- First of all I started gathering the three pieces of data as described below , I started downloading the WeRateDogs Twitter archive. By Downloading this file manually by clicking the following link: `twitter_archive_enhanced.csv` and uploaded it to the jupyter notebook.
- Then I continued with the second step which is The tweet image predictions file, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (`image_predictions.tsv`) is downloaded programmatically using the Requests library and the following URL: https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv
- Then I used the attached tweeter file `tweet_json.txt` in order to complete the gathering process to start assessing the gathered data which will be my second step.
- The second big step in my analysis is Assessing Data , After Gathering the required data I started assessing data visually and programmatically for quality and tidiness issues. In order to Detect and document at least **eight (8) quality issues** and **two (2) tidiness issues** in the `wrangle_act.ipynb` in the Jupyter Notebook.
- In assessing the data I started to display the data and how it was separated and distributed through the dataset in order to get a full image of how to get the full use of the acquired data to assess , clean , analyze and the visualize the insights of our analysis that we reached after our efforts in trying to clean these data as most as we can in order to get an accurate result as much as we could without biasing to a certain direction in our analysis.
- The third big step was cleaning data , according to my assessing work that was documented from the last step.
- Then it comes to storing the gathered , assessed and cleaned data in a new CSV file `twitter_archive_master.csv`.
- Finally , I started Analyzing, and Visualizing Data for this Project using matplotlib .