Wrangling report:

This is the report of the data wrangling I did in this project. Data wrangling is like doing a chore. Let suppose, it is the last day of school, student have played in the classroom and it looks messy and dirty. I need to clean the classroom and I am worried where and how to start. I went to school store and find something to gather all the items (used notebooks, used pens, pencils). I start looking to detect the clutters and dirt. After detecting the clutters and dirt, I divide them into books, pens, electronics.

In this project I did the wrangling on the data, to achieve this I divided the work into three sub work such as gathering, assessing and cleaning the data.

I used the data from the popular Twitter account with twitter handle @WeRateDogs, where various twitter users share pictures of dogs

Here are the three steps I followed in the project

1. Gather
   To start the wrangling, I have gathered the data such as
   - Twitter_archive
     The twitter_archive.csv file have been given by Udacity and I can assess by using the read_csv() module of pandas
   - Images-prediction
     The images-prediction is provided and hosted on Udacity server and I downloaded it programmatically
   - Tweet_json.txt
     To obtain the tweet_json file, I have query the twitter developer API the I separated tweet with and without twitt_id. In the next part, I combined tweets having tweet_id in a dictionary then copied the data into a new file called tweet_json.txt. The last step is to open the tweet_json file and create a dataframe with the help of tweet_id, favorite_count and retweet_count.
2. Assess:
   While accessing the data, I display the data in Jupiter notebook and sometimes in Excel. In Jupiter notebook, I use the info, counts, drop and many more operations on pandas dataframe .
3. Clean
   I created a copy of twitter_archive and imgs dataframe so that my cleaning should not affect the original data. I used define code and test, define indicates what to do, code is the functional part and test is the testing part. I deleted some data that are not relevant.