Introduction :

The dataset that you will be wrangling (and analyzing and visualizing) is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Data gathering :

1 - the first file we used it was twitter archive we called it using pandas load_csv method

2 - the second file we used requests to download it from udacity server and then call it using the pandas load_csv after saving the file as csv

3 - third file was the hardest to gather and get as we used twitter api it was very hard and took some time to gather the data and used alot of research so i can get the data from this api like the website this https://stackoverflow.com/questions/47612822/how-to-create-pandas-dataframe-from-twitter-search-api
And then after using the tweepy method we convert it to json file then from json file we read it line by line then makes list of this and then convert this list into pandas data frame

Ascending data :
We made it with some pandas function like info and sample and value counts and describe so we can know the problems of all the data sets and clean it

Clean the data :
We got this cleaning problems :
  quality issues:
   1 - file1 timestamp change it from object to timestamp
   2 - melt the columns doggo floofer pupper puppo into one column called dog type
   3 - making the rating_numerator for the people of puts fractions in rating
   4 - fixing the rating of the rating_denominator not equal 10
   5- solving the problem of some rating are above 10

   6- drop useless columns

   7 file2 drop duplicated images

   8 - remove retweeted tweets from json fil

Tidness :

1 - making column named rating divide numerator on denumrator

2 - merging the files and drop duplicated.

Then we visualized some of the data