

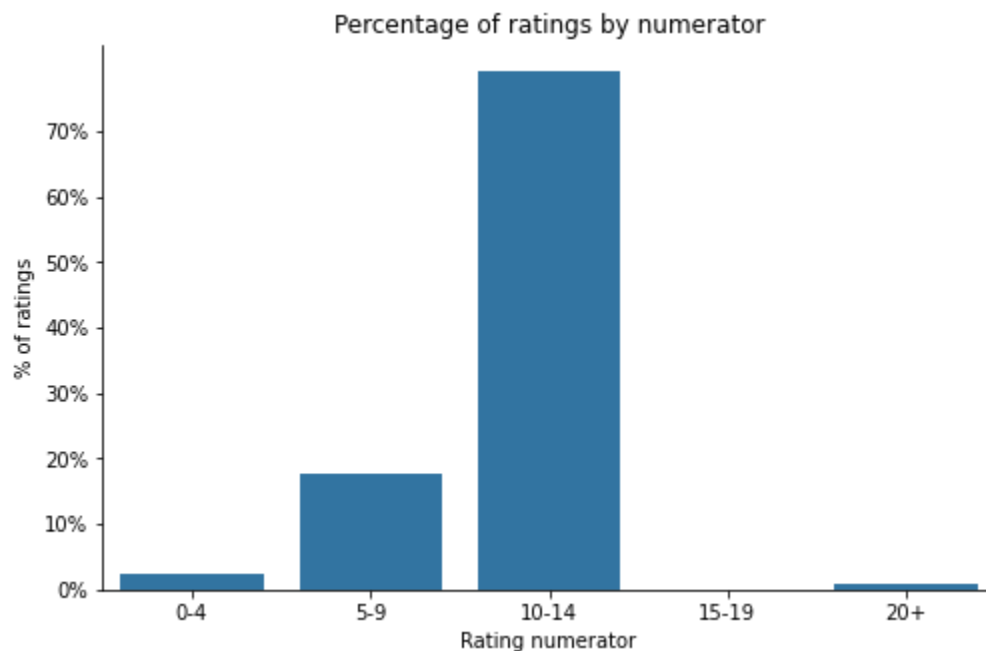
Insights form WeRateDogs Tweet archive

Michael Smales

August 2022

We rate dogs is a popular twitter feed which rates dogs. It uses a unique rating system out of 10, where the numerator is often greater than 10. In this analysis, I combined an archive of November 2015 – August 2017 tweets with image predictions of the dogs to generate a few insights about the dogs on the website.

Firstly, I looked at the distribution of ratings given by the feed. Generally, ratings are given a denominator of 10, so I looked at the numerator only.

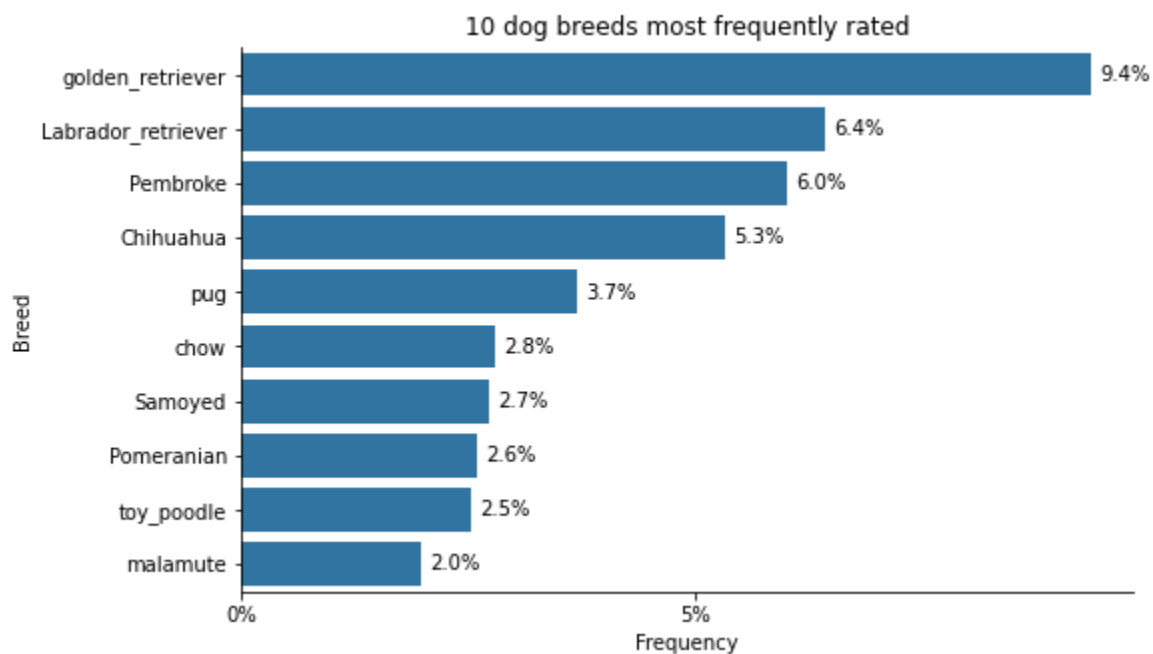


More than 70% of the ratings are in the range of 10/10 to 14/10. There are a very small proportion which have ratings over 20, which are often given for photos of groups of dogs. Surprisingly, given the positively-skewed rating system, the account doesn't always give positive feedback on dogs: around 20% of the ratings are in the 5-9 range, and around 2% of submissions are in the 0-4 range.

The second area I looked into with the data was popular dog names. The WeRateDogs account usually refers to the dog by name in the tweet, so the tweet archive can be used as a proxy for popular dog names. This assumes that the dogs submitted for rating by their owners are representative of the general population of dogs. The 10 dog names which were rated with the highest frequency are below. Charlie and Lucy tie for first place, followed by Cooper and Oliver tied for third place.

Name	Frequency
Charlie	11
Lucy	11
Cooper	10
Oliver	10
Penny	9
Tucker	9
Lola	8
Sadie	8
Winston	8
Daisy	7

Thirdly, I looked at the most common dog breeds included in the ratings. As well as providing the tweet archive data, the Udacity team ran the images of the dogs through an image classification algorithm which guessed the dog's breed. By joining the tweets with the dog breed predictions, a count of dog breed popularity was generated. **Caveat:** For this analysis, the algorithm made 3 guesses of the dog's breed, and assigned a likelihood of the guess being correct for each. I simply used the first guess, regardless of the probability of it being correct. This approach could be refined in the future.



Only the top 10 breeds by popularity are included in this chart. Golden retrievers and Labrador retrievers are the most popular breeds, accounting for over 15% of the dogs rated.

Finally, I looked at which breeds are most likely to be rated highly by WeRateDogs. Breeds were ranked by the mean rating (numerator only). Breeds were filtered-out if they received fewer than five ratings,

to avoid the poll being skewed by a single high rating. **Caveat:** In some cases, denominators greater than 10 are used by WeRateDogs. In future work these could be adjusted for by calculating the rating as a decimal.

Dog breed	Mean rating (numerator)	Rating frequency
Soft-coated wheaten terrier	26.7	10
West Highland white terrier	15.6	14
Great Pyrenees	14.9	14
Borzoi	14.6	8
Labrador retriever	13.9	94
Siberian husky	13.3	20
Pomeranian	12.9	38
Golden retriever	12.3	137
Border terrier	12.1	7
Kuvasz	12.0	15

Soft-coated wheaten terriers are rated the most highly on average, with 26.7/10.

Credits: Thanks to Udacity and WeRateDogs for providing the partially-wrangled tweet archive data and to Udacity for providing the dog image prediction information.