Data Wrangling Report

In this project we worked with three primary datasets that we needed to combine,clean, and gather. One of the datasets was provided to us to me called *rateD*, one was to be extracted using the Twitter API where you had to create a developer account and get the permissions to be able to access twitter data or use the file provided by Udacity which I named *tweets*, and the last dataset was also extracted from the requests API in python where we obtained it through a URL which I named *images*. When I tried to extract the dataset from the twitter API I actually couldn't export it correctly because of a permissions issue or the tweets had been deleted, so I reverted to using the .txt file from Udacity. Once the file had been saved, I converted the file to a dataframe.

In terms of the assessment I made, I made both visual and programming analysis. Both the *rateD* and *images* dataframes didn't have tidy data where there were columns in both where there were multiple columns for variables. Some columns existed that had no use in terms of analysis. In the *tweets* dataframe, the data needed to be reduced because there were a lot of columns that weren't of any use or major analysis.

After we created copies of the dataframe, we had to tidy up some of the data which involved cleaning the data, and testing to see if the data became cleaner or not. I ended up deciding to clean the redundant retweets because we didn't want to count them as actual tweets and inflate the tweet count at the end of the day. The next issue I want to clean up was to move the **"doggo", "flooter", "pupper", "puppo"** to one column because its easier to read and analyze as well. Overall, tidying up wasn't too difficult but really taught me beneficial lessons on having clean data. Lastly insights, after some additional data cleansing. I found that the top 5 most common Dog Names are Lucy, Charlie, Oliver, Cooper, and Tucker. Then I observed the average activity per tweet is growing considerably and has increased activity by more than 20x in nearly 2 years. Lastly, I wanted to know if the increase in tweets is what led to more interaction but I found that tweets dropped by over 149% yet the activity has increased. This was a sign that popularity of the account had grown and followers enjoyed any dog content posted.