Analyzing Wrangled WeRateDogs Twitter Data

Yagmur Dalman

Introduction

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter. [1]

In this report, I provide the analysis of the clean WeRateDogs data with 1908 observations. I mainly analyzed;

- the relationships between ratings, favorite counts and retweet counts

- The distribution of the Ratings based on the number of dogs and it's rating.

- The distribution of Retweets and favorites based on Ratings.

- correlation coefficients show that there is indeed a positive correlation between retweet counts and favourite counts.

Analysis

1.The dog with the highest rating

Cassie has the highest rating 14/10 (1.4). He looks like someone who likes good outfits which explain the haigh rate.
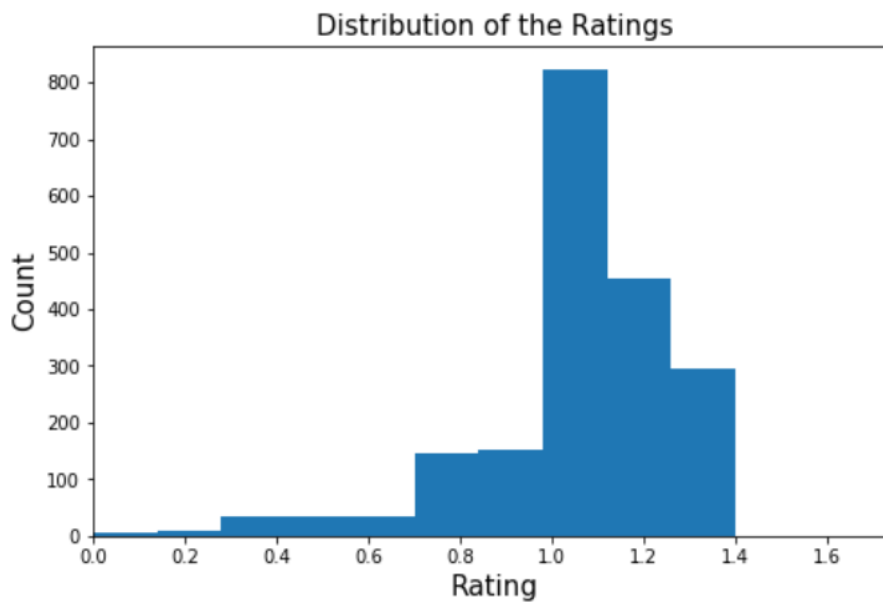
2. The favorite Dog

This puppo without a name has the highest number of likes and it is a Lakeland terrier. He also has the highest retweet count,132810, it can be easily seen from the descriptive analysis.

3. Descriptive Statistics

	tweet_id	rating_numerator	rating_denominator	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf	rating
count	1.988000e+03	1988.000000	1988.000000	1988.000000	1988.000000	1988.000000	1988.000000	1.988000e+03	1.988000e+03	1988.000000
mean	7.358197e+17	11.136821	10.535211	2765.925050	8897.936117	1.203722	0.594046	1.345927e-01	6.027390e-02	1.054152
std	6.751476e+16	8.631842	7.331277	4680.251304	12227.454419	0.561512	0.271908	1.007607e-01	5.093793e-02	0.219203
min	6.660209e+17	0.000000	2.000000	16.000000	81.000000	1.000000	0.044333	1.011300e-08	1.740170e-10	0.000000
25%	6.758399e+17	10.000000	10.000000	623.750000	1978.500000	1.000000	0.362775	5.401683e-02	1.616933e-02	1.000000
50%	7.084132e+17	11.000000	10.000000	1350.000000	4118.500000	1.000000	0.587797	1.175870e-01	4.947920e-02	1.100000
75%	7.878678e+17	12.000000	10.000000	3220.000000	11315.500000	1.000000	0.845599	1.954432e-01	9.162278e-02	1.200000
max	8.924206e+17	204.000000	170.000000	79515.000000	132810.000000	4.000000	1.000000	4.880140e-01	2.734190e-01	1.400000

4. Distribution of the Ratings
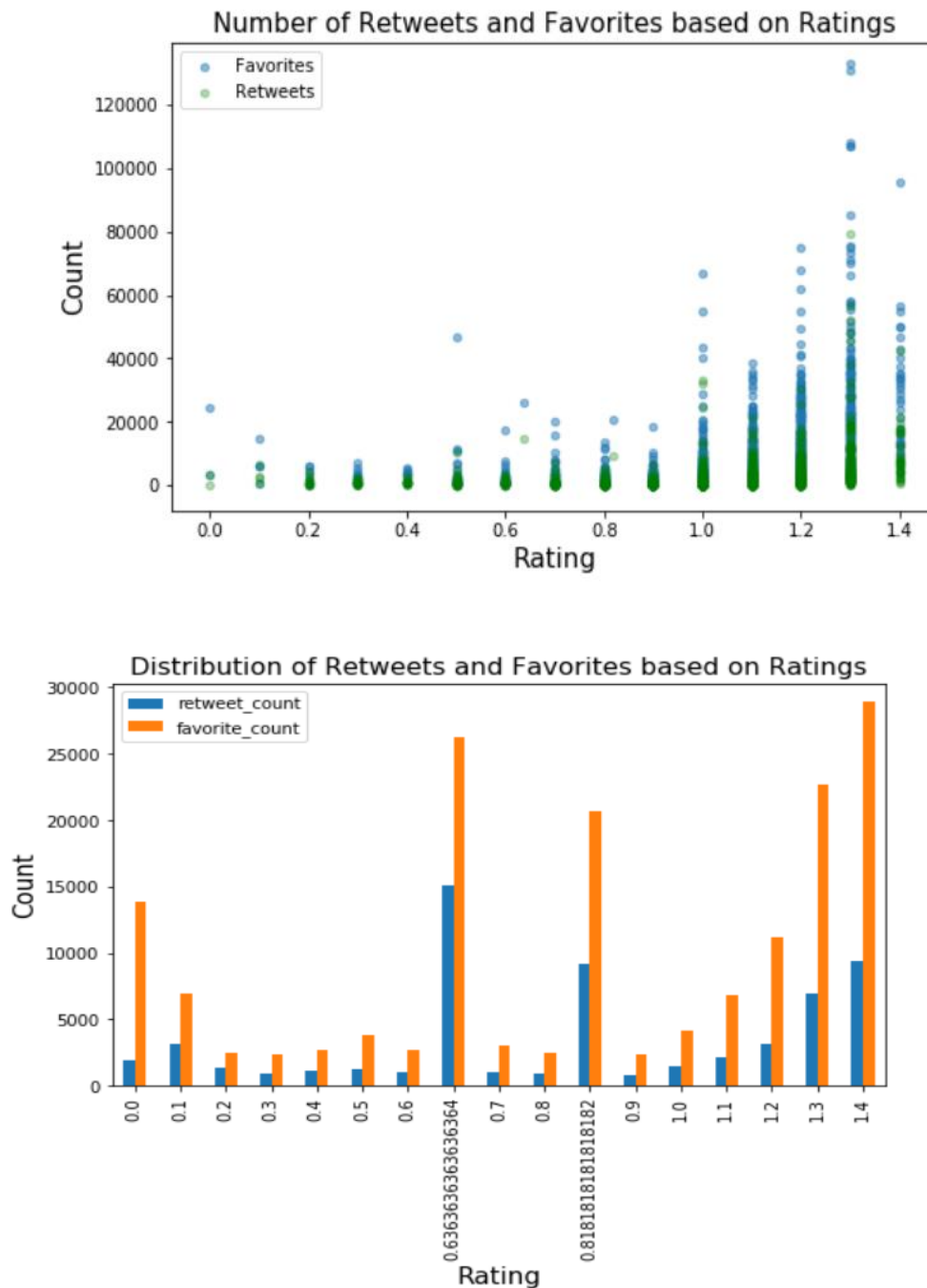

Distribution of the Ratings

- The mean dog rating is 1.06 and the ratings are more frequent between 1 and 1.3, i.e between 10/10 and 13/10.
- 1.2, 12/10, is the most frequent rating.
- The median value for rating is 1.1, 11/10, which means that half of the ratings are greater than or equal to 1.1.

5. Relationship between Favorite Count, Retweet Count and Rating

I analyzed the relationship between the these variables, retweets and favorites, and found out that they are positively correlated.

Number of Retweets and Favorites based on Ratings

Distribution of Retweets and Favorites based on Ratings

- The favorite counts are higher than the retweet counts for each rating.
- The dogs with the ratings between 1.3 and 1.4 are retweeted and liked most.

It seems like there is a positive correlation between retweet counts and favorite counts based on ratings.
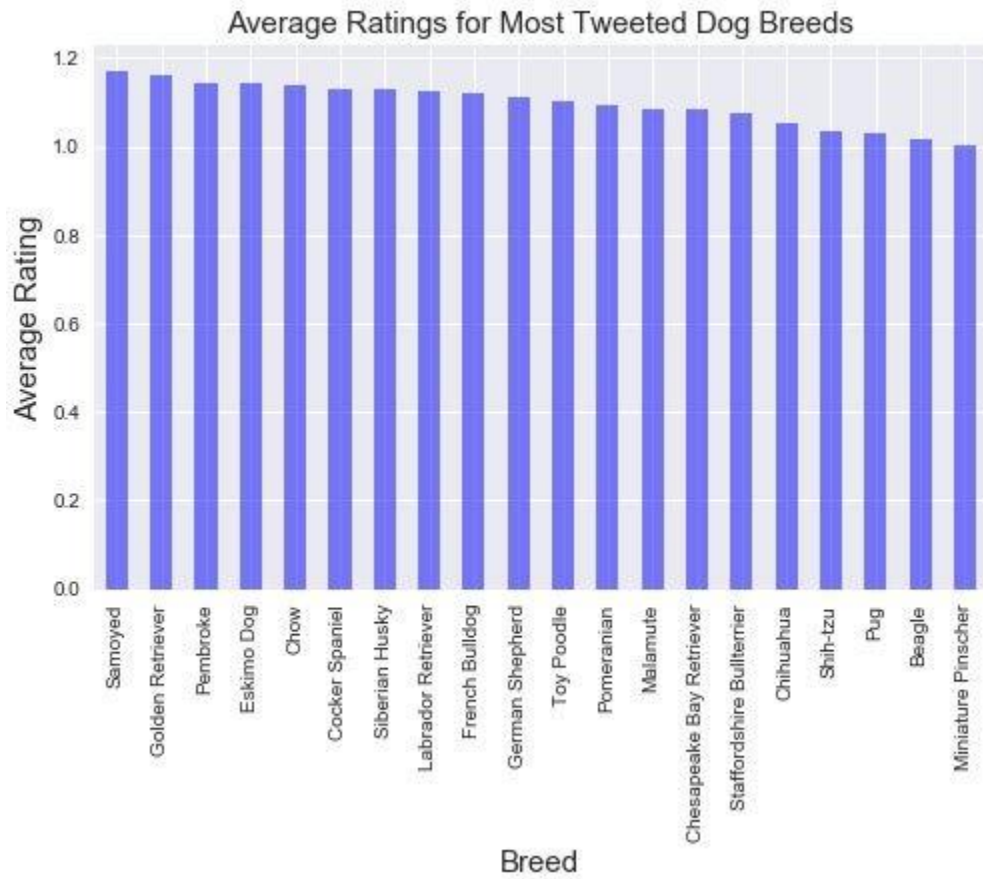
Let's look at the Pearson correlation coefficients:

	rating	retweet_count	favorite_count
rating	1.000000	0.306295	0.406224
retweet_count	0.306295	1.000000	0.930268
favorite_count	0.406224	0.930268	1.000000

- Pearson correlation coefficients show that there is indeed a positive correlation between retweet counts and favorite counts. But there is no correlation between ratings and retweet counts, as well as ratings and favorite counts.
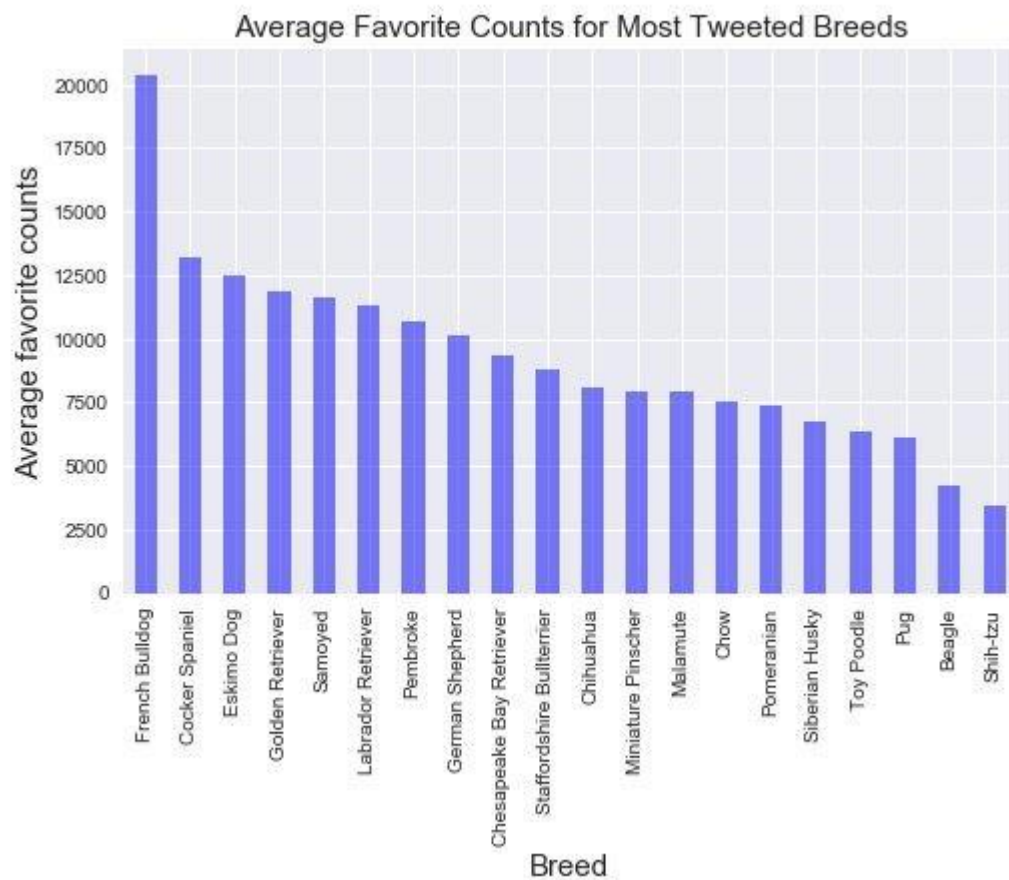
I only took into account the breeds with at least 20 tweets and the most tweeted dog breeds in order: Golden Retriever, Labrador Retriever, Pembroke, Chihuahua, Pug. The number of Golden Retriever is substantially higher.

Average Ratings for Most Tweeted Dog Breeds

Average Ratings for Most Tweeted Dog Breeds

Most tweeted dog breeds with the highest rating in order: Samoyed, Golden Retriever, Pembroke, Eskimo Dog, Chow.

Average Favorite Counts for Most Tweeted Breeds

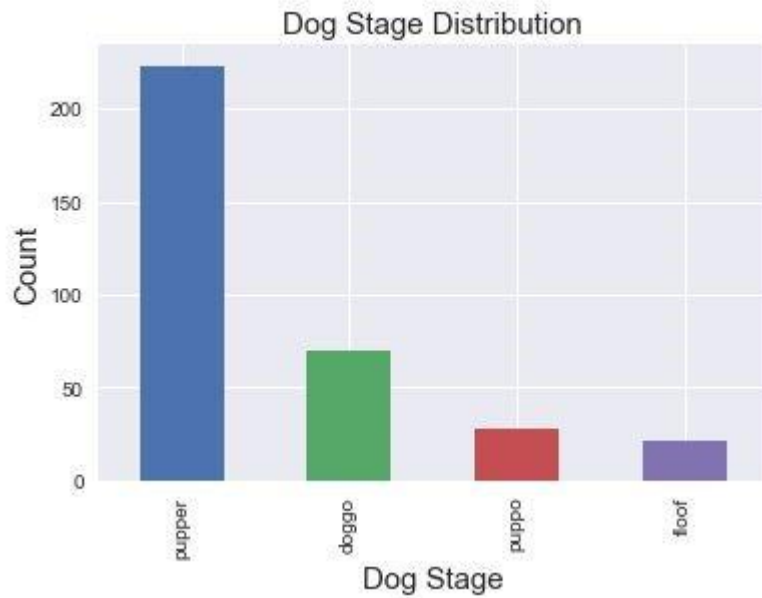Average Favorite Counts for Most Tweeted Breeds

- Most tweeted dog breeds with the highest number of likes: French Bulldog, Cocker Spaniel, Eskimo Dog, Golden Retriever, Samoyed.
- It looks like **French Bulldog** is the most favorite dog breed.

Most liked breeds differ from the breeds with highest ratings. **Samoyed**, **Golden Retriever**, and **Eskimo Dog** are common in the both lists.
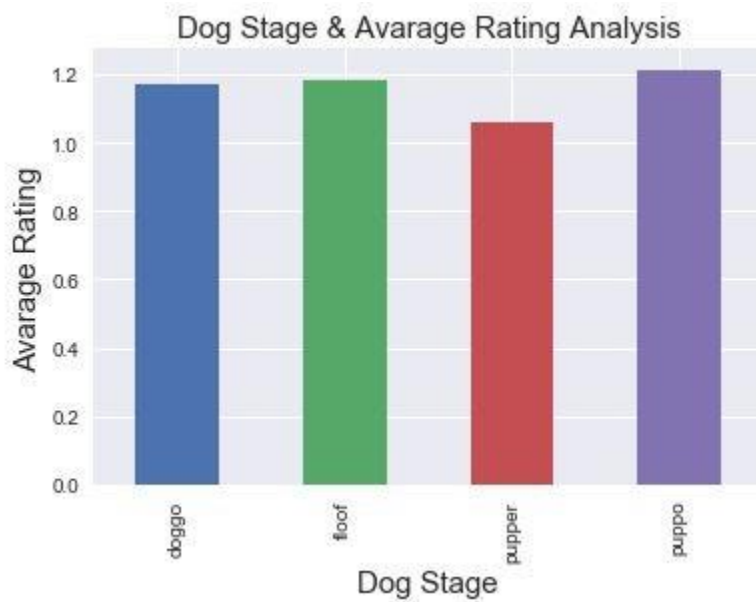
7. Dog Stage Analysis
Dog Stage Distribution

There are 1566 dogs with unknown dog stage. It is a very high number and more time should be spent on cleaning dog_stage part. However, for the current analysis, I excluded the unknown dog stages.
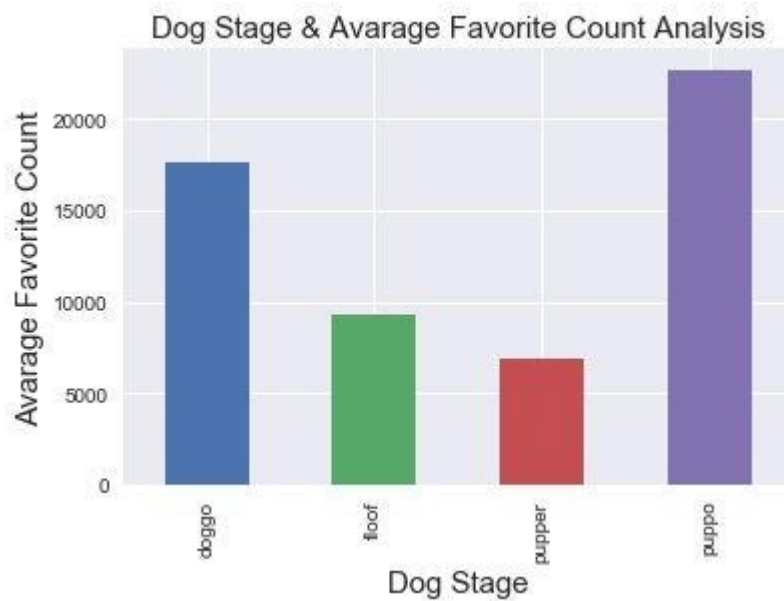

Dog Stage Distribution

Most of the dogs are pupper.

Dog Stage & Average Rating Analysis

The highest average rating is owned by puppos. Since it is an account for fun, it makes sense that all of them have similar and good ratings.

Dog Stage & Average Favorite Count Analysis

● The puppos has the highest favorite counts.

- Although, the puppers are more tweeted about, they have the lowest rating and favorite counts.
- As I stated before it is not surprising to have such differences between ratings and like counts because the ratings are given by a unique account holder WeRateDogs in a subjective way.

Remarks:

- The ratings, retweets and favorites are not only related with the dog images or videos, the humor of the account holder also affects the followers' decisions.
- Image prediction algorithm is not always successful to detect the dog breed. Sometimes it detects another object in the picture rather than the dog, like a bagel.
- I believe I did a good job in re-extracting, cleaning and correcting names and ratings but dog stages needs much more work rather than extraction. Most of the dogs, their stages are not stated in the tweets.
- Some tweets rate other things rather than dogs, I cleaned a significant amount of data but it was not that easy to clean all of them, that is why there are still very low ratings.