Analysis of WeRateDogs

WeRateDogs is a twitter account that rate dogs on the basis of tweets. It gets images of dogs from owners and rate them. It posts the same on its account, comment on it and post it with the ratings they provide for them..They always rate the dogs on scale of 10. At times they increase this scale on numerator side.

For example, 13/10 or 14/10.

It has 6.77M followers and 6,933 tweets at the time of writing this report. Most Recent tweet:
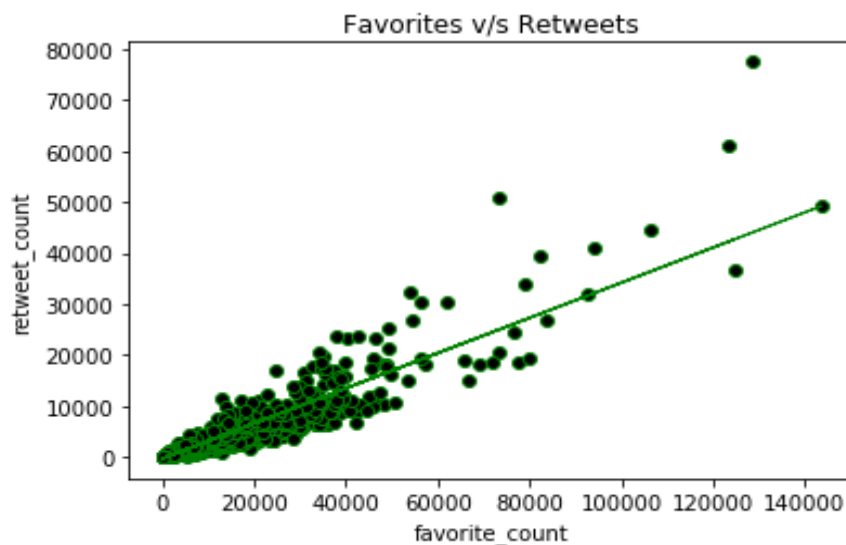
Libraries imported for Analysis in this project

- Tweepy

- Json

- Pandas

- Numpy

- Time

- Scipy

- Os

- Re

- Matplotlib

- Requests

About the Data

To analyze the tweets, we have used different verified sources. The first is an archive of the past tweets provided by Udacity via a csv file named "twitter_archive_enhanced.csv". The second source was using Twitter API tweepy. It fetched further details of the dataset already with us from the WeRateDogs twitter handle. The third source was also provided by Udacity that contained images predicted via a neural network that has classify the breeds of dogs and stored in a csv file named "image_ predictions.csv". There were 2336 entries initially and after combining I ended up analyzing 2174 tweets from WeRateDogs.

Favourites and Retweets of Dog stages, I observed the relationship of favourites and retweets. So a plot was made (shown below) that most dog tweets had more favourites than retweets.
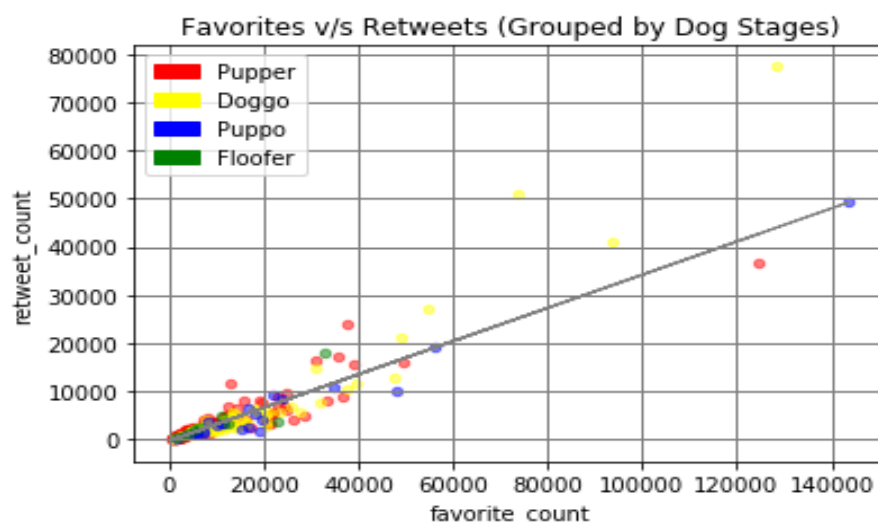

Relationship between favourite_count and retweet
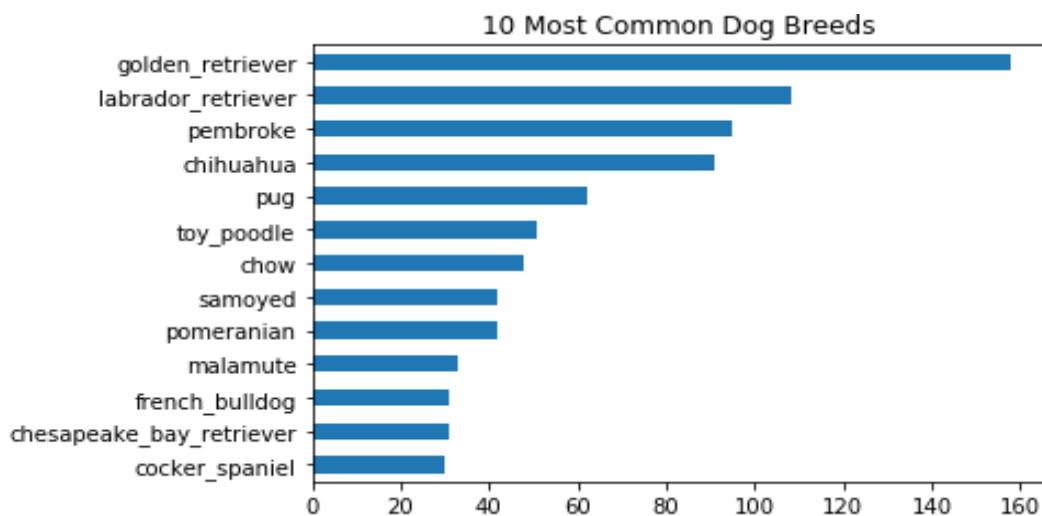
I plotted a scatterplot of all tweets.

It can be seen that most of the tweets have less than 40,000 favorites.However we have seen that as the value increases, the ratio of retweets to favorites also increases.

I scatterplot for the different dog stages .On grouping all the dog stages together I obtained this plot:
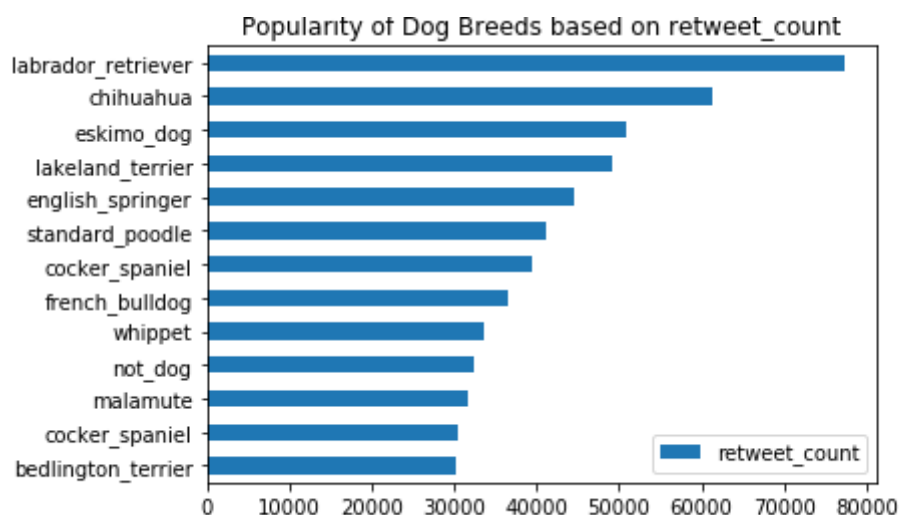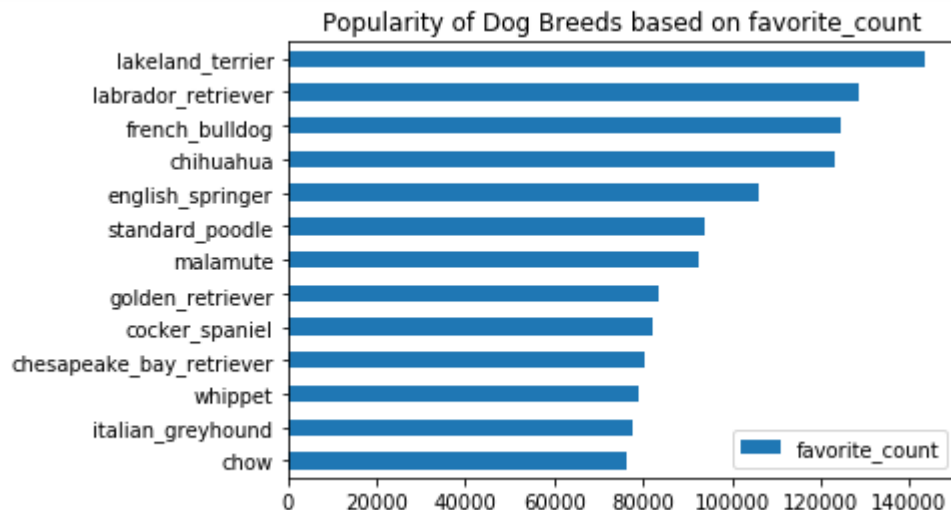
Most common breeds

The 10 most common breeds are:

The 10 most common (actual) dog breeds were plotted. It can be seen that according to this data, the most common dog breed in the tweets was the Golden retriever with 158 tweets. The next five most common breeds were the Labrador Retriever, Pembroke, Chihuahua, Pug and Toy Poodle.

On analyzing the common breeds and sorting them on the basis of retweets and favourite_count ,I have plotted the following plots:

Popularity of Dog Breeds based on favorite_count

On averaging the retweets and favorites for each breed , only the French Bulldog appear in one of these plots as well as the plot of the 10 most common breeds. This suggests that the most common dog breeds on average didn't have the highest retweets favorites. This might make sense since most retweets and favorites are on less common breeds. This relationship could be interesting to investigate in the future.

Most of the breeds appear in both graphs; the dog breeds like Eskimo dog only appear in the top favorite chart and the dog breeds cocker spaniel,and malamulate only appear in the top retweet chart. This suggests that having a high number of retweets is positively correlated with a high number of favorites.

This suggest that Lakeland_terier is most popular breed.