

# We Rate Dogs: Analysis and Insights from Final Data.

by Yukti Sareen, Apr 2019.

The data for analysis was gathered from three different sources, assessed and cleaned as described in the data wrangling report. The result of that is two sets of data that can be looked into more closely to see patterns of popularity, which can then be used for further knowledge discovery.

The rating system formulated in this twitter account is unique and hilarious, but somehow seems fitting once you look at the pictures posted alongside the tweets. How can we just rate them on 10?

Based on the variety of breeds available, I wanted to look at different statistics by breeds to get a better understanding of the feeling among the community of users that are active on this account.

## 1. Most popular breeds by number of tweets

We have the breeds predictions that we can use to count the number of tweets grouped by a breed to have a view on the popularity of different dogs. We use a straight forward vertical bar chart to represent this information.

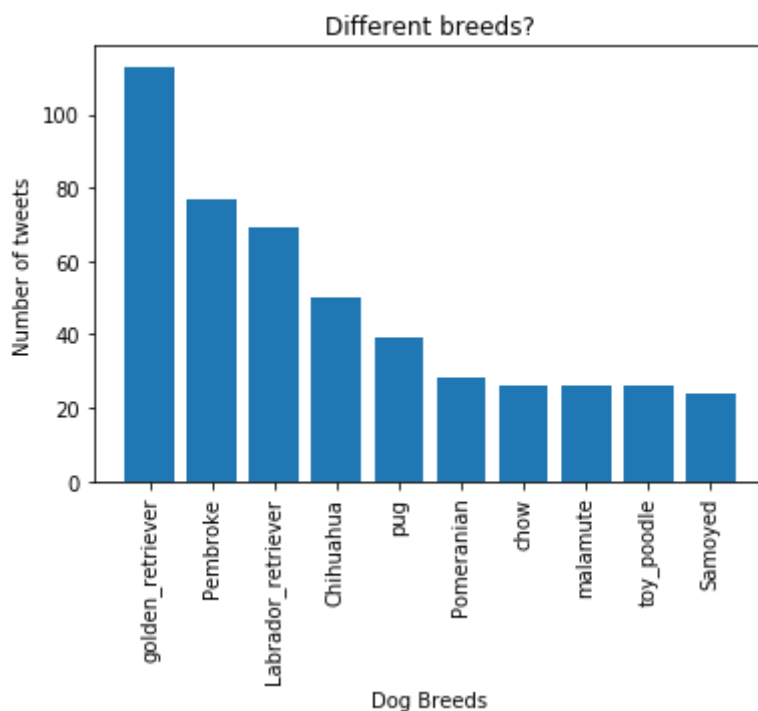


Fig. 1. Popular breeds

Here we can see that the top 5 on this list are Golden Retriever, Labrador, Pembroke, Chihuahua and of course, the Pug. It is noticeable that Golden Retriever leads the way comprehensively. We can see how good this group holds in our subsequent analysis.

## 2. Which breeds have the most Favorites (or likes) on their tweets?

After cleaning of the datasets and merging, we find that the favorites (and retweets) information are in a different dataset to the breeds information (predictions). Hence, we do a merge and store the results in a temporary data frame. We didn't merge everything together during the step as it seemed logical to keep the image predictions data out of the basic tweet information.

Also, here we only consider those rows where the rating\_numerator is less than 20 - to eliminate wild outliers from skewing the average.

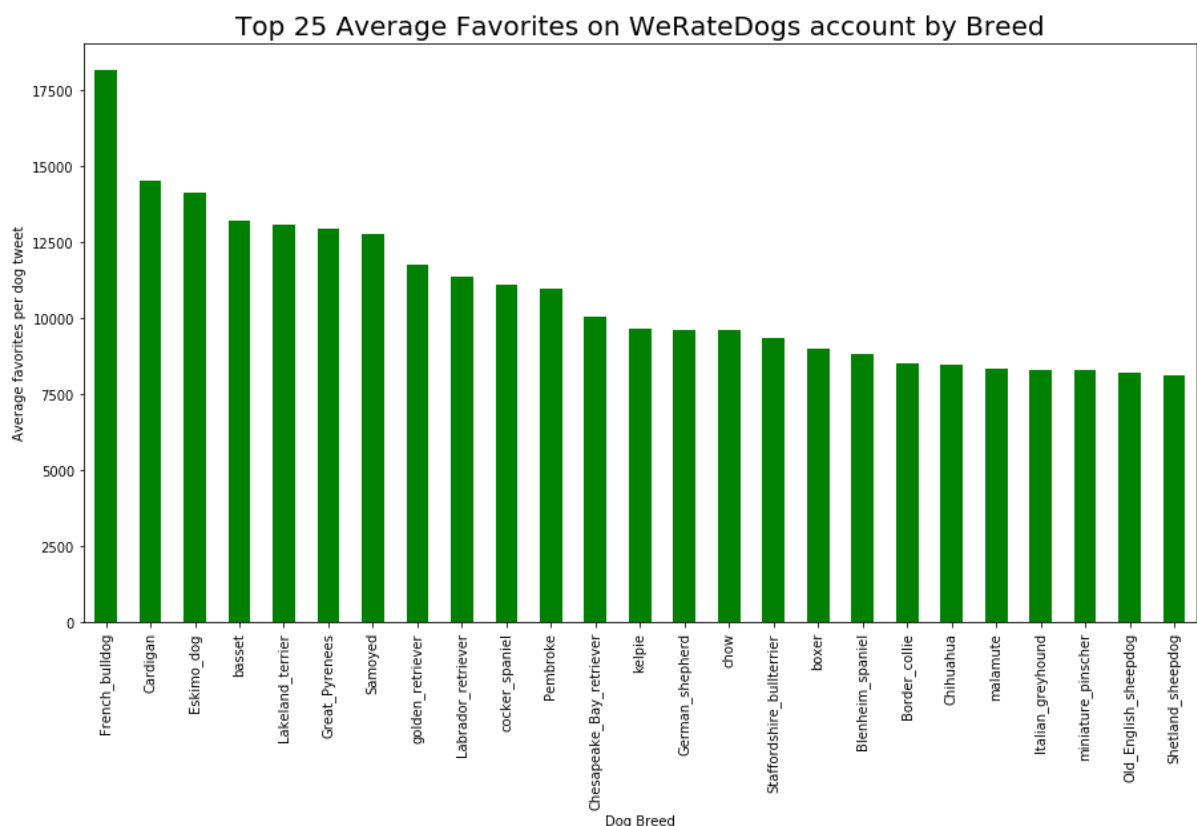


Fig. 2. Top favorites for Breeds.

Here we see the French Bulldog leading the way when it wasn't in the top 5 on number of tweets. The golden retriever, labrador, chihuahua are still around in this list.

## 3. Which breeds received the highest average ratings?

The most interesting thing about this account and the data is the ratings system, and more specifically the rating\_numerator. So, it is almost that we have to see how the breeds line up based on the ratings.

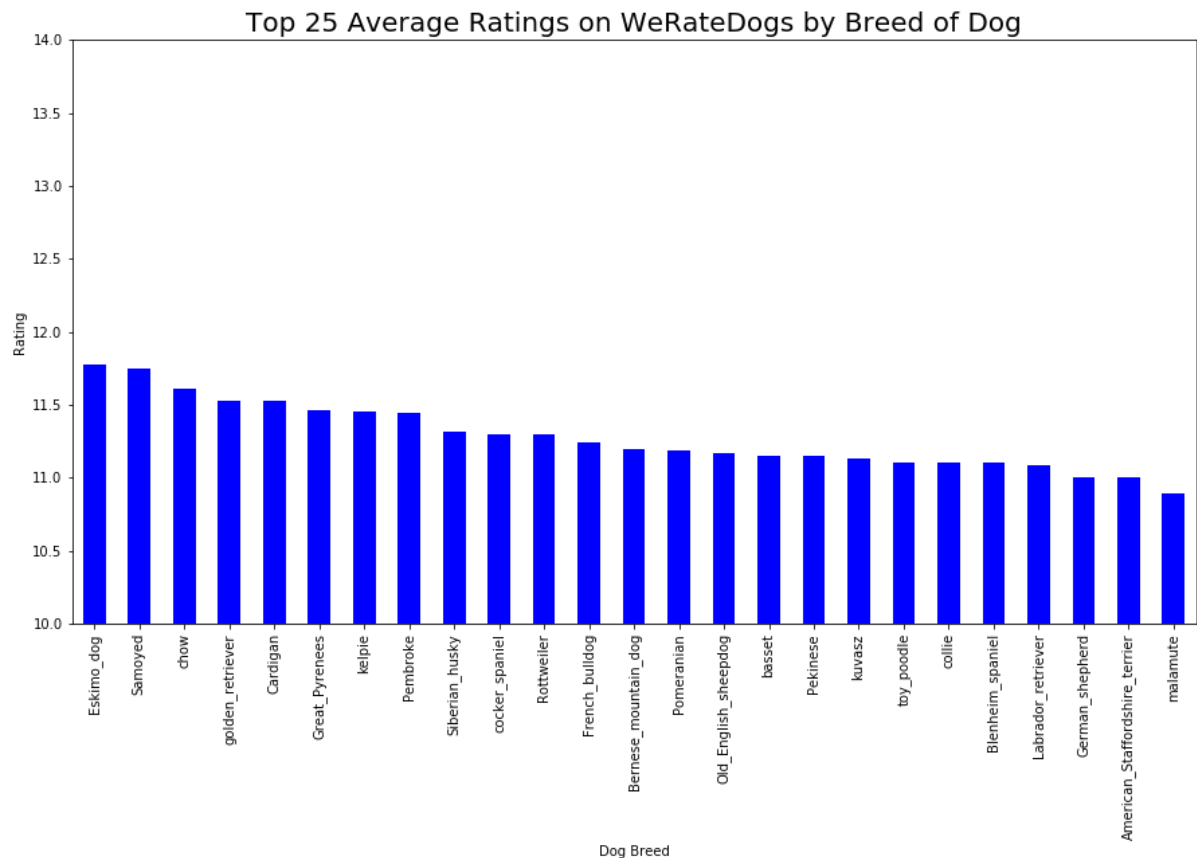


Fig. 3. Highest avg ratings

We notice that the eskimo dog is high on both the favorites and the ratings chart. Also, we see that the French Bulldog while being heavily liked, also has high ratings average.

These charts and insights seem a good place to stop and look at how else we can use the data in, say suggesting a friend or our own family in going for a specific kind. We could definitely use information like average size of these dogs and preference based on locations where they live.

Working with this data is unlike anything else due to the *cute-ness* factor involved throughout the dataset.