INTRODUCTION

I used regular expression (re) library to extract data from 'tweet_json' file because I try to create a developer account on Twitter but is rejected many time.

In this project (Wrangle and Analyze Data), i will use tables from weRateDogs Twitter account with some libraries such as:

pandas - for wranle data from file on jupyter notebook, numpy - used for random, re- regular expression - used for extract data from tweet_json file because we dont have a developer account in Twitter, and datetime. etc.

to wrangle and analyze data step by step as following:

WRANGLE:

Gather - collect the important data from files and another sorcese.

Assess - checking the tiddy and messy data and definde the solveing.

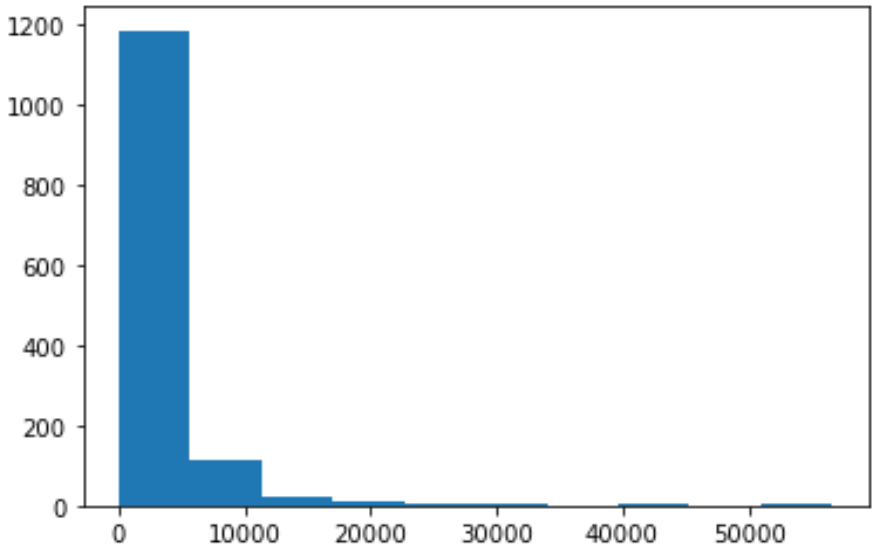
Clean - start to clean and arrange our best data, which we can analyes it.

- I uesed pandas library to wrangel data from 'twitter-archive-enhanced' and 'image-prediction' with re library uesed to extract data from 'tweet_json' file.
- compire between three fils above by using 'tweet_id' column.
- fixed some data from our dataframe to handle the clean data for provide it.

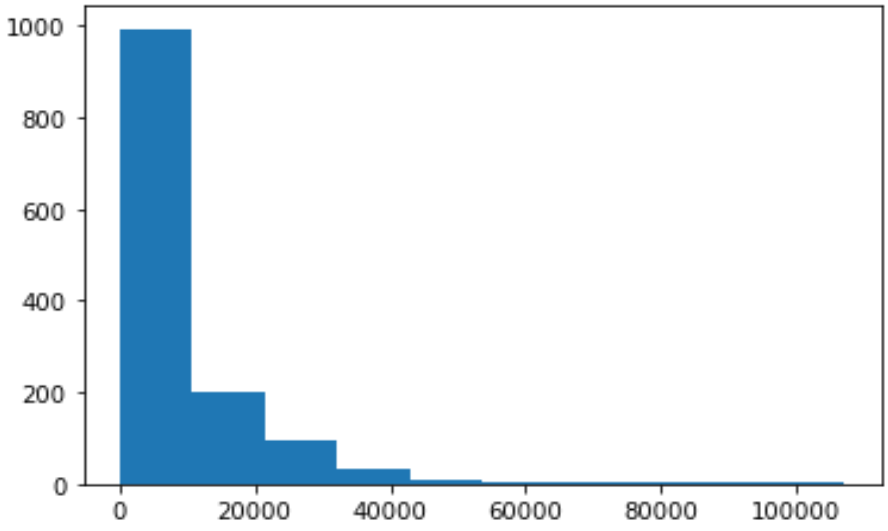
ANALYSIS

- **weRateDogs account on Twitter.**
- The importants on Data Analysis this section after import required library and read dataframe, compare between retweet count and favorite count for weRateDogs account on Twitter.
- on 2017 the favorite count become grow more than retweet count.

- The rating of numerator is grow fast.
- There are some dogs photo url get more favorite count.
- favorite count more than retweet count , it is after 2017 it is cleared by looking the scatter plot.
- we thing the reson of retweet count less than favorite count because many users fast action is by clicking the like button.
- there are few photo make a lot of retweet because it is the most better cute dogs.

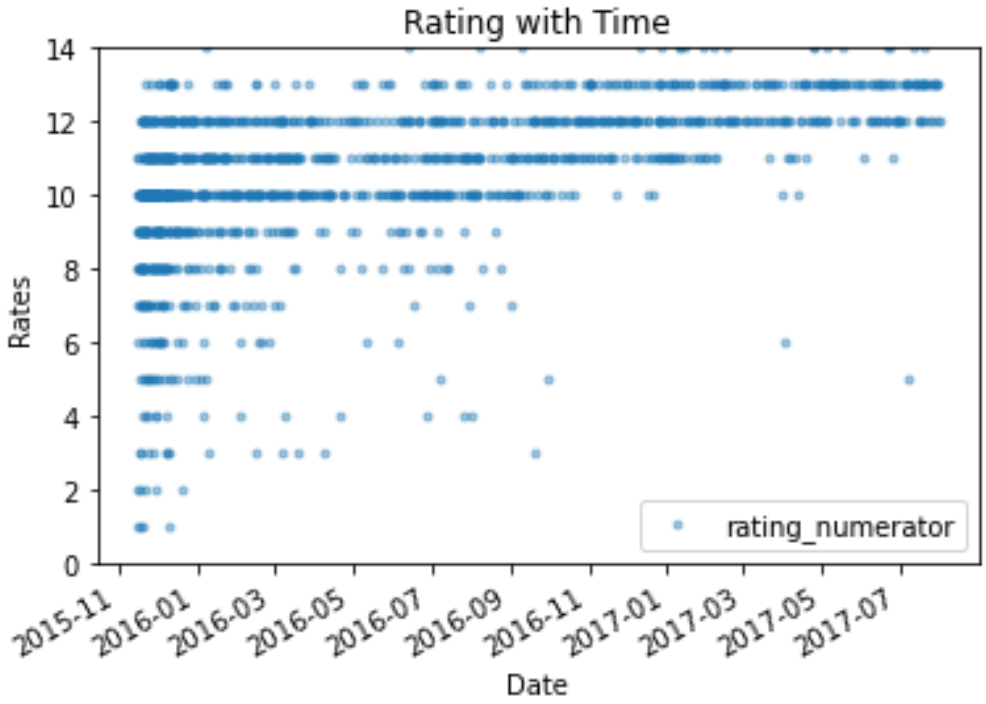


- **This Plot Histogram of retweet_count.**

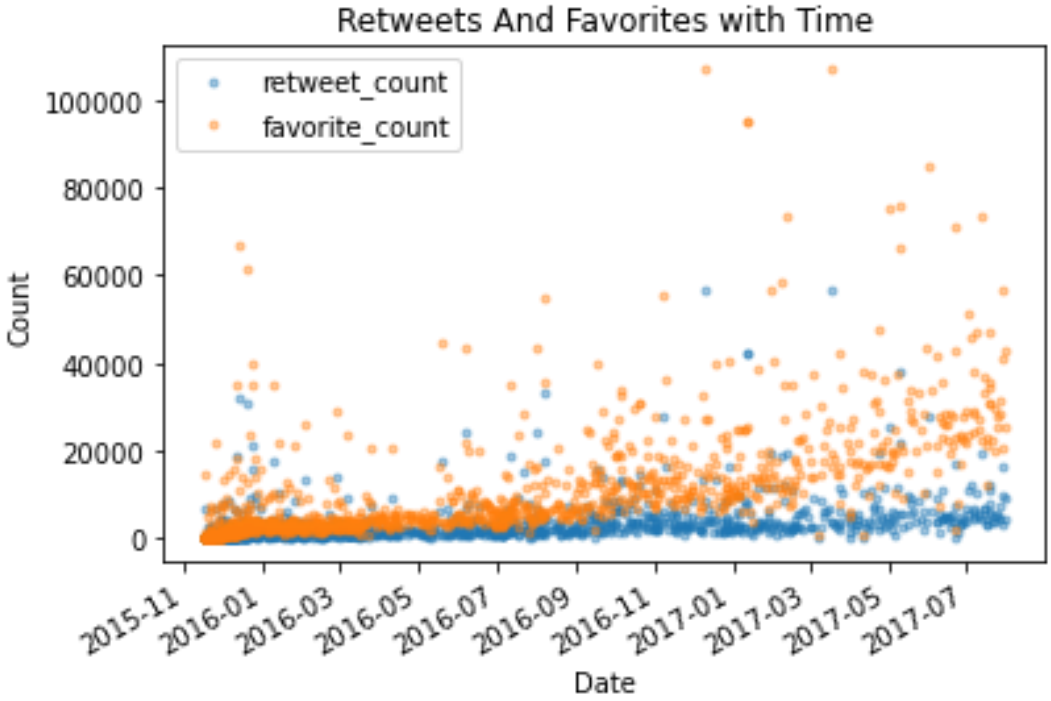


- **This Plot Histogram of favorite_count.**

- **We found plot scater rating with time.**



- **We found plot scater retweet and favorite with time.**



CONCLUSION

We have the beutiful Dogs image with max favorite count rating.

finaly we wrangle the dataset of weRateDogs Twitter account by using Gather data from many sources with defrente ways such as here use read files with Pandas library and extract data by regular exprission library, then merge these files and assessing to find some null value and untiddy column need to clean, after cleaning our data set, we Analyze data by check which photo of dogs get high count by favorite photo tweet.