

Wragling report

We rate dogs udacity

In this work we investigated the tweets data of the WeRateDogs tweeter account, to identify the rating as a function of a time. The rating was taken at certain interval of time and the corresponding data was stored. In order to perform some programming analysis we employ python and jupyter notebook. The initial data was not organized or cleaned so we have to remove unnecessary information and clean the data.

The investigation of WeRateDogs consist on three steps: gathering, assessing and cleaning the data. The final result should contain the information that can be processed and analyzed in order to know the rating of the dogs over the time.

- (a) In the **Gathering process**, we use the dataset recommended by Udacity to get WeRateDogs Twitter archive and downloaded the file `twittwe_archive_enhanced.csv`. The second file is called tweet image prediction it was downloaded by using the program at the address given by Udacity. The third step consisted on using the tweet retweet count and favorite to get the tweeter API and save in the `tweet_json.txt` file.
- (b) To identify the quality of the date we processed to assess the collected data. Here we identified the column names with invalid names for example "None", "a". The `jpg_url` was removed in some `tweet_ids`. The data was additionally manipulated until we get the desired data.