WeRateDogs Twitter Analysis

For this report, I will be investigating the twitter archive of the WeRateDogs twitter page. The data I am using is made up of their twitter archive, image predictions and additional twitter information. I have looked at the following:

- Is a dog stage more tweeted about than another? Does this translate to more favourites/retweets?
- Does more retweets mean more likes?
- What is the most favorite breed?

Is a dog stage more tweeted about than another?

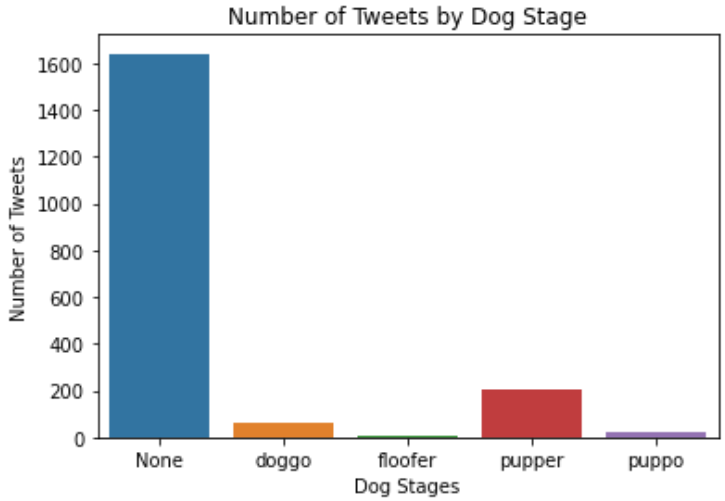

Figure 1: Bar Plot of Number of Tweets by Dog Stage

From the above graph, we can see that a large proportion of tweeted Dogs have not been classified as either doggo, floofer, pupper or puppo. Due to this it makes it hard to look at the 4 stages.

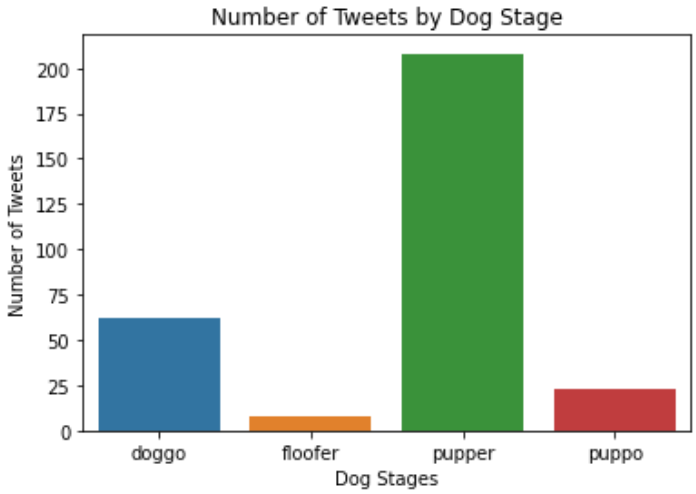

Figure 2: Barplot of Number of Tweets by Dog Stage without None Classification

When taking out the 'None' classification, we can see that the pupper stage was the most tweeted compared to the others. However there are many dogs unclassified so this may not be a good representation of the data over all. Some may argue that if pupper received more tweets they would have more exposure to gaining likes and retweets.

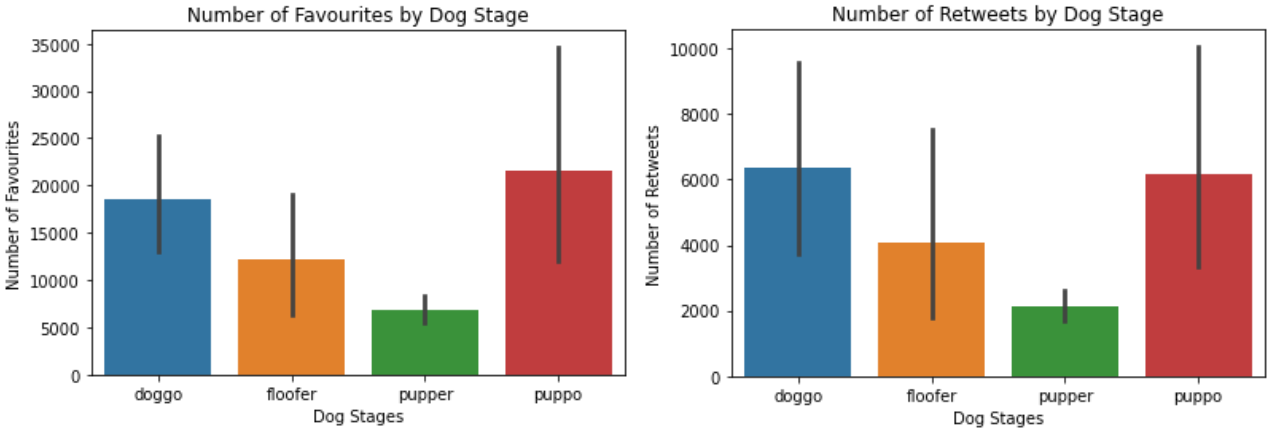

Figure 3 & 4: Number of Favourites and Retweets by Dog Stages

The charts above show is the number of favourites and retweets by Dog Stages. I made an assumption that if pupper was tweeted about more compared to the other stages then they would be more likely to have more favourites and retweets. However the graphs in Figure 3 and 4 tell us otherwise. For both, puppo gained more favourites and retweets than the other stages (have to remember that a large proportion of tweets are unclassified).

So despite being tweeted about more, pupper received less favourites and retweets. I cannot say if there is a definite link between being tweeted about more and favourites/retweets but can be an area explored about more.

Do more Retweets mean more Favourites?

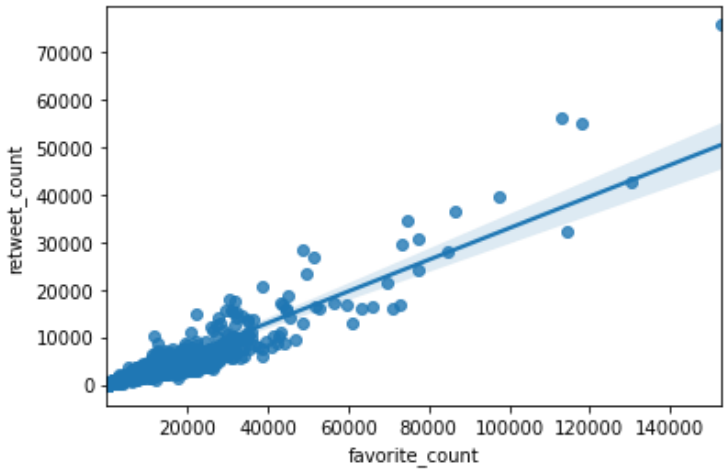

Figure 5: Favourites by Retweets

From the above graph, we can see that there's a strong relationship between the number of retweets and the number of favourites a tweet receives. The regression line is sloping upward showing that it is a positive relationship. To confirm this, we can have a look at the correlation coefficient between the two variables:

	favorite_count	retweet_count
favorite_count	1.000000	0.928448
retweet_count	0.928448	1.000000

Figure 6: Outcome of using .corr()

Using .corr() on these two variables, we can see that the correlation coefficient (by default Pearson's) is 0.928. This solidifies that there is a strong positive correlation between the two variables. This could be due to when an image is retweeted it reaches a bigger audience meaning that there will be more individuals seeing the tweet and hence favouriting.

What are the most favorite Breeds?

By grouping the data by the breeds predicted in p1, the top ten most favorited breeds are:

- Golden Retriever
- Labrador Retriever
- Pembroke
- Chihuahua
- Samoyed
- French Bulldog
- Chow
- Pug
- Cocker Spaniel
- Pomeranian

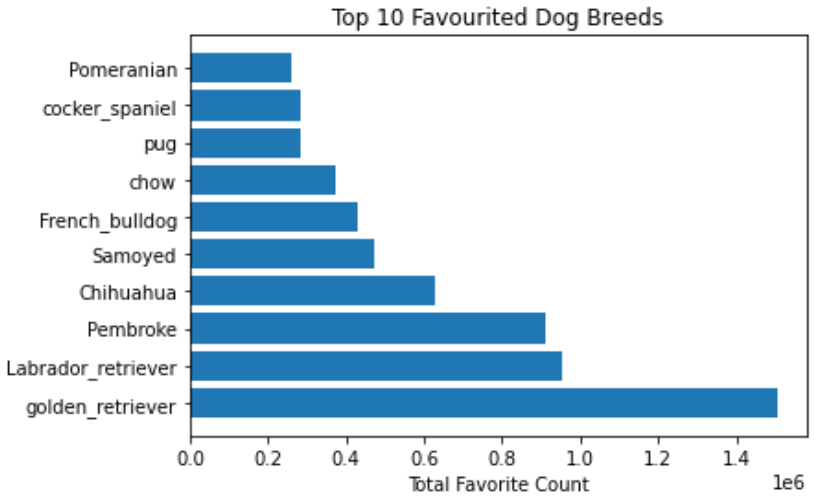

Figure 7: Top 10 Favourite Dog Breeds

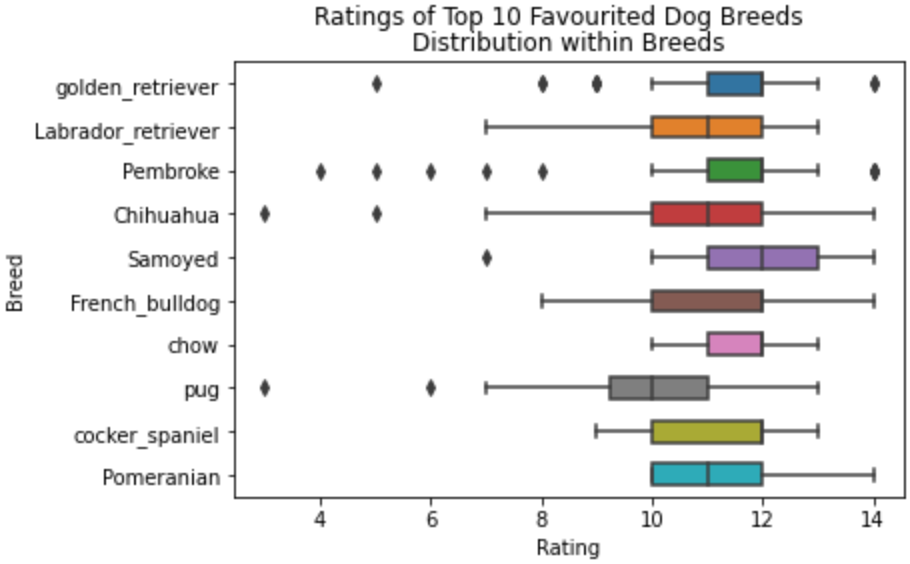

Figure 8: Boxplot showing distribution of ratings for top 10 dog breeds

The top 10 breeds all fall in a rating range of 7 and 14 with several outliers. Our top breed, Golden retriever, had a top rating of 13 (rating of 14 is considered an outlier), however there are several breeds which scored higher. The spread of ratings of the Golden retriever (range of 3) is small compared to some of the other breeds, for example Chihuahua has a range of 7. Overall, across these breeds the ratings are skewed to the left hand side and are typically above 10.

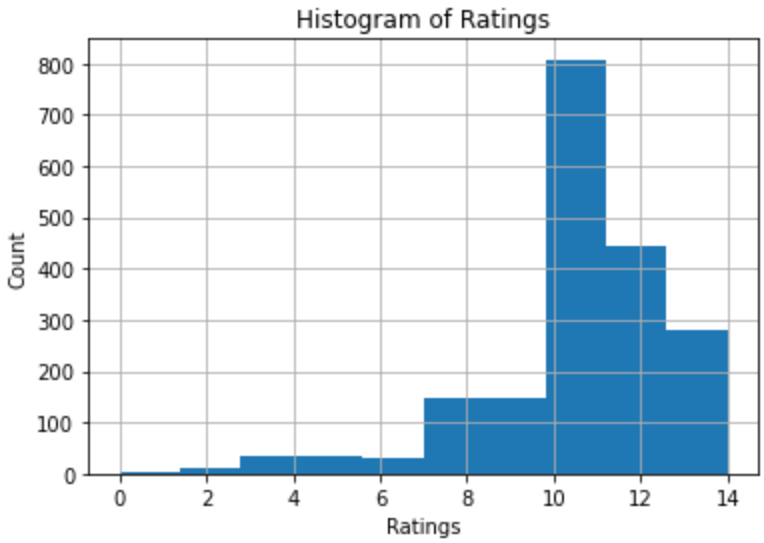

Figure 9: Histogram of Ratings

Ratings across all the tweets are left skewed which is the same as the top 10 breeds. Information provided showed that it isn't unusual that the ratings are above 10. The top 10 favourite breeds follow a similar shape to the data overall.