment analysis function will return the tweet along with its designated sentiment (positive, negative, neutral) along with its numerical polarity and subjectivity score.

**Results:**

We can see that trump’s top 10 words with their counts are as follows:

789 great

765 will

663 s

447 amp

431 rt

398 t

392 people

327 trump

295 president

294 country

Assuming he only uses these words once per tweet, he really loves to use the word great. There are few words that seem to be there for no reason, such as ‘s’, ‘amp’, ‘rt’, and ‘t’, which I assume is the result of twitter specific text.

Since the word “great” appears so often it would be interesting to see what percentage of his tweets are positive versus negative.

Output:

Positive tweets percentage: 59.03538991544003 %

Negative tweets percentage: 22.29877857813968 %

Neutral tweets percentage: 18.665831506420293 %

We can see that according to the analysis, a majority of trump’s tweets are positive, and only 22.3% of his tweets are negative. Given our previous knowledge this is not too surprising. We should also see what tweets were classified as positive, negative or neutral:

Most recent positive tweets:

American Cable Association has big problems with Comcast. They say that Comcast routinely violates Antitrust Laws. “These guys are acting much worse, and have much more potential

for damage to consumers, than anything AT&amp;T-Time Warner would do.” Charlie Gasparino

The California Fire Fighters, FEMA and First Responders are amazing and very brave. Thank you and God Bless you all!

The Florida Election should be called in favor of Rick Scott and Ron DeSantis in that large numbers of new ballots showed up out of nowhere, and many ballots are missing or forged. An honest vote count is no longer possible-ballots massively infected. Must go with Election Night!

.....of money spent on protecting other countries, and we get nothing but Trade Deficits and Losses. It is time that these very rich countries either pay the United States for its great military protection, or protect themselves...and Trade must be made FREE and FAIR!

.....hundreds of billions of dollars, for the great privilege of losing hundreds of billions of dollars with these same countries on trade. I told them that this situation cannot

continue - It is, and always has been, ridiculously unfair to the United States. Massive amounts.....

Most recent negative tweets:

Hopefully, Saudi Arabia and OPEC will not be cutting oil production. Oil prices should be much lower based on supply!

I am in Paris getting ready to celebrate the end of World War One. Is there anything better to celebrate than the end of a war, in particular that one, which was one of the bloodiest and worst of all time?

There is no reason for these massive, deadly and costly forest fires in California except that forest management is so poor. Billions of dollars are given each year, with so many

lives lost, all because of gross mismanagement of the forests. Remedy now, or no more Fed payments!

Jeff Flake(y) doesn’t want to protect the Non-Senate confirmed Special Counsel, he wants to protect his future after being unelectable in Arizona for the “crime” of doing a terrible job! A weak and ineffective guy!

You mean they are just now finding votes in Florida and Georgia – but the Election was on Tuesday? Let’s blame the Russians and demand an immediate apology from President Putin!

Most recent neutral tweets:

The prospect of Presidential Harassment by the Dems is causing the Stock Market big headaches!

RT @WhiteHouse: President Trump Attends the American Commemoration Ceremony at Suresnes American Cemetery https://t.co/P6PgE1zXuE

Trying to STEAL two big elections in Florida! We are watching closely!

Thank you @marcorubio for helping to expose the potential corruption going on with respect to Election Theft in Broward and Palm Beach Counties. The WORLD is now watching closely!

Looking at these tweets, the basic sentiment analysis using TextBlob is not the best, perhaps because the polarity score gives an sum of the positive and negative values of each word rather than a separate total positive or total negative score. It also shows the limitation of sentiment analysis, at least using one straight out of the box, as certain tweets appear negative but are actually positive in intention, and vice versa.

**Reflection:**

We were able to get all the texts and perform simple text analysis on the tweets. However, we used a generalized library of stop words, sentiments, and text processing. Twitter has several peculiarities, such as its use of hashtags, @’s, and smiley faces that should be worth including in the analysis. Furthermore, it may have been interesting to compare trumps tweets to another politicians tweets, like Barack Obama’s. Since TextBlob also gives a number score for polarity and subjectivity, it would be have been interesting to see what is Trumps most positive and negative tweet, and his most subjective and objective tweet, at least according to the sentiment analysis. If I had time I also would have liked to use Markov Chains to generate fake Trump tweets and compare them to real Trump tweets. As what I did for this mini-project is relevant to my main project goal, I will be able to utilize what I learned there. I do wish I knew more about twitter, as I have never used twitter in my life, so I did not know what to do with twitter specific information, such as the use of hashtags and retweets. Frankly I also wish I had more time to work on this as it’s been a busy week and thus I wasn’t able to explore the options I wanted to explore, such as visualizing the data. I can actually do it pretty fast in R now but I haven’t done any visualization of data in Python so this would have been a good opportunity to learn had I more time.