wrangle Report

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 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 are good dogs Brent.
WeRateDogs has over 4 million followers and has received international media coverage.

I work in this project using Jupyter Notebook and importing this libraries :

- pandas
- NumPy
- requests
- tweepy
- json

PROJECT STEPS:

- Data wrangling, which consists of:

    - Gathering data in this project I worked in three separated data resources :
      A. twitter_archive_enhanced.csv local file downloaded from udacity website
      B. `image_predictions.tsv` is hosted on Udacity's servers and should be downloaded programmatically using the Requests library and the following URL:
      https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv
      C. JSON data in a file called `tweet_json.txt` file

- Assessing data : After gathering each of the above pieces of data, i assess this quality and tidiness issues:
  Quality:
  1- Remove tweets with Retweets From df
  2- Drop retweeted_status_id , retweeted_status_user_id , retweeted_status_timestamp From df
  3- Drop retweet_count From additional_data
  4- Ratings that have a decimal in them are incorrectly extracted should be corrected.
  5- Fix type of timestamp to be datetime & Fix type of tweet_id to be str in df table
  6- Fix type of tweet_id to be str in additional_data table
  7- Fix type of tweet_id to be str in img_pred table
  8- Replace Wrong Names in name column in df table to be None

Tidiness:
1- Group dog names in one column

2- Merge the three separated tables in one

- Cleaning data : All the Quality and Tidiness issues solved in the cleaning part programmatically and tested to make sure of that.

- Storing: The clean copy saved in csv named as witter_archive_master.csv .

- analyzing and visualizing : The last steps in this project is Analyze and visualize the wrangled data .