Data Source and Gathering

WeRateDogs is a twitter account that care about dogs rating and they have more than 5,000 of their tweets on Aug. 1,2017. They have sent their archive to Udacity contains many features (tweet_id, text, timestamp, ratings, etc). The data is consisted of many features also like number of retweets or number of loves. To download more additional data, used Tweepy to authenticate to my account to get data. Eventually, I gathered all data frames in one data frame we can get more information from it.

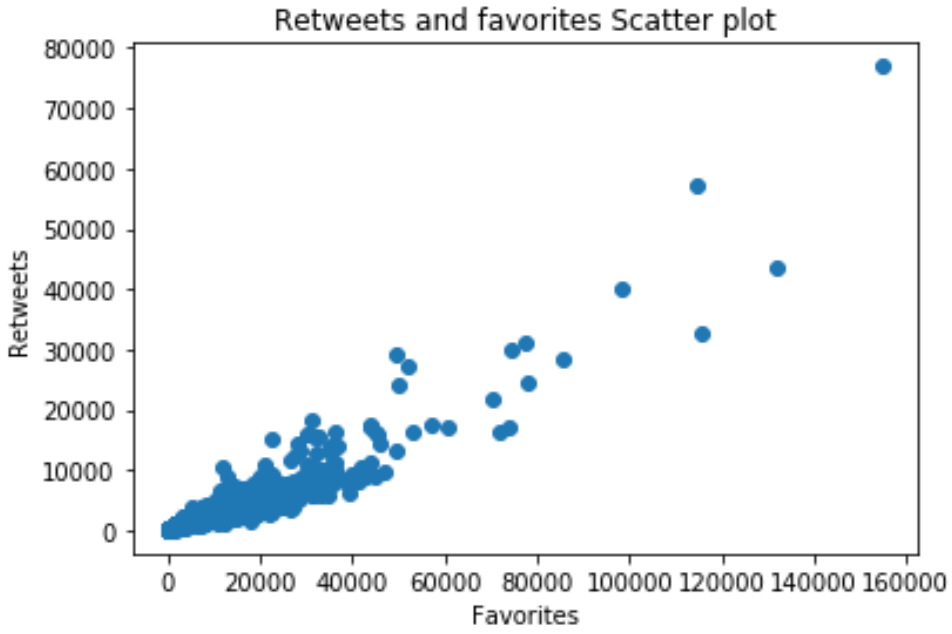
Wrangling Phase

Before going deep into Analyzing data, I tried to assess the data for quality and tidiness. Quality as we know, is measured by how valid and accurate the data are. Deleting the columns that will not add anything. Deleting duplicated tweets ID or any column that is consider a Primary key or unique. After many edits, I merged the data frames to result the final data frame in which we can answer our questions.

Insights

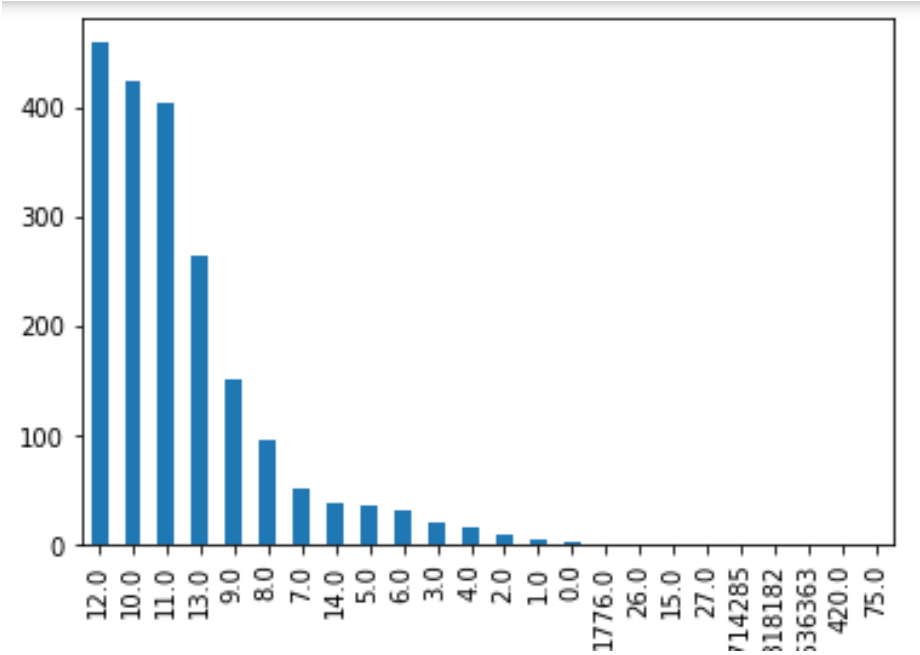
It is obvious that favorite count is more that retweets count

1-



The relation between number of Retweets and Favorites is strong relation that refers to a proportional, seems linear, relation.

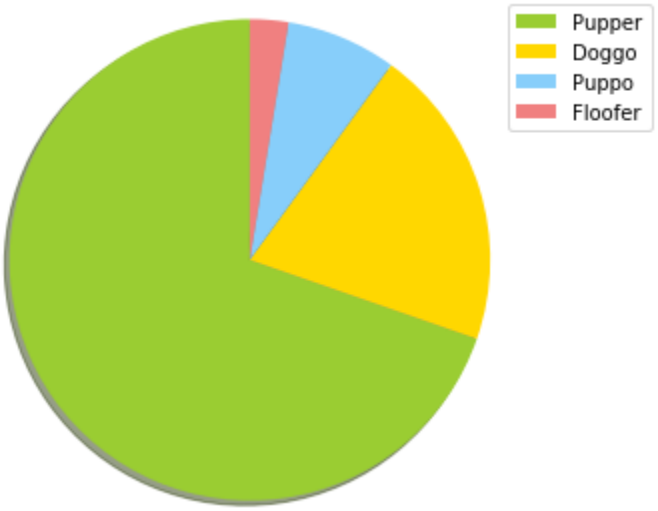
2-



The bar plot shows that the descending order of ratings of dogs. Noticing that Rating total parameter is defined through the next equation.

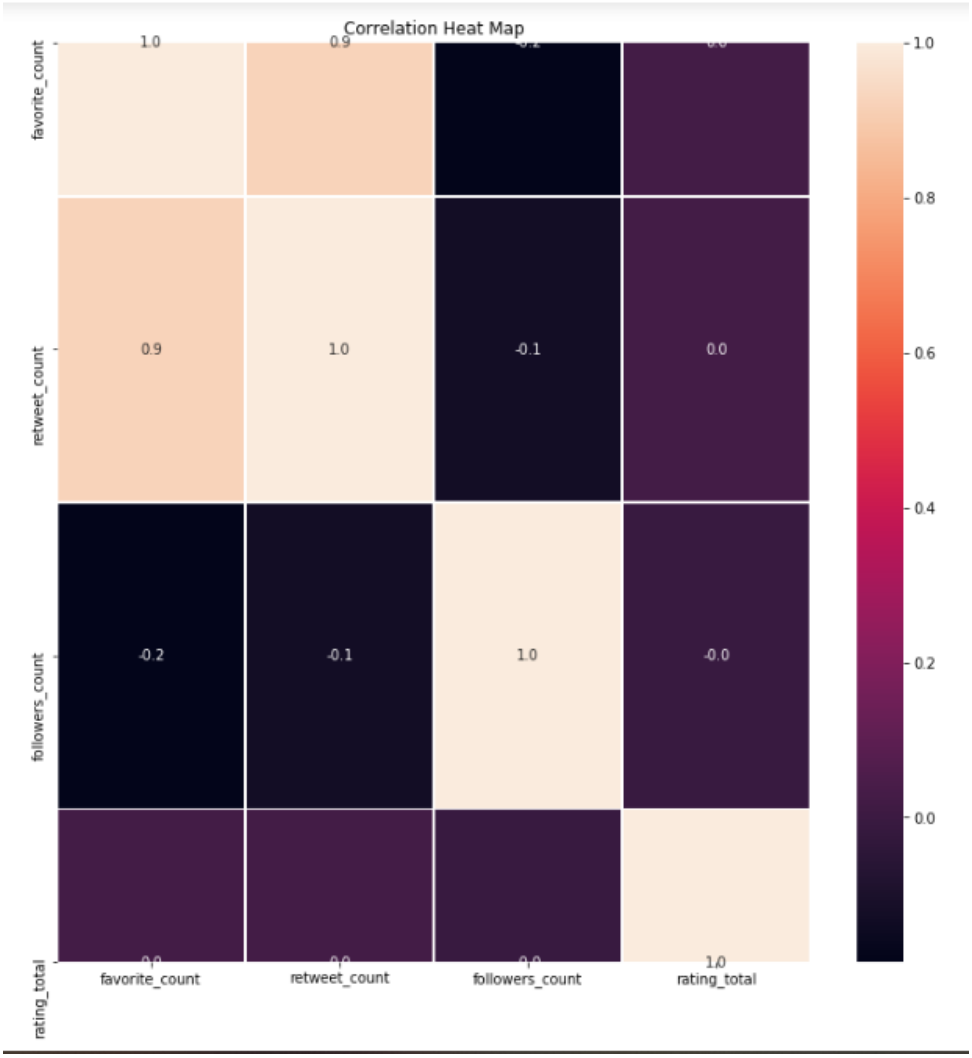
$$rating_{total} = \frac{rating_{numerator}}{rating_{denominator}}$$

3-



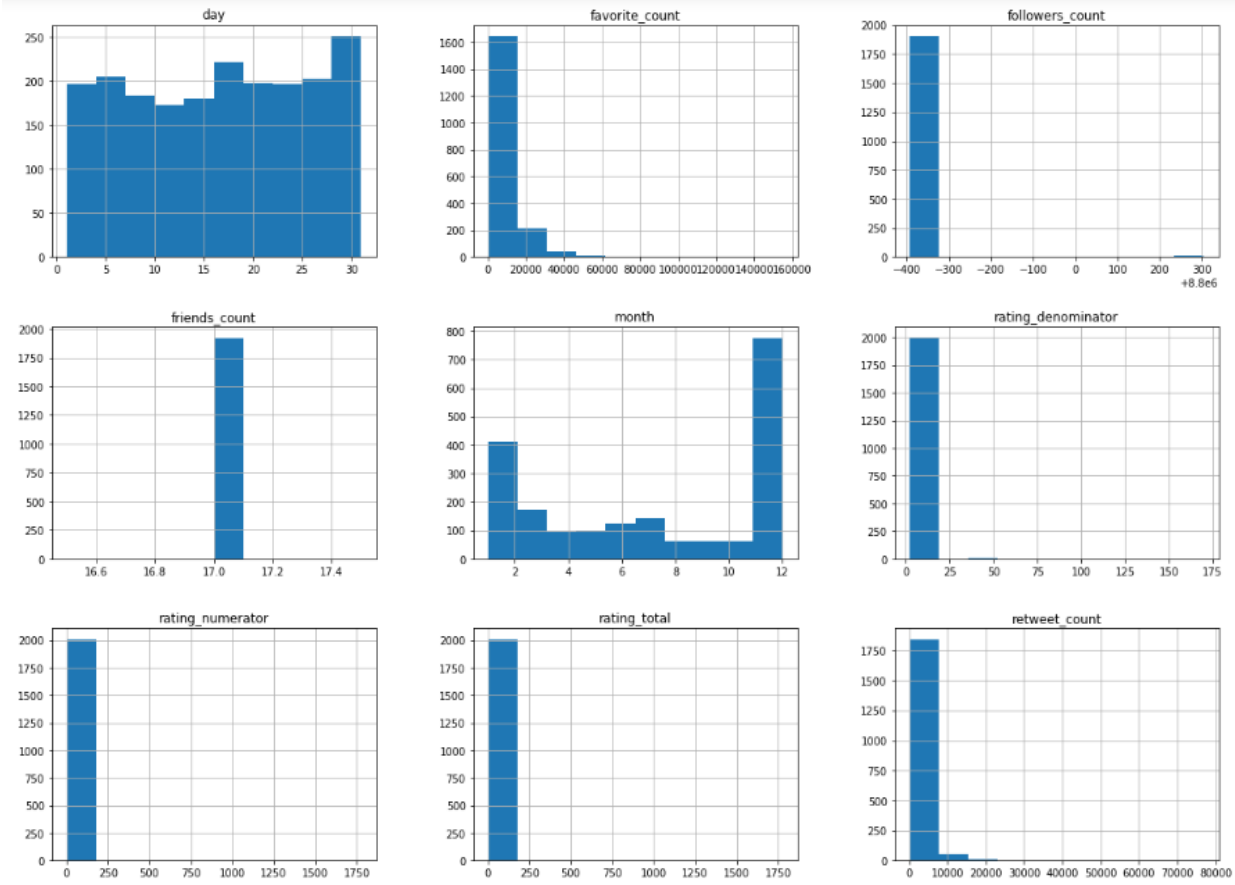
The Pie plot refers to how each part represents dog main type in a pretty representation.

4-



The above heat map represents the relation between four parameters and each other like a matrix 4x4.

We can also represent it in histograms of all columns of our data to see the relationship among each other.


day
favorite_count
followers_count
+8.8e6
friends_count
month
rating_denominator
rating_numerator
rating_total
retweet_count