The project was a to take a twitter data from a twitter account by the name @WeRateDogs and do the data wrangling steps to gather, assess, clean the data to prepare it to store it and visualize it.
the steps taken to prepare this data is:

quality assures

- renaming columns (p1, p1_conf, p1_dog, p2, p2_conf, p2_dog, p3,p3_conf, p3_dog) to (algo_A_prediction, algo_A_confident, algo_A_dog, algo_B_prediction, algo_B_confident, algo_B_dog, algo_C_prediction, algo_C_confident, algo_C_dog)

- Delete columns from df1(in_reply_to_status_id, in_reply_to_user_id, source, text, retweeted_status_id, retweeted_status_user_id, retweeted_status_timestamp, expanded_urls), from df2(jpg_url, img_num), from df3('contributors', 'coordinates', 'created_at', 'display_text_range','entities', 'extended_entities', 'favorited','full_text', 'geo', 'id_str', 'in_reply_to_screen_name','in_reply_to_status_id', 'in_reply_to_status_id_str','in_reply_to_user_id', 'in_reply_to_user_id_str', 'is_quote_status','lang', 'place', 'possibly_sensitive', 'possibly_sensitive_appealable','quoted_status', 'quoted_status_id', 'quoted_status_id_str', 'retweeted', 'retweeted_status', 'source', 'truncated','user') by using the drop() method in the pandas libaray

- dropping all the rows with rating_denominator > 10 or < 10 by making a query that gets all the rows with denominator bigger than 10 or smaller than 10

- change the id type from int to string in all the dfs by using the astype() method with parameter 'str' for string

- change the timestamp to date by using the pandas method to_datetime().

- divide all the rating_numerator by 10

- change the wrong predicted names of dogs (a , an , the) to null by using the replace method

- change the name of column id in the df3 to tweet_id to match the name with df1, df2.

- I must chose 1 algorithm since having 3 is inassure, depending on the mean of the confidence of the algorithm

- drop rows that aren't dogs according to our chose algorithm

tidyness

- type of dogs should be in 1 column.

- merge all dfs to one df

After doing this the data was clean and I was able to store the data in my machine for future use or sharing it with other