Analysis of WeRateDogs

We Rate Dogies is a twitter account that rates dogs. It receives images of dogs and rates it. It posts the same on its account, comment on it and post it with the rating they provide for each of them. They always rate the dogs out of 10. At times the ratings go beyond 10 on the numerators side.

For example, 13/10 or maybe 14/10.

It has 6.55M followers and 6853 tweets at the time of writing this report.

Most recent tweet!

About the Data

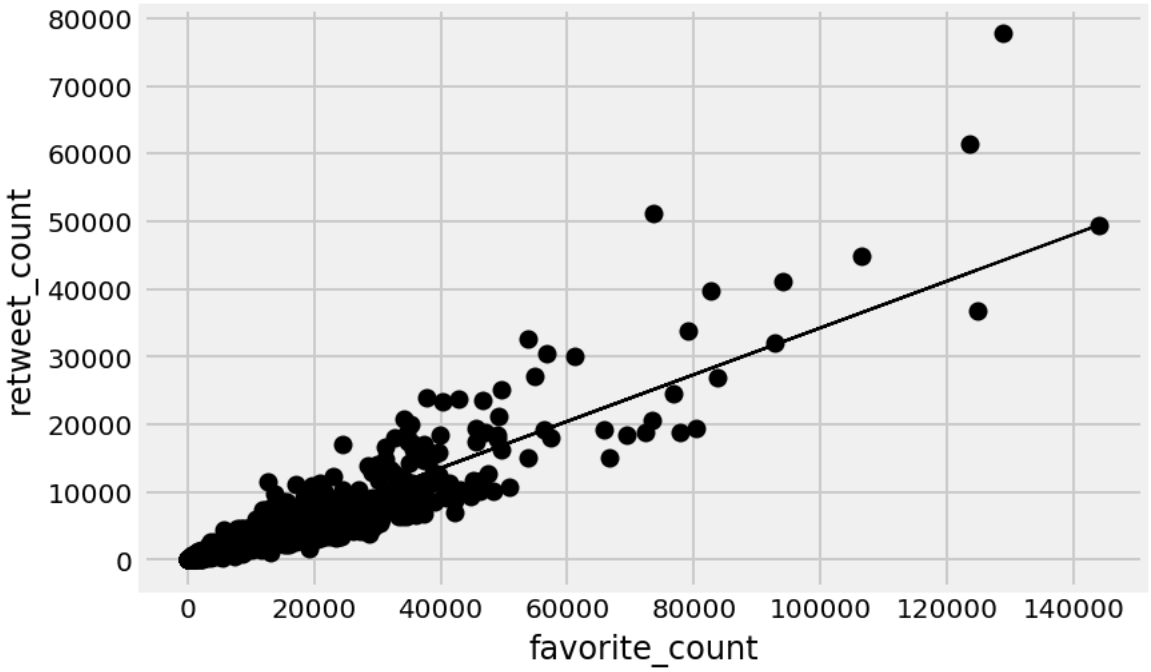
To analyze the tweets, we have used different verified sources.

The first is an archive of the past tweets provided by Udacity via a csv file. The second source was retrived using Twitter API. It fetched further details of the dataset already with us. The third source was also provided by Udacity that contained images predicted via a neural network. (The accuracy of the prediction is not discussed.)

There were 2336 entries initially and after combining I ended up analyzing 2174 tweets from WeRateDogs..

Favorites and Retweets of Dog Stages

I first observed the relationship of favorites and retweets as I wanted to know how these metrics later related to the most popular dog. So a plot was made (shown below) that most dog tweets had more favorites than retweets.



Relationship between the number of favorites and retweets

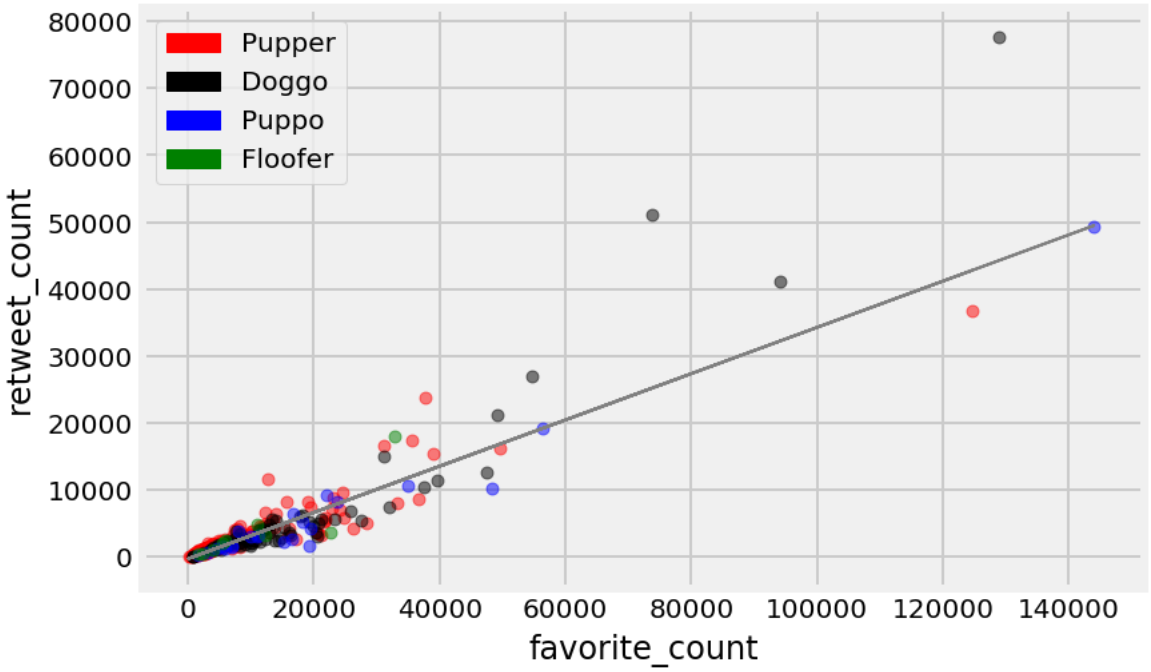
with x-axis as the number of favorites and y-axis as the number of retweets

This makes sense since more people will favorite a tweet instead of taking ownership of a retweet, so this information wasn't unexpected. A line of best fit was made (equation: retweets = (0.35)*favorites + -293) and had a correlation r of about 0.92. It should be noted though, that the data may be more quadratic especially for a high number of retweets and favorites. But this line will be good enough for now since there are not many high valued data point

I plotted a scatterplot of all tweets.

It can be observed that most of the tweets have less than 40,000 favorites. It can also be observed that from the equation for the line of best fit that there tends to be 3 times as many favorites as retweets. However we notice that as the values get larger, the ratio of retweets to favorites get larger too.

On grouping the plot with the dog stages (for those data is available), smaller values stick-close below the original line of best fit.

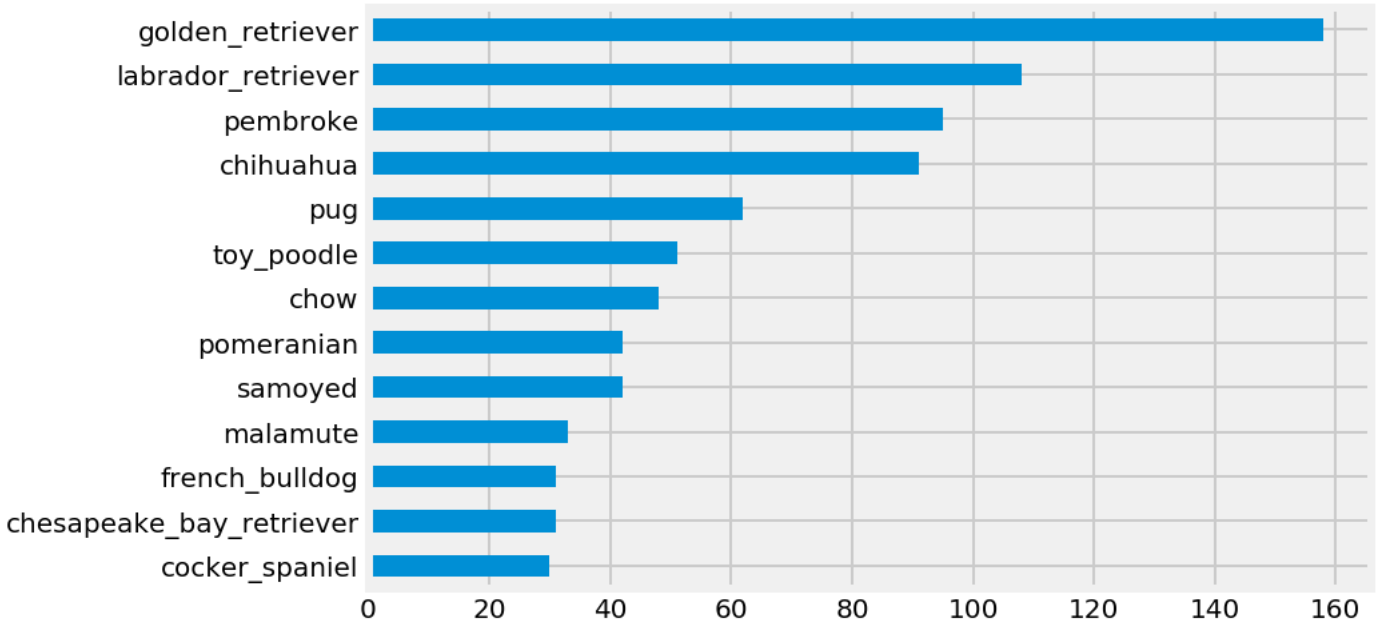


After zooming in, it can be observed that there are more tweets classified as puppers than any other existing classification. Most of the puppers are lower valued and tend to have a higher ratio of retweets to favorites than the line of best fit. Almost every other classification tends to have a lower ratio of retweets to favorites (more favorites) than the puppers classification.The tweets classified withing doggos tend to have higher values as compared to the pupper tweets.

Most Common Dog Breed

Initially, 30 most common dog breeds was plotted. It can be observed that nearly 500 tweets were classified as not a dog (not_dog). This is about 3 times more than the highest numbered actual dog breed. This could be that many tweets from this ratings archive are not dogs, however this seems quite strange. It is possible that the image predictions has misclassified many of the images as not dogs. Next,12 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.

However, concluding too much from that data since must invoke a sense of extra care since, we are relying on the image prediction code to inform us of the dog breed. It is possible that certain dog breeds were harder to predict.

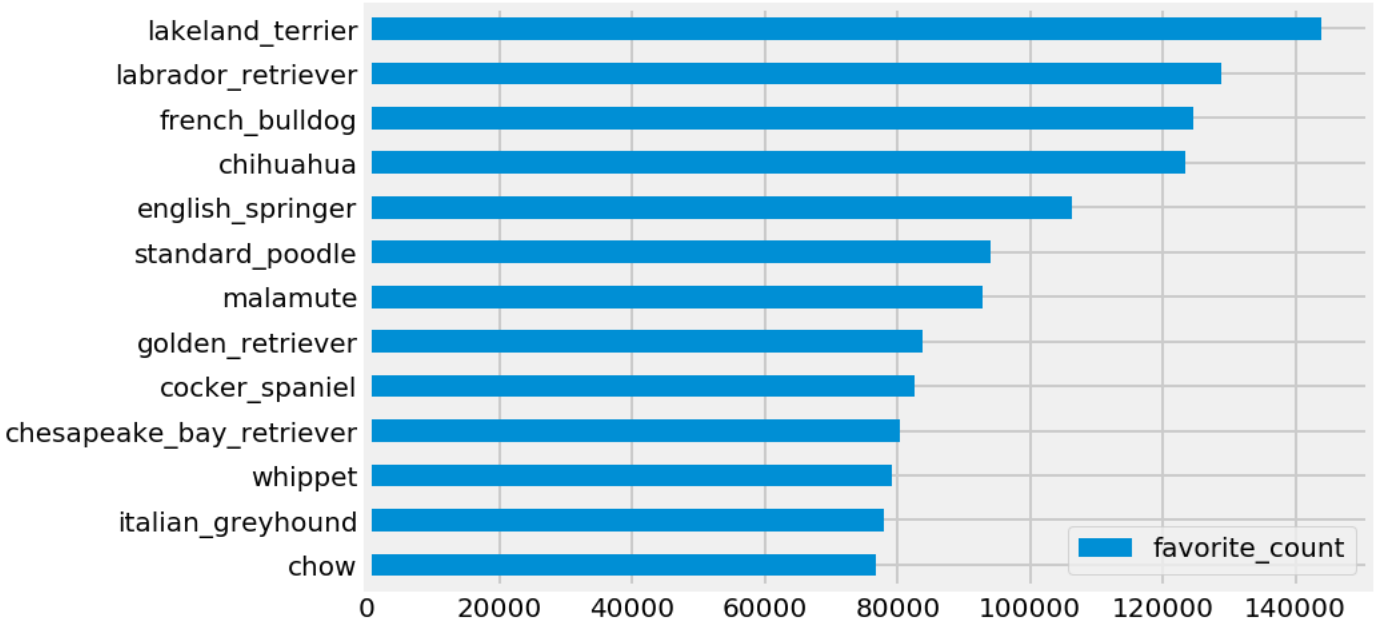


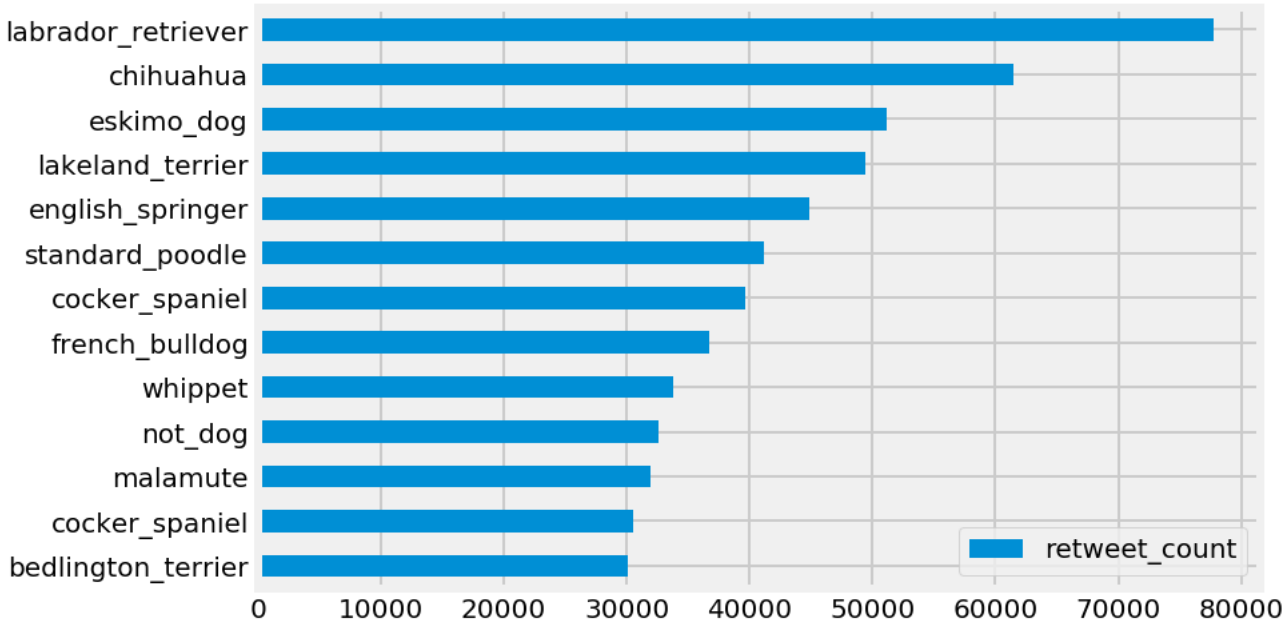
We see that the most popular dog breed is the golden retriever, followed by the labrador retriever, the pembroke (aka the much internet beloved corgi), the chihuahua, the pug, and so on. Those five most common dog breeds seem to me to match what you see on the internet. So I would say that this distribution of dog breeds seems pretty representable of what I would expect.

So going on to our final part we investigate which furry friend is the most popular. And if what we learned from our earlier plotis true, we would expect that most of the most popular dogs to not be a common breed that is tweeted.

Which Breed is more Popular?

So coming equipped with new knowledge, we plot our which dog breed tweet has the most favorites and retweets on average. Remember that these attributes should be correlated and we expect that the most common dogs are unlikely to show up on our list.





On averaging the retweets and favorites for each breed , only the French Bulldog and the Samoyed appear in one of these plots as well as the plot of the 12 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 leonberg, black and tan coonhound, and bouiver des flanders only appear in the top favorite chart and the dog breeds cocker spaniel, samoyed, and great pyreness 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.

Although most of the breeds appear in both the plots, the bedlington terrier is amongst the top. This dog breed is likely the most popular breed among the tweets.