# Analysis of WeRateDogs

WeRateDogs™ is a twitter account that tweets out pictures of (mostly) dogs with their associated "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. Why are so many dogs so "over-rated"? There could be plenty of reasons, but WeRateDogs™ does have at least one reason; "they're good dogs Brent" (https://twitter.com/dog_rates/status/775410014383026176).

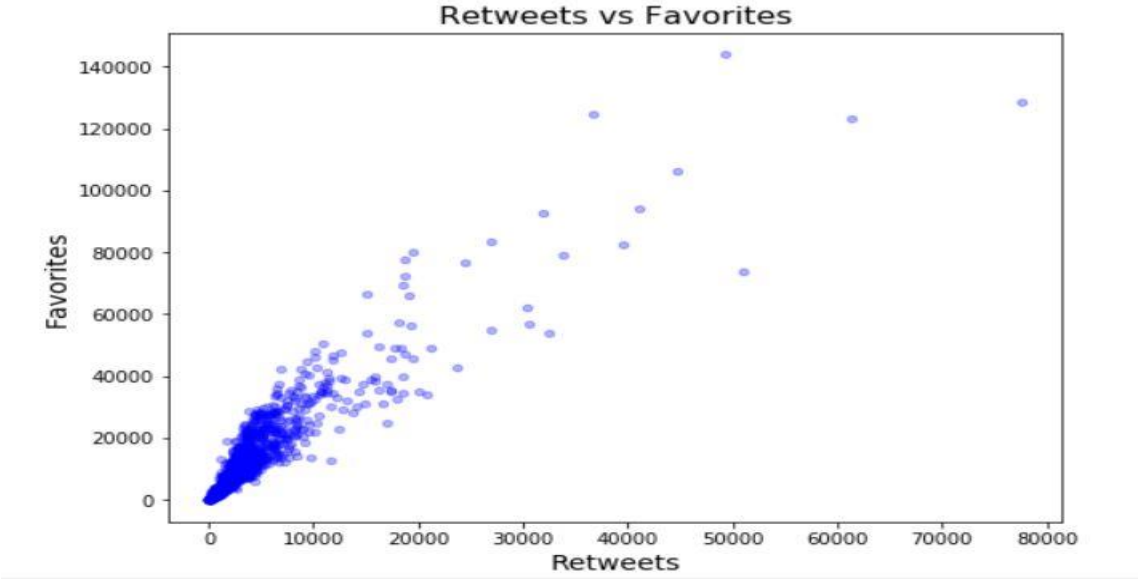
## About the Data

To analyse 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 source is from the Twitter API used to retrieve more information about the tweets like number of each was retweeted. The third 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. This third source is particularly important as it was not determined how accurate the predictions were.

Combining this data from these sources, we ended up analysing 2174 tweets from WeRate-Dogs. These tweets were from before August 2017 and were used to analyse the relationship between retweets and favorites amongst the different "dog stage" (floofer, pupper, doggo, etc.), the distribution of dog breeds in the tweets, and the popularity of the dog breeds based on retweets and favorites.

## Favorites vs Retweets

I wanted to observe the relationship of favorites and retweets. So a plot was made (shown below) that most dog tweets had more favorites than retweets.



Relationship between the number of retweets and favorites

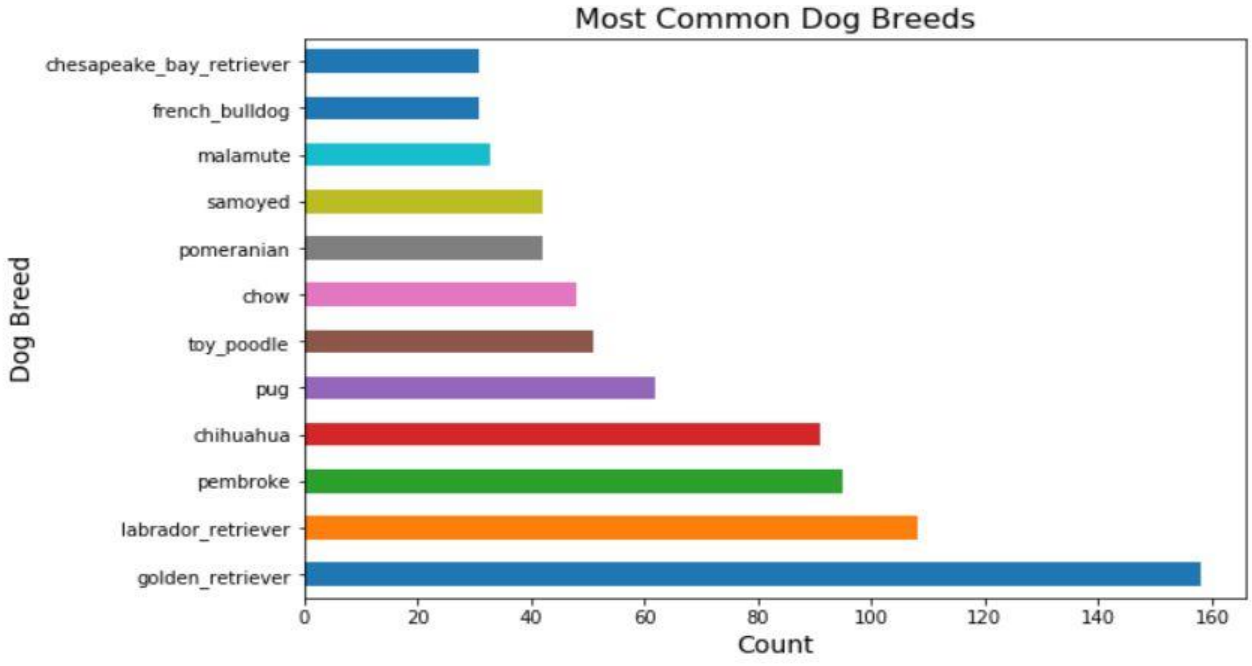
with x-axis as the number of retweets and y-axis as the number of favorites

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 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 points.

## Most Tweeted Dog Breed

Coming from our previous graph and observing that perhaps popularity of a tweet is related to its commonality amongst its fellow tweets. When sorting for the most common dog breeds, I found that nearly 500 tweets were not labelled as a dog breed. This might be perplexing. Were so many tweets not of dogs?

Though it is possible some of these tweets weren't dogs, it is more likely that the data for the dog breeds is not accurate. This could come from the fact the neural network did not recognize some breeds. Maybe some dog breeds were harder to classify for the neural network. But since we don't have access to the neural network and I think it would be a better use of our time to use what we have, we'll ignore this issue. We'll just keep this in mind as we investigate further and maybe revisit this another day to improve the data's accuracy. So for now, we'll just sweep this small issue under the rug (which I never do while cleaning the house of course).

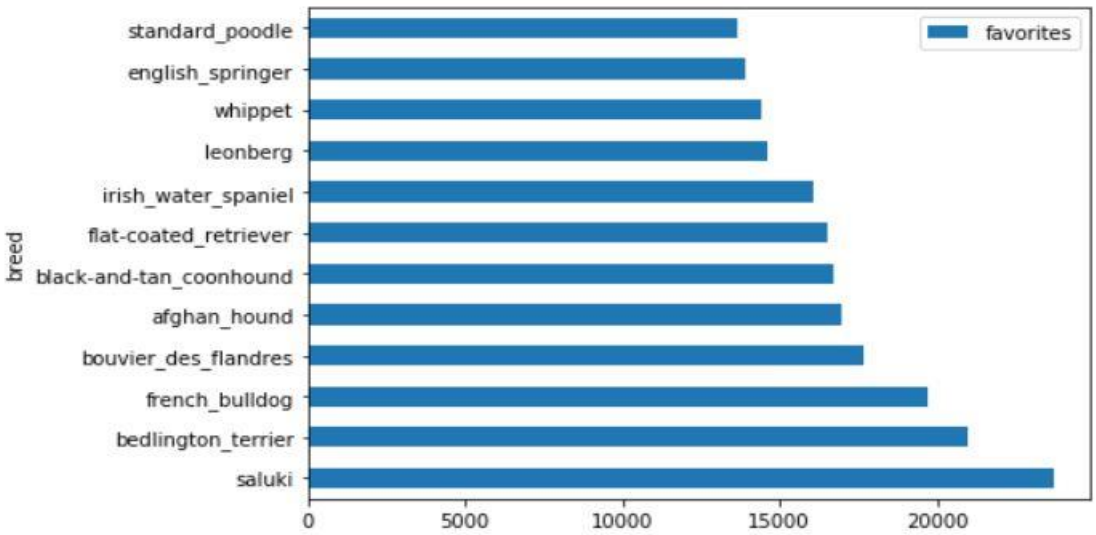


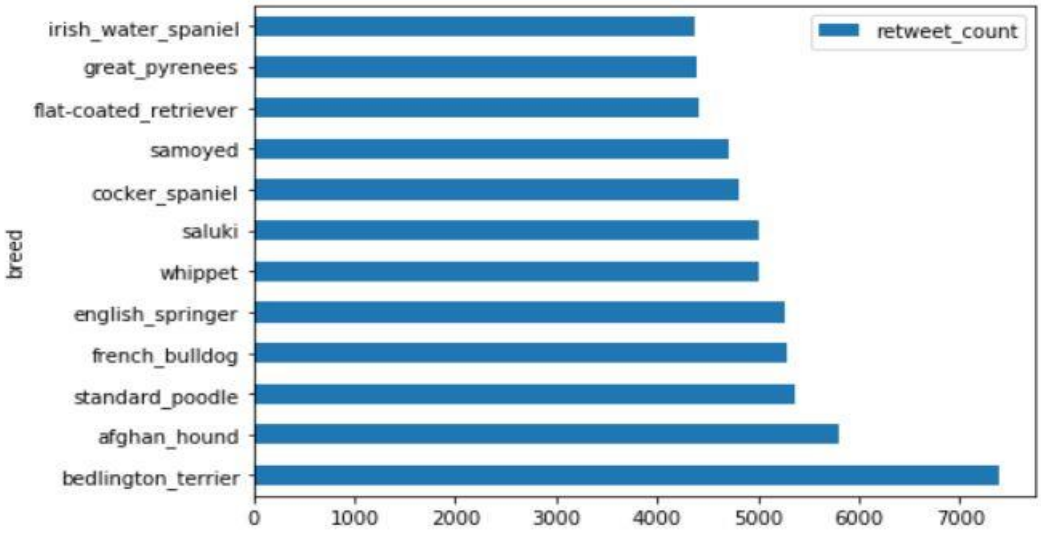
We see that the most popular dog breed is the golden retriever, followed by the labrador retriever, the pembroke (aka the much internet beloved corgi), the chihuahua, the pug, and so on. Those five most common dog breeds seem to me to match what you see on the internet. So I would say that this distribution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

## Most Popular Breed

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.



What magnificent graphs! So the first thing I wanted to know was that if any of our common breeds were on either chart. And we do! Wait, what? Our predition was wrong then? Well maybe not. The only dog breed from the most common chart that appears in either of the favorites or retweets chart is the french bulldog. The french bulldog is only the 10th most common dog and it only appears on the favorite chart. So despiste this exception, most common dog breeds are not the most retweeted or favorited. So it looks we might be right that the internet loves those rare puppos.

We also notice that nine dog breeds are found in both the top retweets and favorites charts. This matches well with what we saw observed in that retweets and favorites are positively correlated with one another. However, there is one dog breed that is 1st on the average number of retweets and 2nd on the average number of favorites. And that is the noble bedlington terrier! And looking at how common the dog breed is in our data, we find that only 6 tweets had the bedlington terrier making it the 70th most common dog breed in our data out of
114 different dog breeds.