Wrangle Report

1- Gathering :-

Following instruction from Udacity project detail page first I downloaded Twitter archive file 'twitter_archive_enhanced.csv' which was provided for download, second file 'image_predictions.tsv' download programmatically from Udacity server using request library and third file was tweet_json.txt file using twitter API for @WeRateDogs , I read the file to create data frame for at least three columns id(tweet) , favorite and retweet counts.

2- Assessing :-

Asses Data Frames visually and programmatically and found lots of Quality and Tidiness issue but done minimum requirement for project 8 quality and 2 tidiness issue as below for cleaning .

Quality Issues :-

- **`twitts` table :-**

1- change time_stamp column to day,month and year

2- drop retweets that have images

3- delete columns that won't be used for analysis

4- remove twitt with rating_numerator = 960

5- edit twitts with not valid rating_denominator

-

- **`image_perdections` table :-**

6- remove duplicated

7- delete columns that won't be used for analysis

- **`twitts_json` table :-**

8- change id to tweet_id

Tidiness Issues :-

1- make one column for dog status

2- merge the three tables in one table

3- Cleaning :-

First I made a copy for each file then used basic python function like duplicates , drop, value_counts ,describe , info and others to comply with above mentioned point. I struggle with few issues and had to spend a lot of time to get my understand.