Analysis and Insights of WeRateDog Twitter Archive
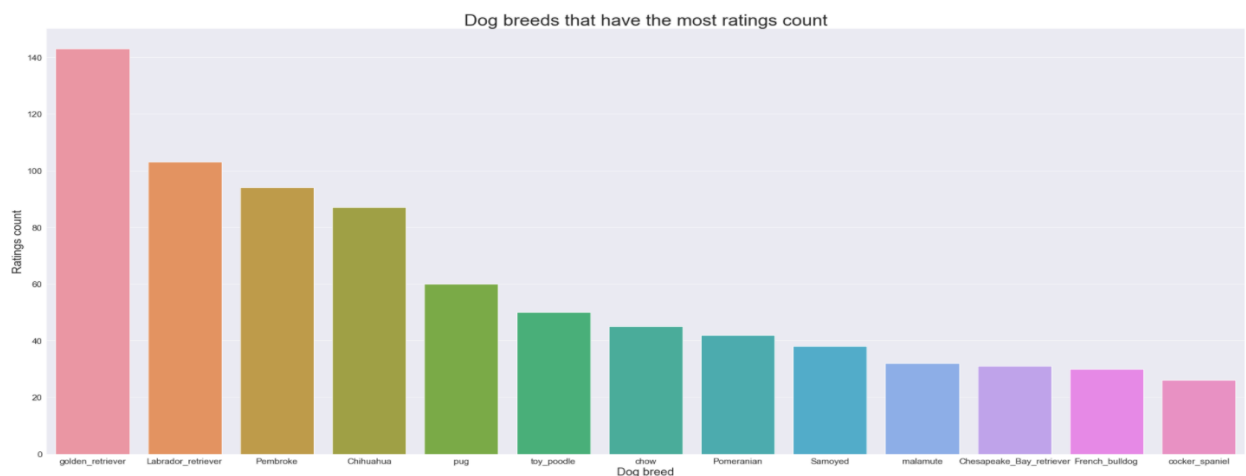
By: Mohammed Hani

Introduction

WeRateDogs is a Twitter account that rates people's dogs with humorous comments. These ratings almost always have a denominator of 10, however, numerators can be any number, even greater than 10. *WeRateDogs* admits that all dogs deserve 10 over 10 and sometimes greater than that "*They're good dogs Brent*".

So in the following discussion we will gain some insights about the data presented in the *WeRateDogs* account and try to answer some questions such as : What is the most common dog breed? What is the relationship between the retweet counts and the dog ratings? What is the trend of tweeting on *WeRateDogs* account from 2015 till 2017? What is the distribution of the rating of dogs and what dogs have the maximum and minimum rating value?
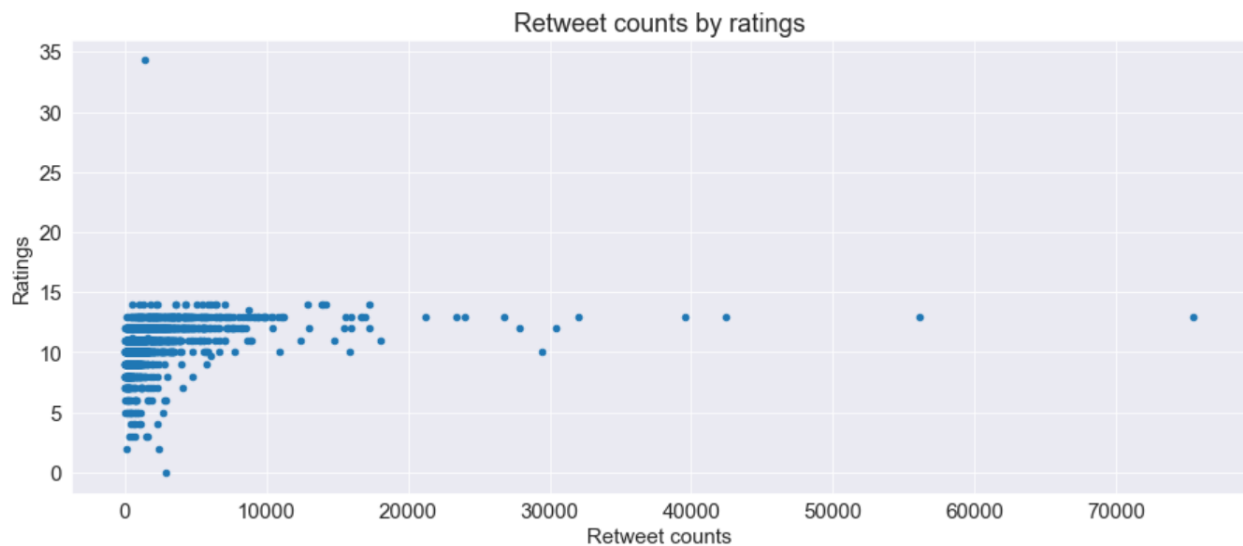
The most common dog breed

According to the figure below, Golden retriever is the most common dog breed in the archive with about 143 ratings from total about 1600

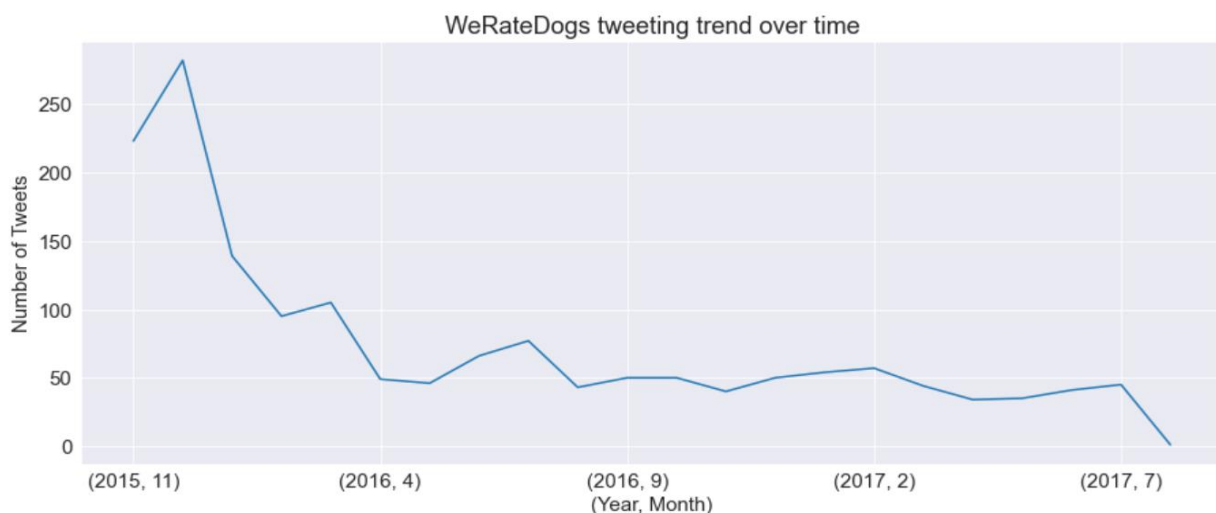Dog breeds that have the most ratings count

Retweet counts

According to the scatterplot below, there is no clear relationship between the ratings of the dogs and the count of retweets that the pictures of the dogs get.
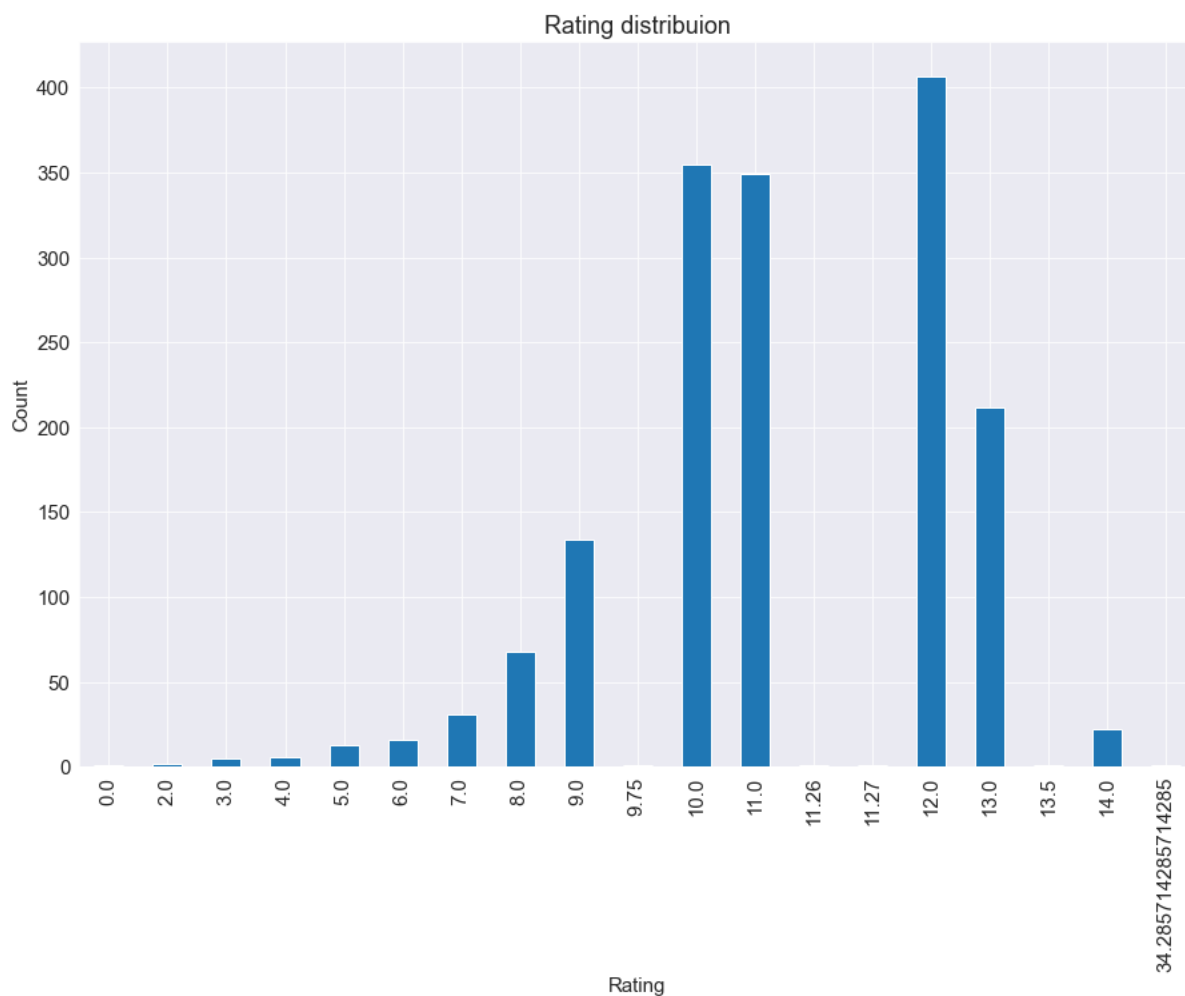

Tweeting trend over time

The trend of tweeting in *WeRateDogs* account is declining sharply as demonstrated in the graph below as it was about 260 tweets in November 2015, but this number is declining until July 2017 as it is being on the verge of being zero.

Rating distribution

Most of the dogs are rated 12/10 (about 407 out of 1,626 records after cleaning). The rating in this dataset is given for each dog, as some tweets are giving rank value accumulated for all dogs (not per each dog)The highest rating that any dog earned is 34.29/10, which is for the dog breed of Golden Retriever, and the minimum value is given to the American Staffordshire Terrier breed.
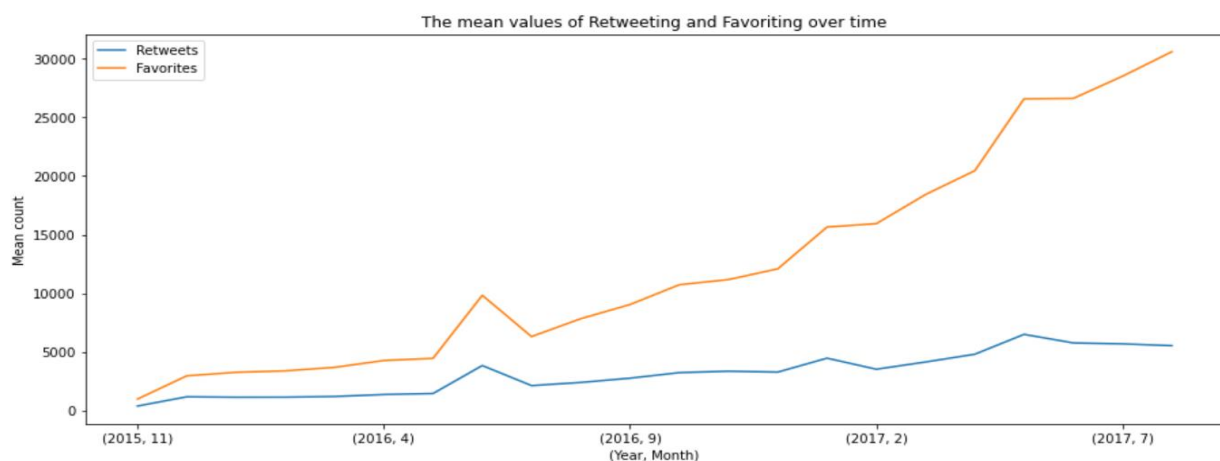
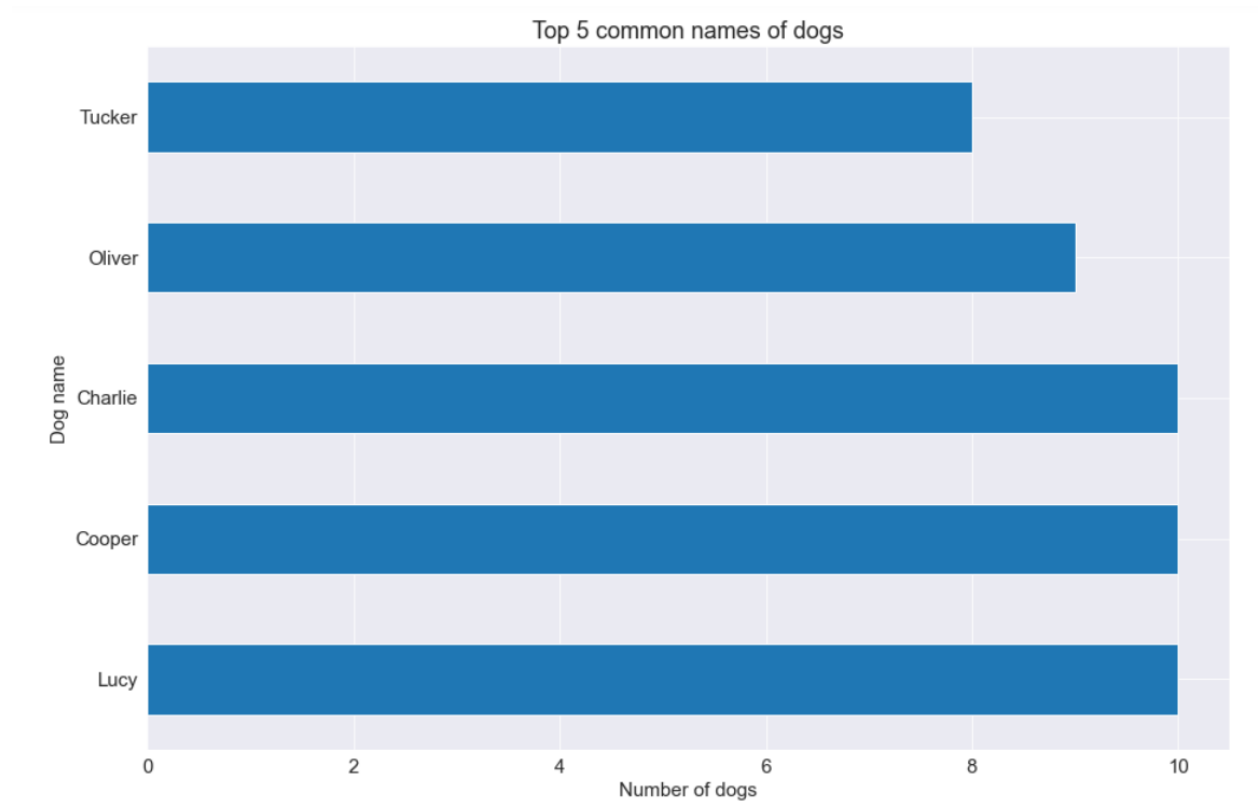And this is the lovely dog that gets the highest rating.

Retweet and favorite analysis

The line graph below demonstrates that the both the retweeting and favoriting are increasing over time since November, 2015. However, the favoriting trend is increasing at a higher rates compared to the retweeting trend.

Top 5 most common names

The figure below demonstrates the top five common names among the rated dogs on *WeRateDogs* account. Three names share the first rank. These names are Lucy, Cooper, and Charlie.

Top 5 common names of dogs

Prediction model for dog breed

This part is an extra analysis as we try to utilize the dataset that is cleaned to make prediction about the dog breed based upon the selected columns. These columns are:

- Retweet count
- Favourite count
- Dog stage
- Rating(out of 10 per dog)
- Confidence level

The data is split into two datasets, the first one is for training the data (80% of the data), and the rest is for testing the model. Unfortunately, the accuracy of the model is not high enough to be reliable, but this maybe generated from the nature of the dataset, as it needs more columns such as the age of dogs, price of the dogs or the county of origin.