

WeRateDogs - act report

Tzu-Yu CHIU

12th August 2020

The Data

1. The WeRateDogs Twitter archive which was provided by Udacity. Download this file manually by clicking the following link:

https://d17h27t6h515a5.cloudfront.net/topher/2017/August/59a4e958_twitter-archive-enhanced/twitter-archive-enhanced.csv

2. The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (image_predictions.tsv) is hosted on Udacity's servers and should be downloaded programmatically using the Requests library and the following URL:

https://d17h27t6h515a5.cloudfront.net/topher/2017/August/599fd2ad_image-predictions/image-predictions.tsv

3. Using the tweet IDs in the WeRateDogs Twitter archive, query the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet_json.txt file. Each tweet's JSON data should be written to its own line. Then read this .txt file line by line into a pandas DataFrame with (at minimum) tweet ID, retweet count, and favorite count.

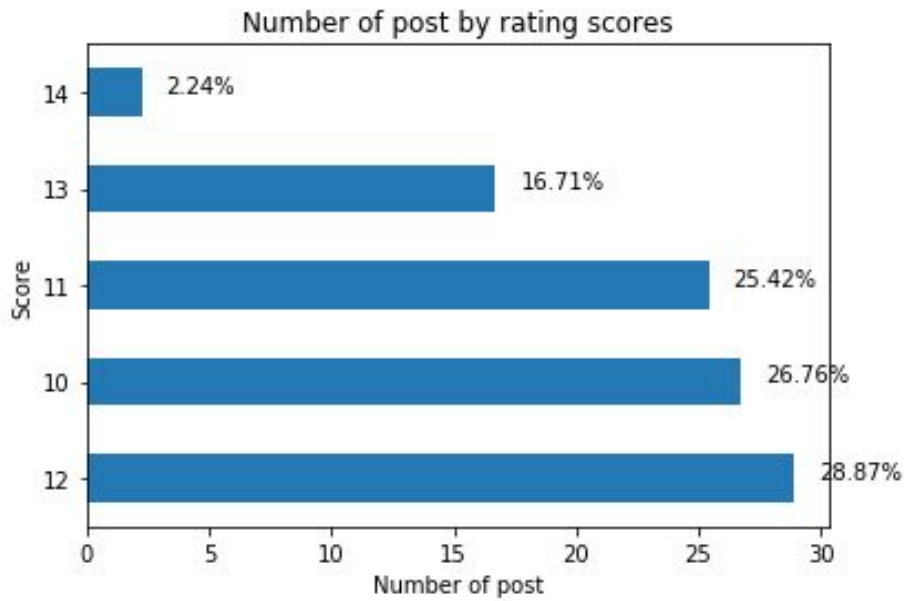
Wrangling data

Before going to analysis, I assessed and cleaned the data to help me for better analysis and visualization. After cleaning, there are 1562 tweets with good quality of data.

Insights

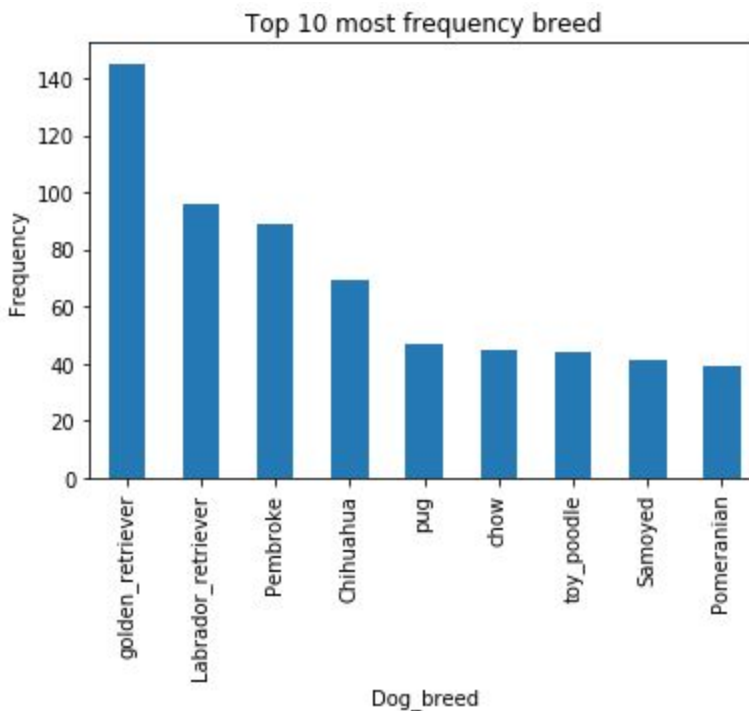
1. What is the frequency of each rating?

Most of the scores for dogs in our data is 12 (29%), then the frequency of 10 and 11 is more or less the same. According to the graph above, we assume that WeRateDogs rarely gave more than 14. I checked the dog breed for rating 14, Pembroke received rating 14 more than other breeds.



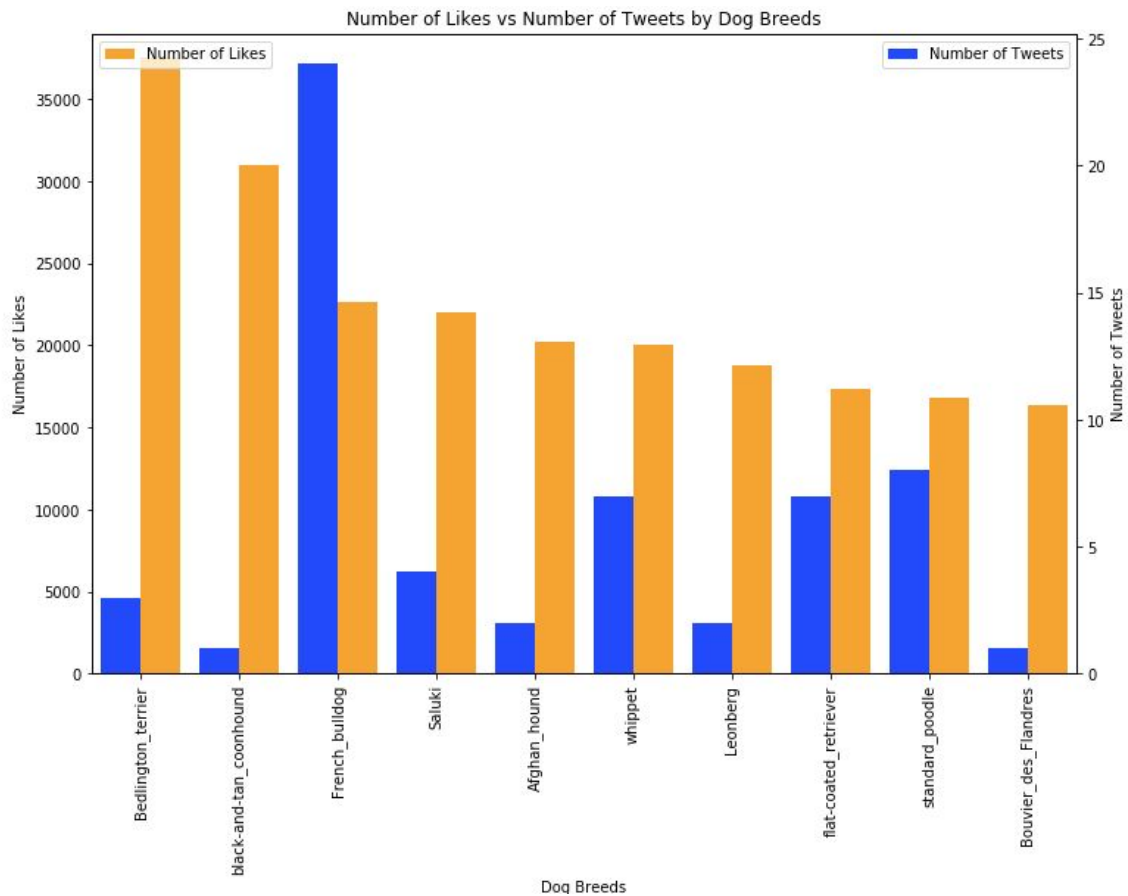
2. What are the top 10 most frequent breeds of dogs that appear on WeRateDogs?

The most common dog species from WeRateDogs over the time period 11/16/2015 and 8/1/2017 is golden retriever.



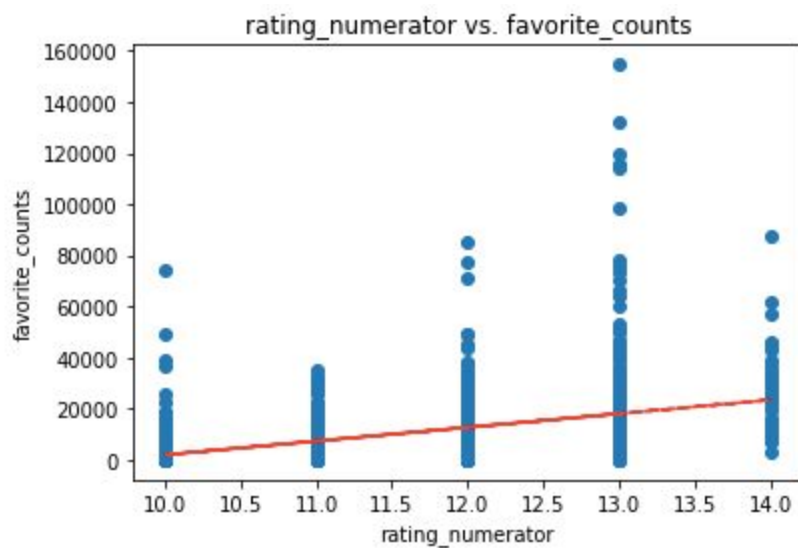
3. What breeds of dogs get the most likes per tweet?

On average, the dog breed "Bedlington_terrier" got most likes per tweet.



4. Does rating and likes have positive correlation?

We can assume that there is a positive correlation between rating numerator and favorite counts.



Visualization: WordCloud

- stopwords doesn't work for my visualization
- I can't put the words to the image (dog shape)

