Analysis of WeRateDogs™

WeRateDogs™ is a twitter account that humorously reviews pictures of dogs and give them ratings. Seems simple enough though, WeRateDogs™ has what one might call a different approach to a normal judge. Almost all of this Twitter handles' ratings are over 10/10. These ratings almost always have a denominator of 10. The numerators, however, are almost always greater than 10 (11/10, 12/10, 13/10, etc.).

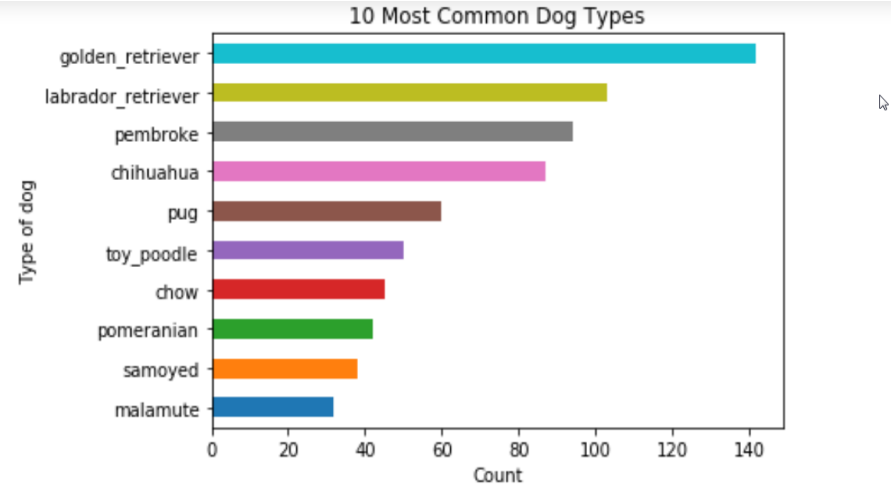
So what does this rating do? Which dog type (breed) is the most common dog in the tweet dataset? What is the most popular dog type in terms of favorites and retweets? What is the most appropriate metric to measure the popularity of dogs? These questions are discussed in the following report.

About the Data

To analyze the tweets from WeRateDogs™, we have used three different sources. The first source is an archive of the past tweets from @dog_rates (https://twitter.com/dog_rates) provided via a CSV from Udacity. The second data source provides us the predicted dog breed in each tweet's image, programmatically determined from a neural network. This was also provided by Udacity. The third source is from the Twitter API used to retrieve more information about the popularity of tweeted dogs. By gathering, accessing, and analyzing the data from these three sources, we ended up figuring out three insights and visualizations as followed:

Most Common Dog Types

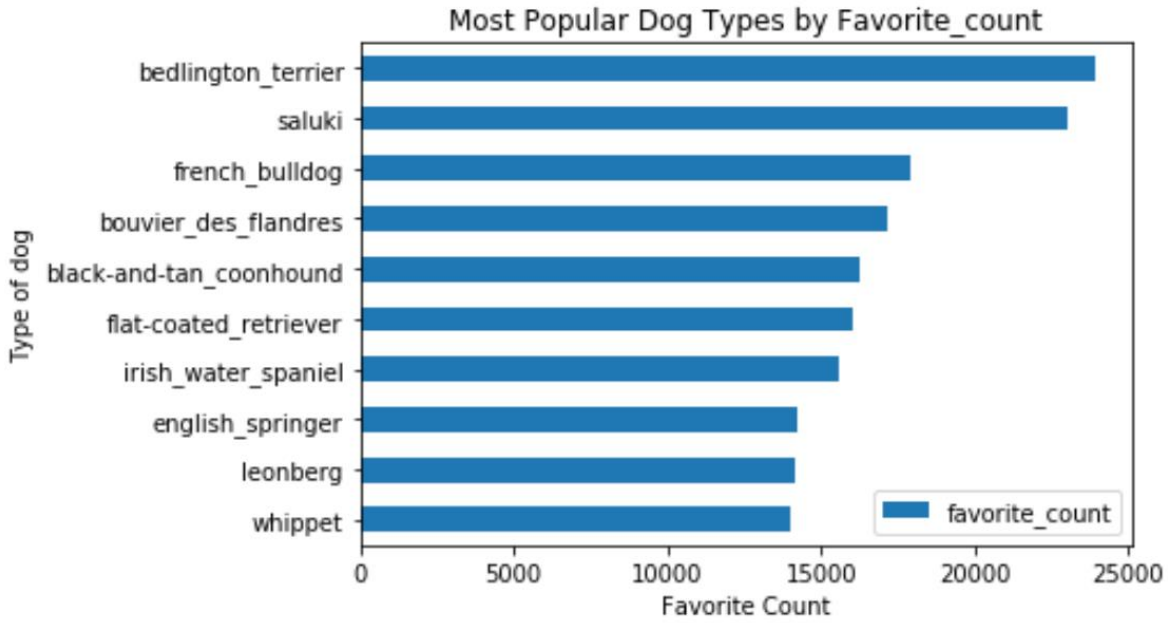
The 10 most common dog breeds are plotted in a bar chart. The most common dog breed in this dataset is the golden retriever with over 150 tweets. The next four most common breeds, in descending order, are the labrador retrieve, the pembroke (corgi), chihuahua, and the pug.

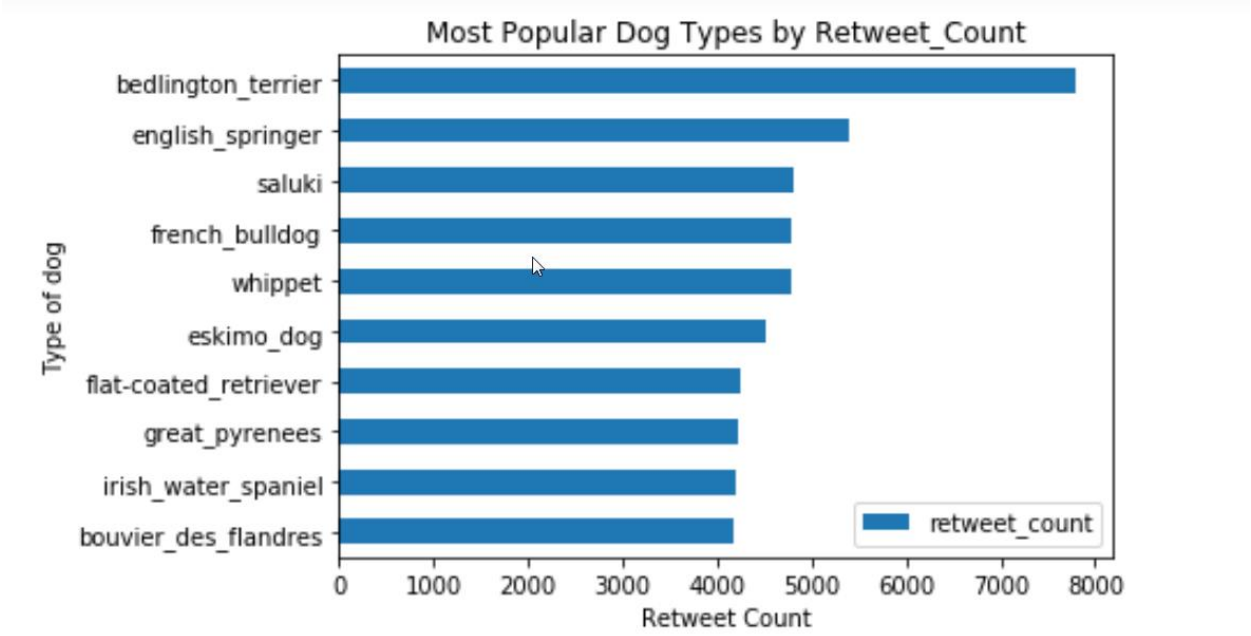


This is an example of the most common dog type – Golden Retriever.

Most Popular Dog Types

The 10 most common dog breeds are plotted in a bar chart. The most common dog breed in this dataset is the golden retriever with over 150 tweets. The next four most common breeds, in descending order, are the labrador retrieve, the pembroke (corgi), chihuahua, and the pug.





We averaged the retweets and favorites for each breed and plotted the top 10 highest breeds for each factor. None of the top 10 most common dog types appears in these two plots. This suggests that the most common dog types on average didn't have the highest retweets or favorites. This might make sense as perhaps the most retweets and favorites are on less common breeds since they stick out. This relationship between common and popularity of dog breed could be interesting to investigate in the future.

Also we notice that most of the breeds appear in both graphs, except for the dog breeds leonberg and black-and-tan coonhound appearing in the favorite chart only and the dog breeds eskimo dog and great pyreness appearing in the retweet chart only. The highly consistent results between these two factors indicate that a positive relationship may exist between favorite_count and retweet_count. This will be discussed further in the following section - Insight three & Visualization.

We see that although most breeds appear in both charts, the bedlington terrier is consistently standing at the top of both charts, showing that this dog breed the most popular dog breed. It would be interesting to see how this differs with the different dog stages (pupper, floofer, etc.)
We did the same analysis for different stages. The results show that for doggos, eskimo_dog is the most popular type, with both favorite_count and retweet_count three times higher than the second popular type – Labrador_retriever. For floofer, Samoyed and chow are the most popular types. French_bulldog is the most popular dog type for pupper, and Lakeland_terrier for puppo.

Measures of Popularity – Favorites vs. Retweets

A linear regression was run. The scatterplot of all tweets as well as a linear line of best are plotted. The high R-value and significant p-value reveal that the linear regression fit the data well. The scope value of 0.345 shows that favorite count is approximately three times as many as retweet_count. However, we notice that as the values get larger, the ratio of retweets to favorites get larger (further above the line of best fit).