We Rate Dogs Analysis

The dataset wrangled, analyzed and visualized is the tweet archive of WeRateDogs, a twitter account that rates people's dogs alongside comments and ratings for the dogs. Most of the dogs were rated above 10 because they're good dogs Brent. One of such dogs is Logan with a rating of 75/10. The dog had 17387 likes and 5745 retweets.

Figure 1: Image of Logan (75/10)

For the purpose of this analysis, three datasets were gathered and wrangled to create the main dataset. The data collected were:

- The Twitter enhanced archive dataset
- The Image Predictions dataset used for predicting dog breed
- The twitter data collected online using Tweepy.

After wrangling, the main dataset had 1963 rows and 11 columns containing original tweets from November 15th 2015 to August 1st 2017.

Most of the tweets were sourced from an iPhone (Twitter for iPhone). About 98% of the original tweets had their source from an iPhone indicating a possible preference of twitter users for iPhone.

Table 1: Table showing the frequency of tweets' source

Source	Frequency	Proportion
Twitter for iPhone	1925	98.0642
Twitter for Web Client	28	1.4264
Tweet Deck	10	0.05094

With the help of the Image Prediction file, and using the images of the dogs, the dog breed was predicted using Neural Network. However, some dogs were not correctly classified as dogs. Out of the 1963 dogs, 305 were not correctly

classified. For the classified dogs, the bar chart below shows the top 10 most predicted breed.

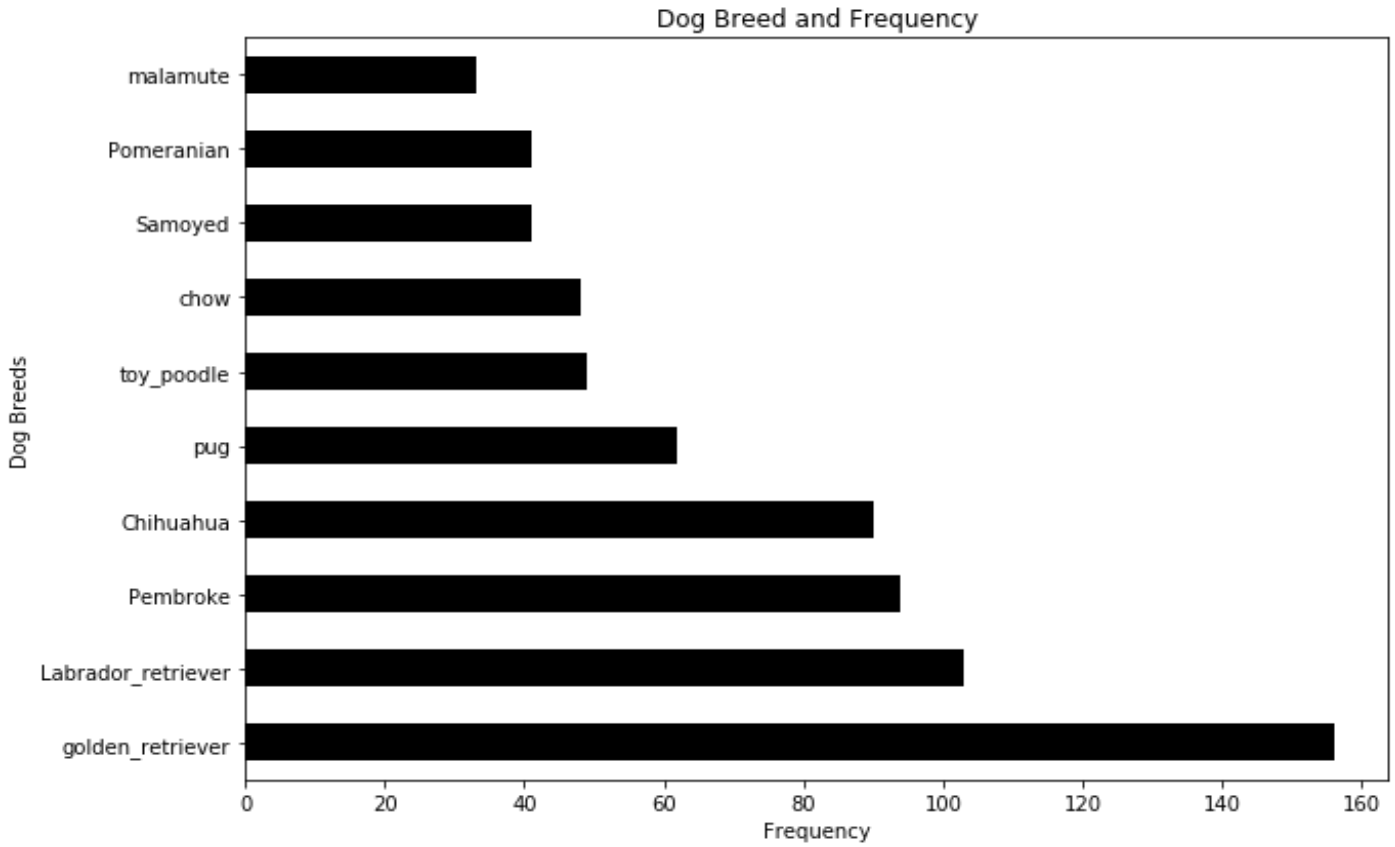

Figure 2: Frequency Distribution of top 10 dog breeds

The four stages of dog breeds were plotted using the Pie Chart in Figure 3 below. From the Pie-Chart, the greater proportion of dog-stage was **pupper** accounting for 67% of the dog population in the dataset.

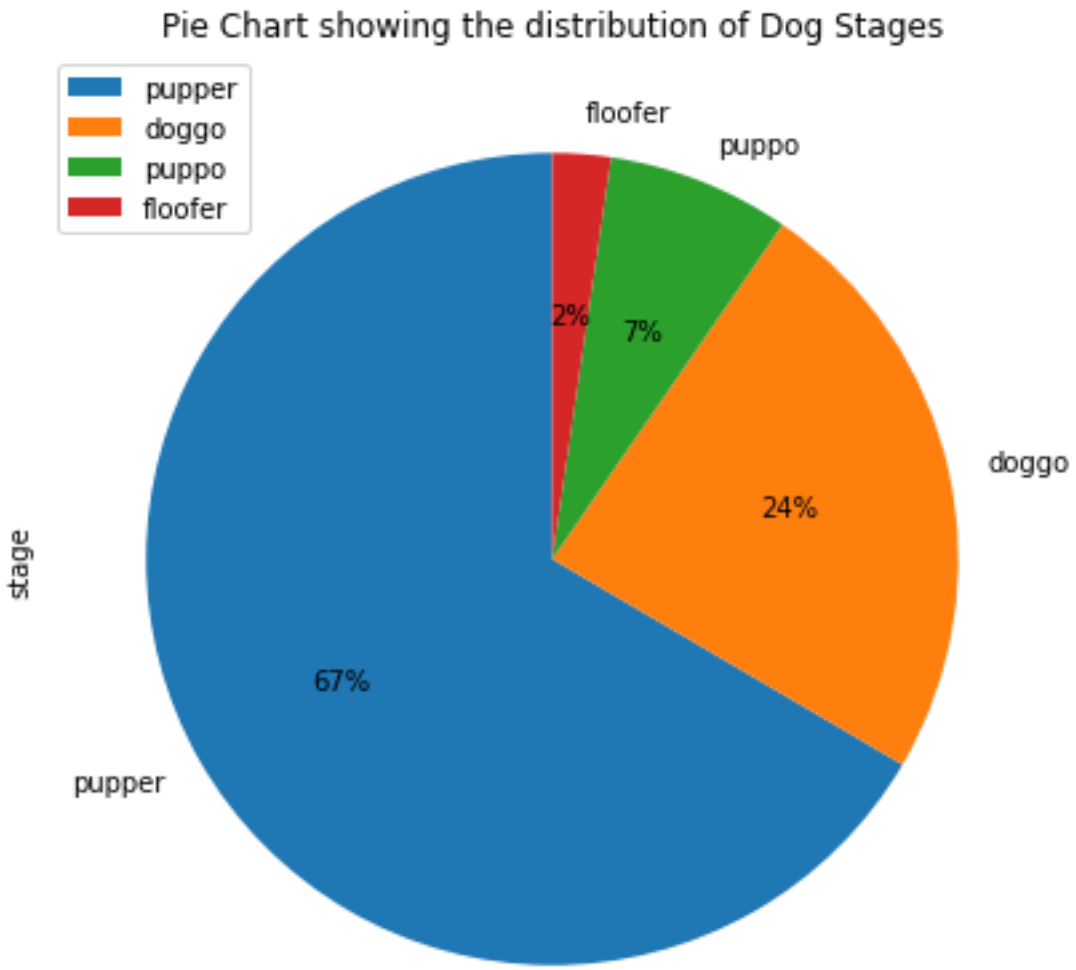

Figure 3: Pie Chart showing the distribution of dog stages

In most cases, dog images with high number of likes usually have a corresponding high number of retweets (and vice versa). To confirm this claim, the correlation between Retweet Counts and Favorite Counts was computed and found to be approximately 0.93 indicating a strong positive correlation between Retweet counts and Favorite counts. To further ascertain this claim, a Scatterplot of Retweet count and Favorite counts was plotted.

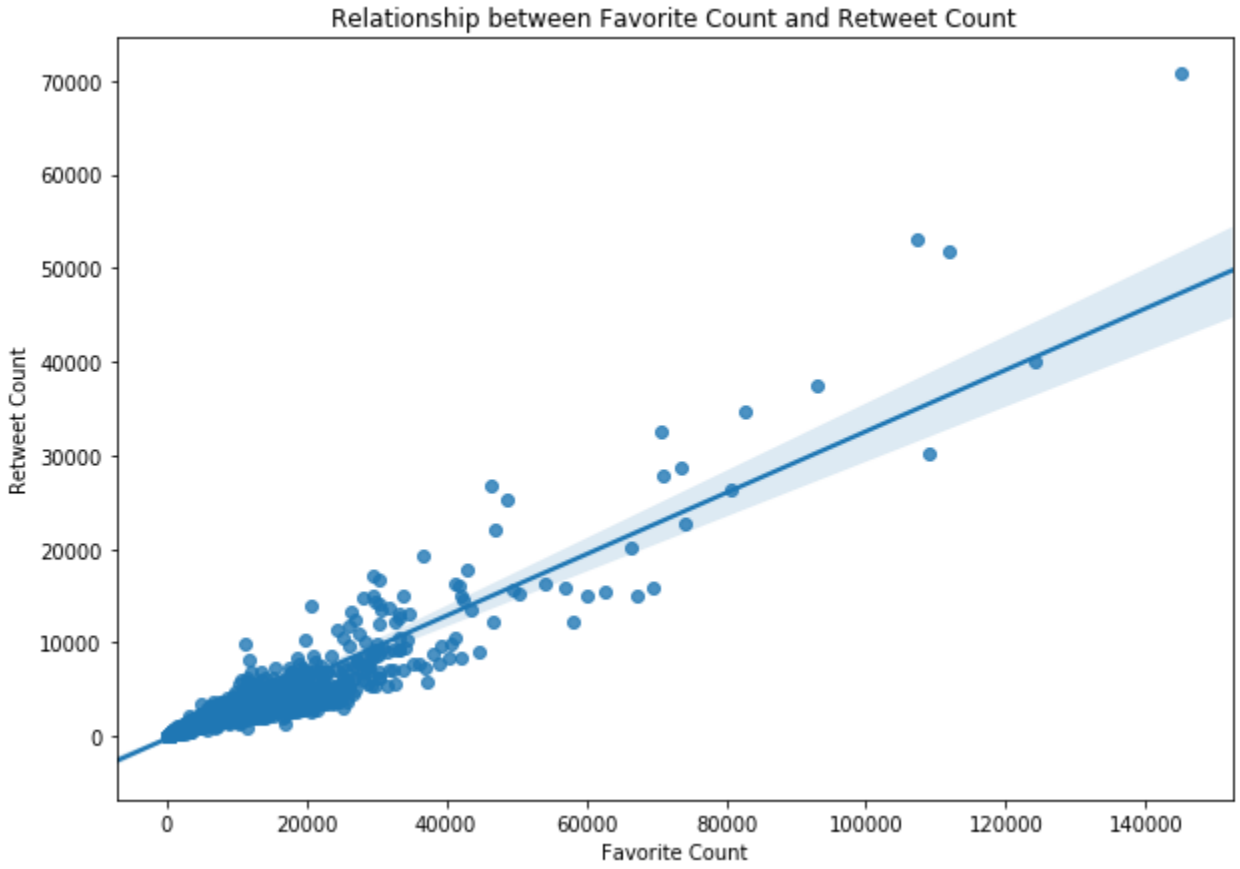

Figure 4: Scatterplot showing the relationship between Favorite and Retweet Counts