Wrangling act :-

By: *Mohammed A.M. Yassin*

This course provided me with a lot of information that form the solid base for my data analysis career and to make sure we (students) understood the concept the project was very helpful

How I have done my project
the solution divided into 3 part
1- gathering

2- assessing

3- cleaning

1- Gathering

I have gathered the data from 3 different sources :

1- twitter_archive_enhanced.csv : a csv file contains basic tweet data for all 5000+ of their tweets
2- Twitter API : I have used the the twitter API using Udacity twitter_api.py as I don't have developer account and I saved a JSON file into a data frame df_tweets
3- I have downloaded `image_predictions.tsv and saved images into a dataframe image_df`

2- Assessing the data frame had a lot of issues the issues I have found and cleaned :

1- Qualities
   1-1- tweet_id is integer
   1-2- timestamp are strings not datatime
   1-3- timestamp has +0000
   1-4- retweet present in data
   1-5- (in_reply_to_status_id , in_reply_to_user_id) columns aren't necessary
   1-6- tweet_id integer (in image prediction)
   1-7- 66 duplicate jpg_url (in image prediction)
   1-8- id is integer (in json data frame)

2- Tidiness
   1- many columns for stage of dog (name doggo floofer pupper puppo)
   2- Merge dataframe in twitter_archive_master

Cleaning

1- tweet_id change to string
2- timestamp have +0000 remove them using (strip method)
3- timestamp are strings not datatime change it to datatime using (to_datatime)
4- retweet present in data remove them (by using isnull)
5- (in_reply_to_status_id , in_reply_to_user_id) columns aren't necessary remove them by (drop)
6- tweet_id integer change to string by using (astype(str))
7- merge all column into one column dog_stage (used extract to choose the 4 columns and assign them to dog_stage column then drop them
8- Merge dataframe in twitter_archive_master (using merge)

Sorting

I have sorted data after merge them into a new data frame called *twitter_archive_master* and saved it into a new csv file *twitter_archive_master.csv*