Wrangle report

- Real-world data rarely comes clean. Using Python and its libraries, I gathered data from a variety of sources and in a variety of formats, assessed its quality and tidiness, then cleaned it. This is called data wrangling.
- The dataset that I wrangled (and analyzed and visualized) is the tweet archive of Twitter user **@dog_rates**, also known as **WeRateDogs**. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have adenominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "**they're good dogs Brent**." WeRateDogs has over 4 million followers and has received international media coverage.

1. Gathering each of the three pieces of data as described below in a Jupyter Notebook titled **wrangle_act.ipynb**:
    i. The **WeRateDogs Twitter archive** which I Downloaded manually
    ii. The **tweet image predictions**, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (image_predictions.tsv) is hosted on Udacity's servers so I downloaded it programmatically using the Requests library on the following URL: https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv
    iii. Each tweet's retweet count and favorite ("like") count. Using the tweet IDs in the WeRateDogs Twitter archive, I queried the Twitter API for each tweet's JSON

data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called **tweet_json.txt** file. Then I read this .txt file line by line into a pandas DataFrame with tweet ID, retweet count, favorite count and user count.

2. Assessing Data for the Project
    - After gathering each of the above pieces of data, assessing them visually and programmatically for quality and tidiness issues. I tried to Detect and document nine (9) quality issues and three (3) tidiness issues in my wrangle_act.ipynb Jupyter Notebook.
3. Cleaning Data for the Project
    - I cleaned each of the issues I documented while assessing. Then I performed this cleaning in wrangle_act.ipynb as well. The result is a high quality and tidy pandas DataFrames.
4. Storing, Analyzing, and Visualizing Data for the Project
    - I stored the clean DataFrames in a CSV files with the main one named twitter_archive_master.csv and the other file called Image_predictions_clean.csv.