Reporting: wrangle_report

- Create a **300-600 word written report** called "wrangle_report.pdf" or "wrangle_report.html" that briefly describes your wrangling efforts. This is to be framed as an internal document.

Gathering Data

I downloaded the first dataset twitter_archive_data and uploaded it in the jupyter notebook, then i used the request library to download the tweet image prediction programatically, then i used the tweet.json txt to import the json file.

ASSESSING Data I used the info, describe to assess the three dataset forntheir quality and tidiness issues. I provided 8 quality issues which are:

1. The retweeted status_id, retweeted status user_id and retweeted status timestamp columns will be removed since they are all retweeted and we do not need the retweeted values. And these columns would be dropped after removing it.

2. The timestamp column that has a datatype of object should be changed to a timestamp datatype

3. All the different dog names that are not correct should be removed

4. Change all the tweet_id from the tables to a datatype of string or object

5. Ratings without images should be removed due to the project rules

6. none values should be removed and replaced with null

7. some ratings are not correct

8. some dog names are represented as none in the df_1 table and should instead be changed to null

9. The in_reply to status_id and in-reply to user-id should be removed and dropped

I also got 2 Tidiness issues from the tables which are:

1.According to the project rules the retweeted columns will not be needed after we get rid of the retweeted portions

2. In the df_1 table, the dog stages should be in one column because variables should be in columns and observation in rows

I made copies of the three dataset,then worked on cleaning the quality issues and the tidiness issues.