Introduction:

The dataset that I 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 6 million followers and has received international media coverage.

wrangling and Analyzing data:

summary:

- Gathering Data.
- Assessing Data.
- Cleaning Data.
- Storing.
- Analyze & Visualize.

Data sources:

- Enhanced Twitter Archive (.csv).
- Additional Data via the Twitter API.
- Image Predictions File.

process details.

1-Gathering Data:

I gathered data from three different sources and read

With three different format.

- **twitter_archive_enhanced.csv.**

  I downloaded it manually then read into a dataframe with pandas library.

- **image_predictions.tsv.**

  I downloaded it programmatically using the Requests

  library and the URL which supplied.

- **Data via the Twitter API**
  Tweet_json.txt

  Note: I failed to have dev acc so I used the code proved by udacity, and download the result from the classroom.

2-Assessing Data:

after gathering data, i displayed all data frames.

and assess it manually & programmatic.

i used pandas' functions and methods are used to assess the data. and found some quality and tidiness issues like (datatype, values are column name, ..., etc).

3- Data Cleaning

after assessing data, I going to the clean with the target issues, that I found it in assessing.

first I copied all the original pieces of data. And going to start the clean, and these are some examples of the issues that I cleaned:

- Merged data frames together.
- Datatype in ('rating_numerator', 'rating_denominator') changed to 'float' .
- Datatype in 'tweet_id' changed to 'str' type.
- Correcting names.
- (doggo, floofer, pupper, puppo) columns melt into dog stages.
- Deleted all duplicates.

4- Data Store

after cleaning I stored the clean DataFrame

in twitter_archive_master.csv, and make it ready for analysis.

5- Analyze and Visualize.

after wrangle data, I found some interesting insights and visualization

like (the most favorite tweet and the most retweeted, The most common dog breed, ...etc).