Final Project Readme Template – Ibrahim A. Rasheed

1. Type of project:
An API mashup project

2. Brief description of your project:
This program takes in an input of three movies from the user. It searches each movie through Twitter by adding a hashtag through it, and collects 100 tweets with that hashtag. It then removes redundant "Retweets"/RT tweets. The remaining tweets are then split into individual words and a sentiment score is ran through the words. The positive and negative scores are collected and then divided by each other to receive a ratio. The program then accesses OMDb (The Open Movie Database) for the same three movies, and collects the IMDb critic score and Metacritic score of each movie. The results are displayed, and the user is then able to see whether a movie's sentiment score on Twitter is correlational to its official IMDb and Metacritic score, and by comparing one movie to two others to establish a baseline and relativity.

3. Files being turned in:
Final.py is the main file with the code.
creds.txt is a text file with the Twitter API credentials.
positive-words.txt is a text file with positive words to compare to in sentiment scores.
negative-words.txt is a text file with negative words to compare to in sentiment scores.
Rasheed_ReadMe is a PDF with instructions and other administrative work.

4. Python packages/modules:
requests_oauthlib
webbrowser
json
requests

5. Data sources used:
positive-words.txt
negative-words.txt

6. Instructions:
1. Run Final.py
2. Enter Twitter verification if prompted.
3. Enter name of first movie to be searched. Include year if specification is needed.
4. Repeat Step 3 for the next two movies.
5. Read through analysis on screen.

7. Appropriate line numbers:
Sorting: 194
Class definition: 8
Creating instance of class: 180
Calling methods on a class instance: 192, 196-198
Importing a python module: 3-6

Using module in code:
requests_oauthlib: 53, 67, 87
webbrowser: 63
json: 44, 78, 83, 146, 162
requests: 145
Filter: 104

8. Rationale/Reflections:

I became more interested in my project as I began to write it because I was excited to use what I learned to collect data and numbers that can actually be analyzed. There were a few inconsistencies with my theory which I would have liked to work on if given more time and if I had more experience. One, if a movie name is referenced in a tweet, it's not necessarily used as a movie. For example, the movie name "Frozen" is used in contexts not referring to the movie. This would make my sentiment score less accurate. I attempted to fix this by adding a hashtag to the movie searches because I noticed if people hashtag a simple word like "frozen", then it is probably referring to something specific, often a movie. This also means sentiment scores will be more accurate if the movies are recent since many people will be tweeting their opinions on the movie. Second, an optional "year" search is needed for movies with the same title but multiple productions. "The Jungle Book", "2016" is a great example for a title and year search in addition to being recent and acclaimed. Last, I would have liked to search through more tweets and parsed through unhelpful words like spam or those written in other languages. Overall, I'm proud of what I learned to do in this class and the opportunities available through API. It was exciting to compare critic scores to real time sentiment scores and even more exciting to see when the correlations matched up. I'm excited to explore more APIs and methods in later courses.