

## Brief insights and visualizations from WeRateDogs Twitter data

WeRateDogs is the name of a Twitter account that started on November 15, 2015. It asks people to send photos of their dogs, then tweets selected photos with a rating and a humorous comment. The rating is usually on a scale of one to 10, but dogs are often given ratings that are outside of this scale, such as “15/10.” Dogs are also sometimes classified in tweets as a “doggo”, “floofer,” “pupper” or “puppo.” These loosely refer to stages in a dog’s life.

As of August 2021, WeRateDogs has 9 million followers on Twitter. There are also WeRateDogs accounts on Instagram and Facebook.

For this project, I had three sets of data from WeRateDogs from November 15, 2015 to August 1, 2017:

**twit\_archive\_clean**, which is the WeRateDogs Twitter archive. It contains the tweet id (which is used to refer to a specific tweet), the text of the tweet, the numerator and denominator of that dog’s rating, the dog’s name (if available), and if the dog is classified in that tweet as a doggo, floofer, pupper, puppo or a combination of those classifications.

**image\_pred\_clean**, which contains the top three tweet image predictions (i.e., what breed of dog (or another object, animal, etc.) for each tweet according to a neural network. Image\_pred\_clean also contains tweet id and the image URL.

**tweets\_clean**, which I downloaded from Udacity because I was unable to get a Twitter API key. This dataset includes tweet id, the timestamp of the tweet, the text of the tweet as well as each tweet’s retweet count and favorite (“like”) count.

Each dataset has roughly 2,000 rows.

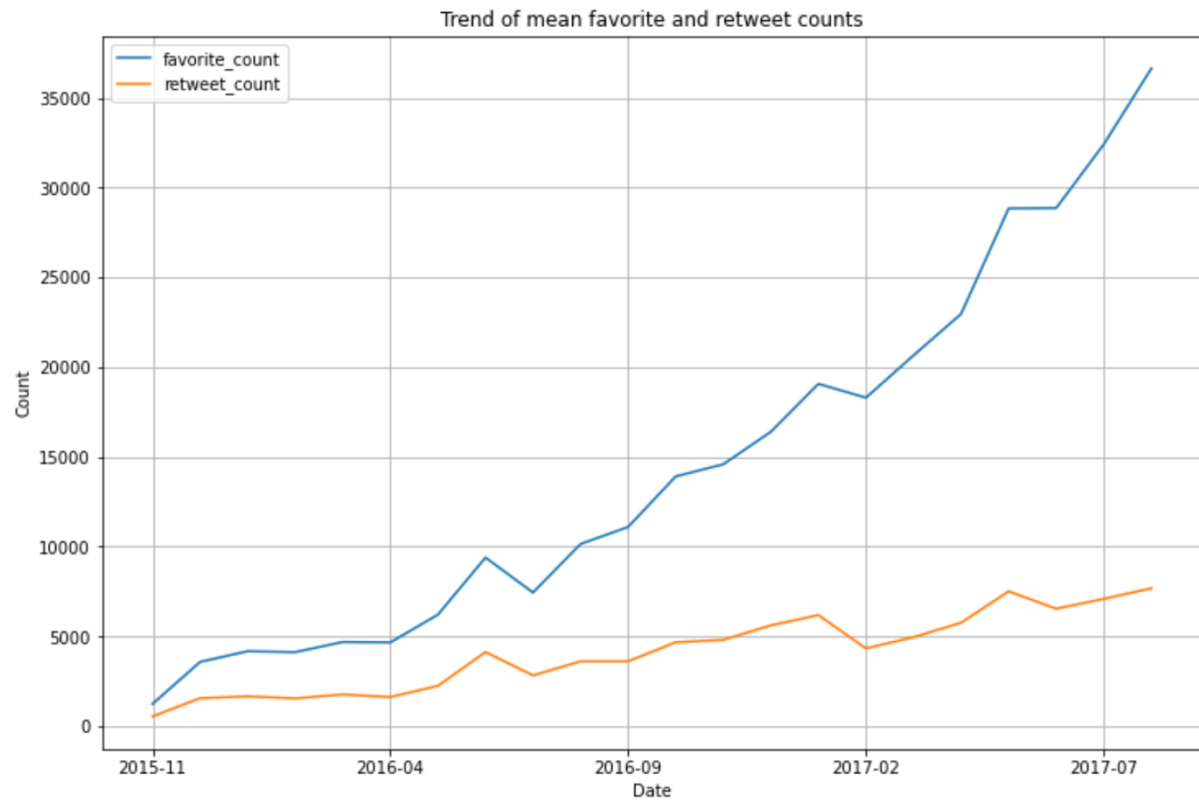
Given the limited amount of data that was available, I kept my analysis relatively straightforward after cleansing and tidying the datasets. (See wrangle\_report.pdf for a description of my wrangling efforts.)

First, I wanted to see the trend between November 2015-August 2017 for the mean favorite count and mean retweet count. I suspected the mean favorite count and mean retweet count increased over time but wanted to see whether one value increased more quickly than the other.

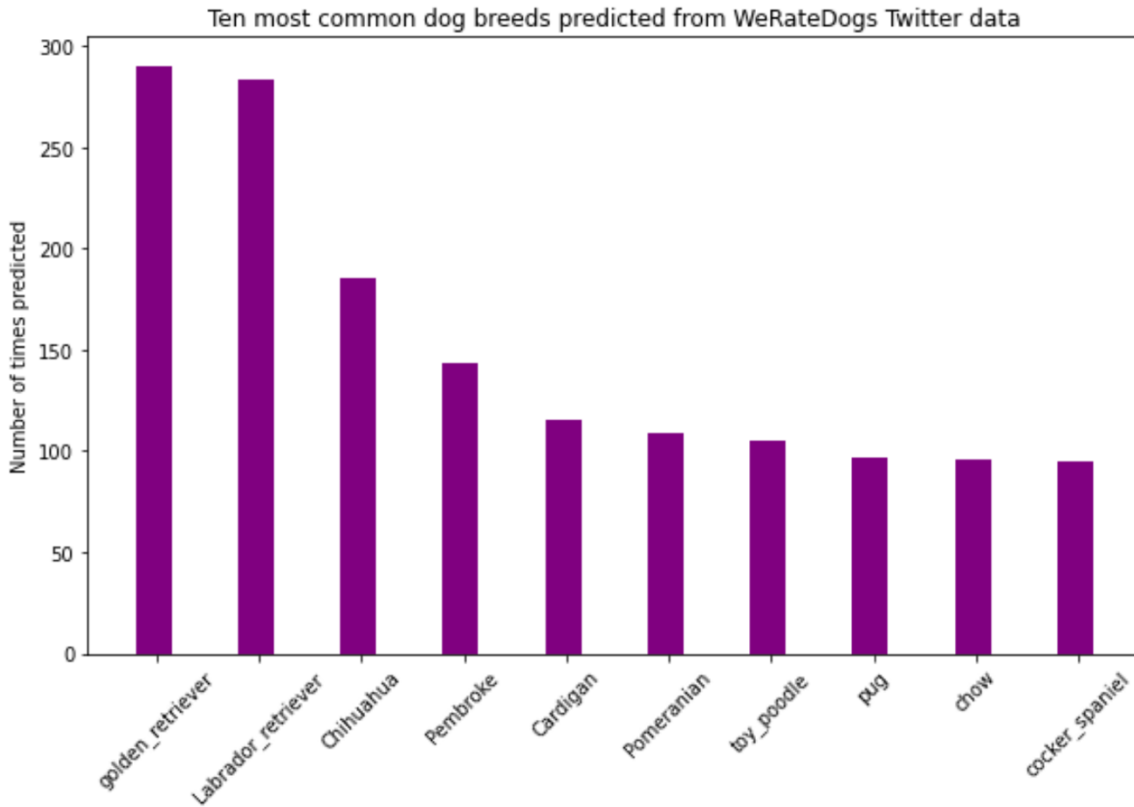
Second, I wanted to know the top 10 dog breeds predicted by the neural network in image\_pred\_clean.

Finally, I wanted to know the top 10 dog names that appeared in the tweets.

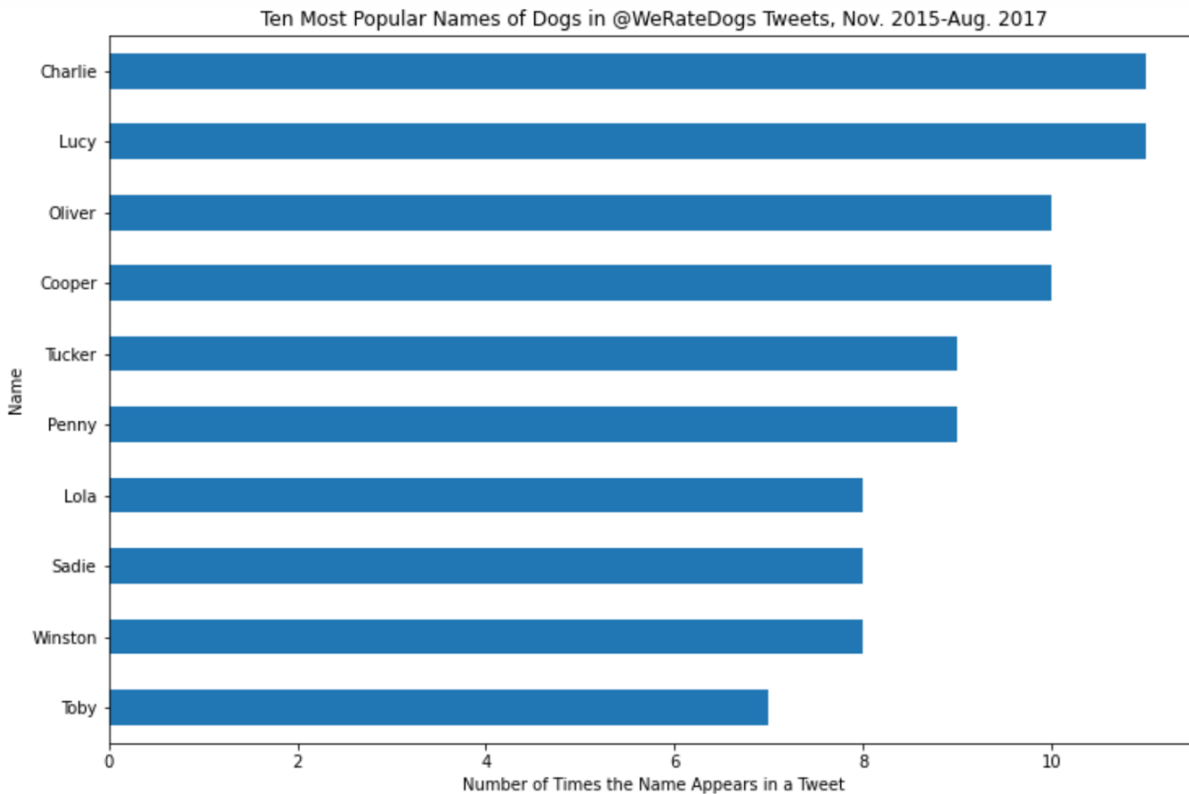
Here's what my preliminary analysis showed so far:



Surprisingly, the monthly mean favorite count increased over time more quickly than the monthly mean retweet count. This is also reflected in the results from the `describe()` method, which indicates a higher mean favorite count (mean of all values, not by month) compared with the mean retweet count (again, mean of all values, not by month). This suggests that people who view WeRateDogs tweets between November 2015 and August 2017 were much more likely to “like” a tweet than they were to retweet it.



Based on the three predictions made by the algorithm for each image analyzed, the two most predicted dog breeds – the golden retriever and the Labrador retriever -- are also two of the most popular dogs in the U.S. (This is according to the American Kennel Club). Without knowing more about the algorithm, the frequency with which golden retriever and Labrador retriever are predicted (and assumed to be predicted correctly) could also reflect that the people who submit dog photos to WeRateDogs tend to submit photos of retrievers or that the people at WeRateDogs who pick photos to tweet and rate tend to pick photos of golden retrievers and Labrador retrievers.



Interesting, the 10 most popular dog names that appeared in WeRateDogs tweets are names that would also be suitable for people, such as Charlie and Lucy. “Traditional” dog names such as Fluffy, Rover, Fido or Snoopy are nowhere on this list. Also, more names in this list (Charlie, Cooper, Oliver, Tucker, Winston and Toby) are male names, suggesting that of the dogs whose names are mentioned in tweets, more are male dogs.