README

The ultimate goal of these files is to store and analyze Twitter data. These various files, some obvious by their names, were used to create database tables, query Twitter and store the processed data, and perform the end analysis. All files were programmed in Python v. 3.6.

The final output is listed at the end of this document, though to alter files or rerun the program with updated Twitter data, first update the db\_utils.py file to create a local SQLite database and update the path to this file. A file named config.py must be added into the same directory as these files with your personal Twitter API keys. You can then run the below files as such-

**Create\_tables\_db.py**

* Used to create initial SQLite tables- one which will store all Trump tweets, one to store a sample of random tweets, and one to store a sample of tweets with mention @realDonaldTrump

**Create\_response\_tweet\_tables.py**

* Used to create additional tables which will store tweets directed at Trump. One table will hold general tweets directed at Trump, one will hold tweets directed at Trump including keywords from the list of top keywords from positive Trump tweets, and the third will hold tweets directed at Trump including top keywords from negative Trump tweets

**Load\_general\_tweets.py**

* Queries tweets (no specific query- only queried keyword is “a” since a term must be entered). Processes the responses and stores them in SQLite table

**Load\_response\_tables.py**

* Queries the following kinds of tweets, processes the data, and stores the data in their corresponding tables
  + Tweets that are directed at Trump with no keywords
  + Tweets that are directed at Trump including keywords from the list of top keywords from positive Trump tweets
  + tweets directed at Trump including top keywords from negative Trump tweets

**Load\_trump\_tweets.py**

* Accesses following data files from Kaggle which contain all Trump tweets from the past two years
  + Condensed\_2017.json
  + Condensed\_2018.json
* Processes the data and stores it in SQLite table

**Identify\_keywords.py**

* Queries table of trump tweets to identify top keywords in his positive, negative, and neutral tweets. Writes these keywords to text files in the same directory
* Removes stop words (take from MySQL) and counts the most mentioned words under each sentiment

**Find\_average\_followers.py**

* Creates a SQLite table to store followers per random Twitter user, queries Twitter for random tweets, stores the followers per user of each tweet, then outputs the average number of followers for this sample

**Tweet\_analysis.py**

* Queries the SQLite tables and outputs all final statistics

**Final Output**

Percent of positive Trump tweets: 55.6 %

Percent of negative Trump tweets: 22.5 %

Percent of neutral Trump tweets: 21.9 %

Percent of positive public tweets: 43.6 %

Percent of negative public tweets: 21.8 %

Percent of neutral public tweets: 34.6 %

Average polarity of Trump tweets: 0.14

Average polarity of general public tweets: 0.09

Number of Trump tweets analyzed: 3254

Number of general public tweets analyzed: 2004

Average polarity of tweets mentioning Trump: 0.09

Average polarity of tweets mentioning Trump with positive keywords: 0.18

Average polarity of tweets mentioning Trump with negative keywords: 0.05

Sample size of tweets mentioning Trump: 2209

Sample size of tweets mentioning Trump with positive keywords: 6050

Sample size of tweets mentioning Trump with negative keywords: 4040