Analysis Report

Based on a data frame of several tweets from WeRateDogs(provided by twitter_archive_master.csv file), I would like to investigate some of the data variables.

1- Univariate Analysis:

First, I explored the rating distribution, and I found it left skewed, was most values about 79% between (10 & 14).
That was expected as it is a main characteristic of the page rating system and the funniest thing about it.

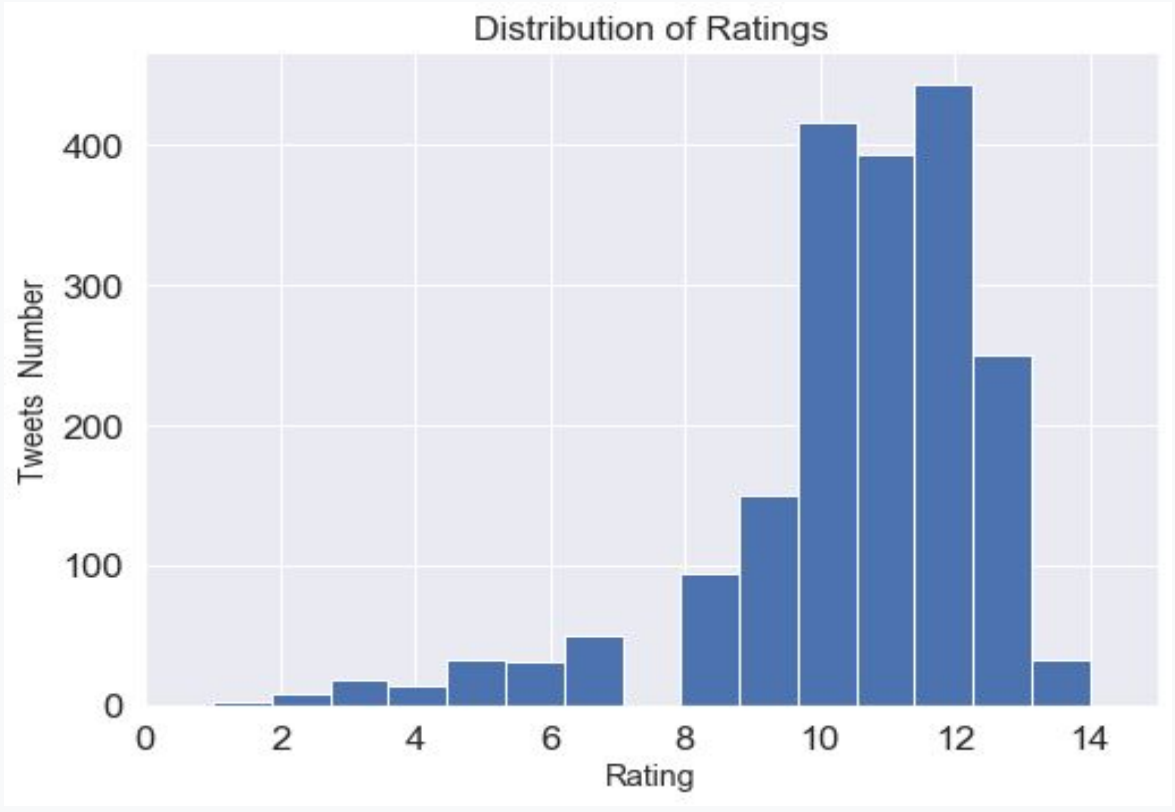

Figure.1

Second, I tried to figure out which breeds are present the most in our dataset, so I got the following plot that showing, ‘Golden- Retriever’ breed as the most tweeted breed.

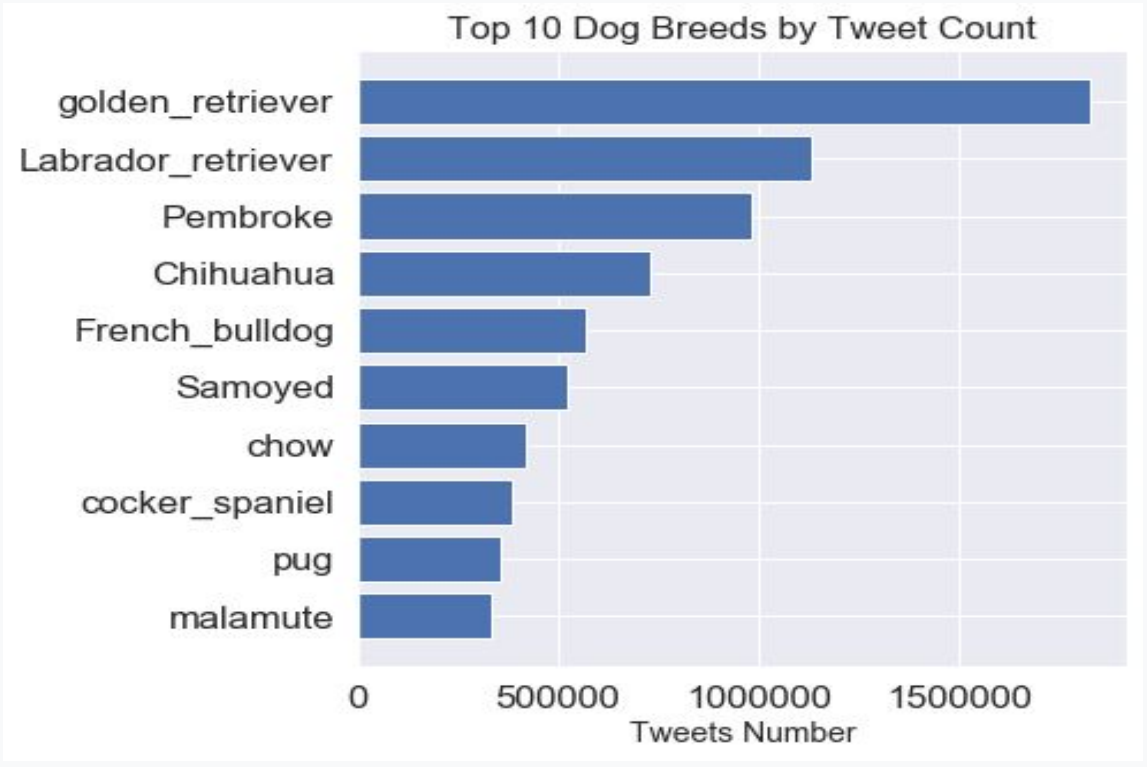

Figure.2

2- Bivariate Analysis:

Later exploration will reveal that it's also the most favourite breed through the archive. The following plot shows that.

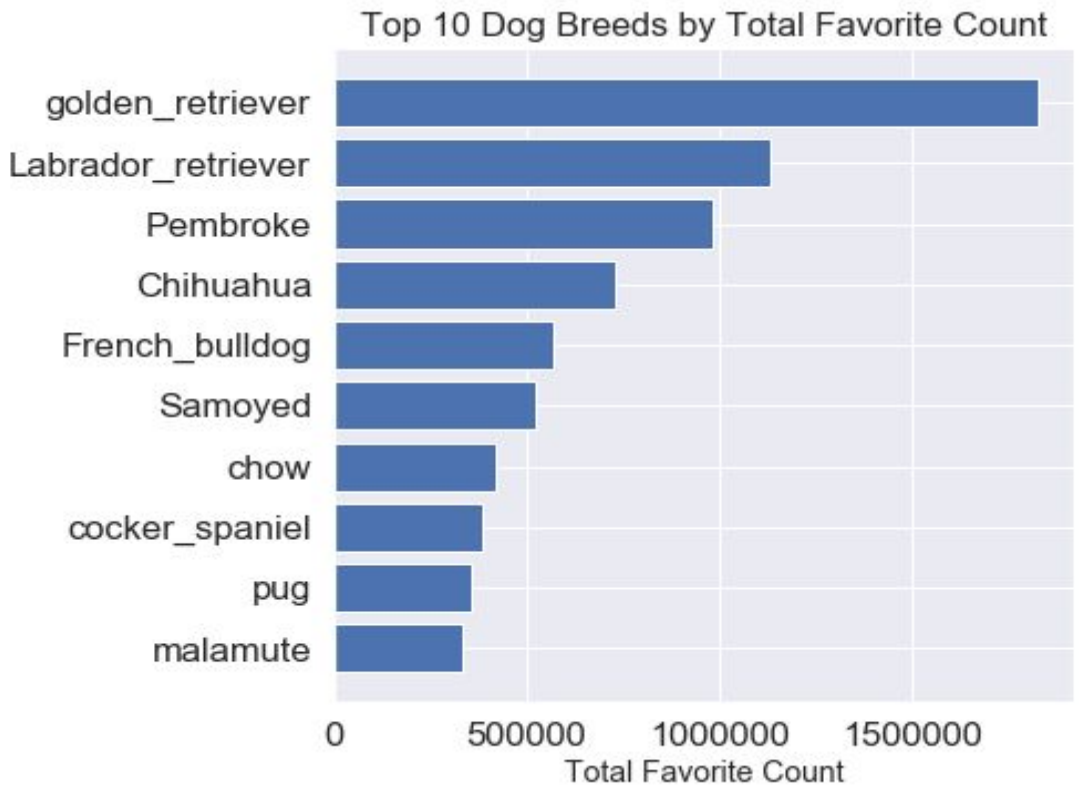

Figure.3

3- Multivariate Analysis:
In this section I explored the relation with time colored by rating values.

- In the following plot, I explored the favourite counts over time, colored with dog rating.

Figure.4

Clearly, the more recent tweets (on the right side) are rated consistently higher (blue and green) than much older tweets (yellow and green).

The slope of the line is also increasing at a steady rate from around january 2016, reflecting the increase in tweets being favorited. This clearly shows that the Twitter account followers were active, and possibly growing, during the timeframe represented in the archive.

- In the next plot, I explored the relationship between favourite and retweet counts colored with dog rating values.

Figure.5

This plot is showing that there is a strong positive correlation between retweet counts and favorite counts, and actually that does make sense, as if you 'like' tweet (favourite it), you are more likely to retweet it as well.

The tightly packed points follow a distinct linear trend with a positive slope, so as favorite count increases so does the retweet count.

Also evident, tweets with high favorite counts and high retweet counts also have higher ratings, as shown by the concentration of blue points.