2019

Our Best Friends Ever

Osama Osman

Data Wrangling Act Report

4/3/2019

WeRateDogs Insights for Dogs' Lovers

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "**they're good dogs Brent**." If you are a dogs' lover like me; the insights gleaned from this twitter account data introduces some interesting facts about dogs and dogs' lovers behaviour towards their posts that mostly features dogs' photos in it.

1. How is the interaction with the account's posts doing over time?

The interaction is usually defined in terms of favorites and retweets. In the line plot below, there's an evident upward trend in the average monthly count of retweets and favorites attracted by WeRateDogs posts over the time of the analysis. People engagement with the posts is a great indicator of appeal of the content of the posts.
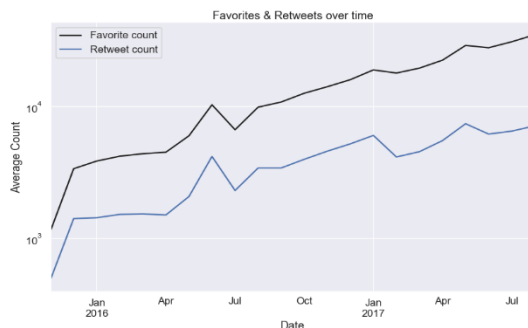

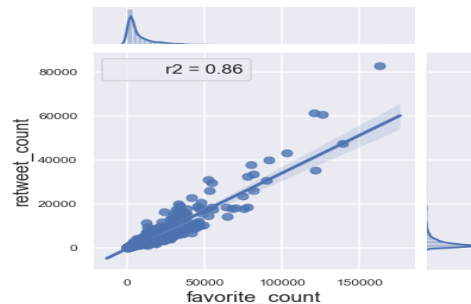*Figure 1 - The average monthly retweets and favorites count*

Figure 2 - Association between the favorite counts and retweets count

There's also an apparent association between the retweets and favorites counts that is illustrated in Figure – 2 that features a considerably high coefficient of determination of 0.86.

2. But, hold on for a second, this might be due to larger number of posts tweeted out by WeRateDogs not the interesting and engaging content!

That's where the tweets monthly count over time will clarify in Figure – 3, the average number of posts tweeted out by WeRateDogs on monthly basis with a bit lower than 100 pots per month as represented by the blue horizontal dotted line. The Overall Monthly tweets released were consistently below this number most of the time. This indicates that the posts content rather than counts is the driving force behind attracting increasing retweets and hiking favorites over time.

Now that we know that the content is the intriguing factor that leads to retweets and favorites, let's investigate the main components of that content in the upcoming insight.
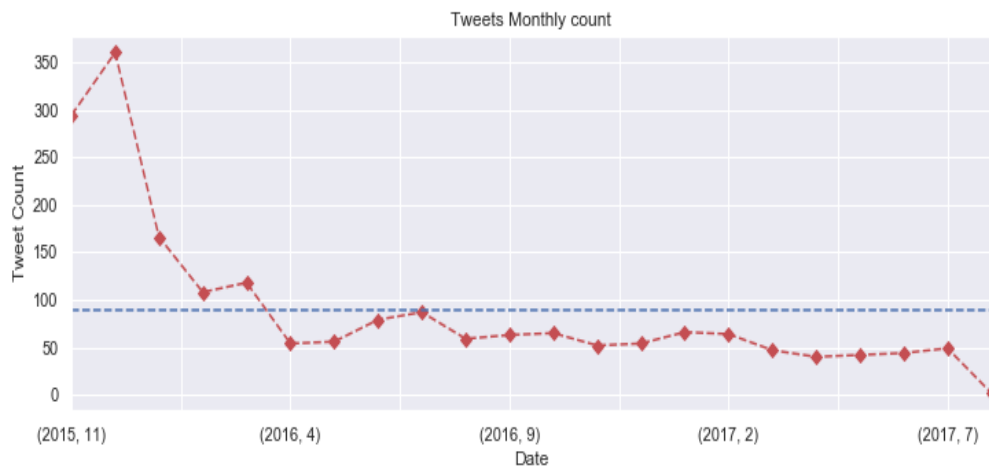

Figure 3 - Monthly Tweeted out tweets by WeRateDogs

3. Do Posts featuring dogs attract more favorites and retweets than others that don't?

The boxplots show that tweets that featured dogs in their images typically generated more retweets and favorites interaction and were more engaging to the account followers.
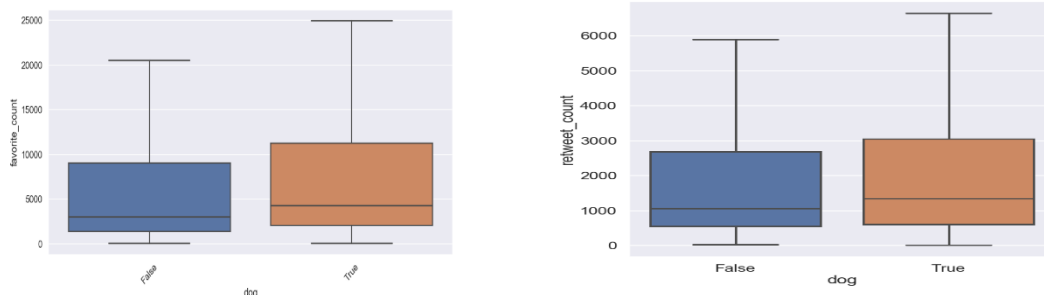

Figure 4 - Retweet and favorite counts boxplots for posts with dogs and without dogs

4. Another measure for the content quality is the rating of the posts, let's have a sneak peek on it.

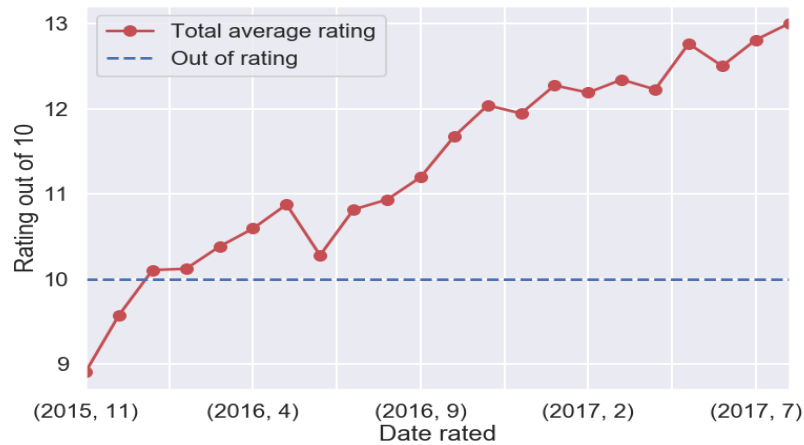In general, the average rating for the posts was growing higher over time as illustrated in figure 5.

*Figure 5 - Average rating of the tweeted-out tweets over time*

But again, what about the ratings for tweets that feature dogs in comparison with those that don't. The following boxplot highlights the fact that rating for dogs' featuring posts is basically higher than those that don't across the years included of the dataset.


*Figure 6 - Boxplots for the rating distribution across the years for posts with and without dogs*

5. Delving into the dogs' characteristics and Breeds that are more engaging, the upcoming plots leads to the following insights:

a) Pupper dog stage is the most featured in the posts as per the following bar chart.
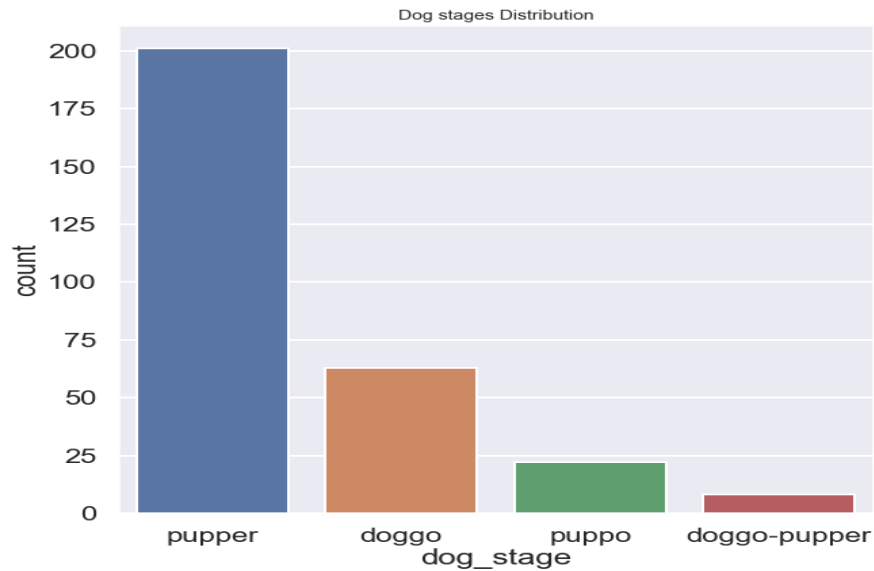
b) Even though pupper stage is the most featured, it's not the best in terms of generating favorites and retweets. Meanwhile, the Puppo and Doggo are performing better in terms attracting retweets and favorites
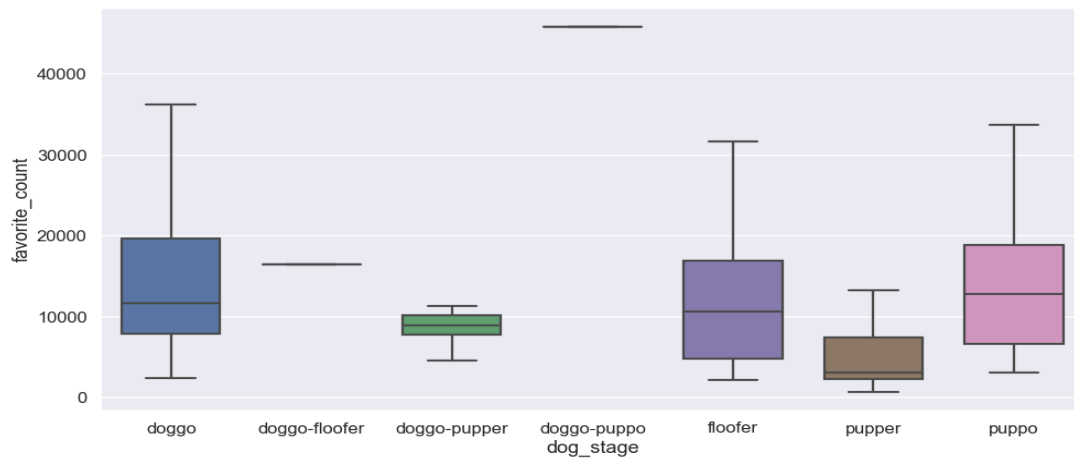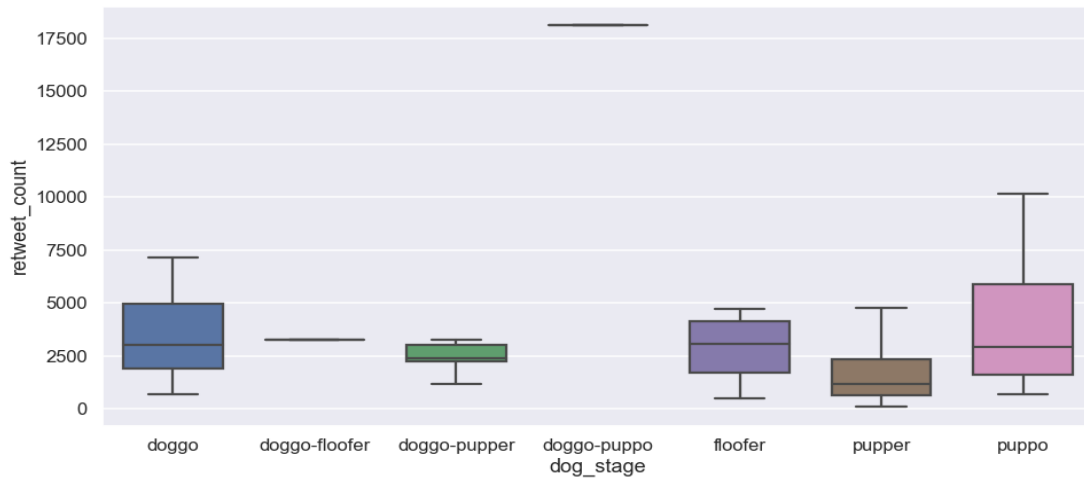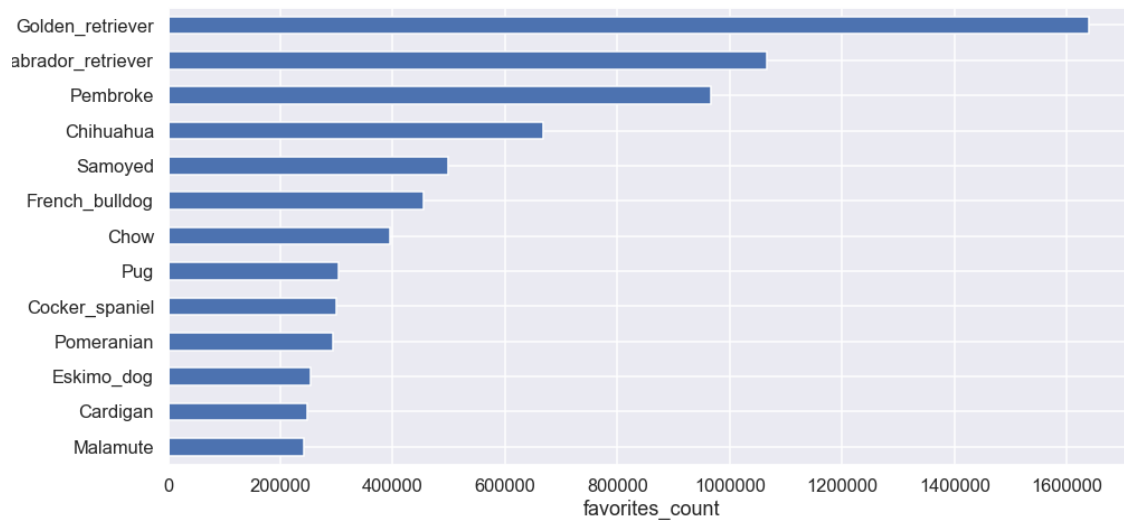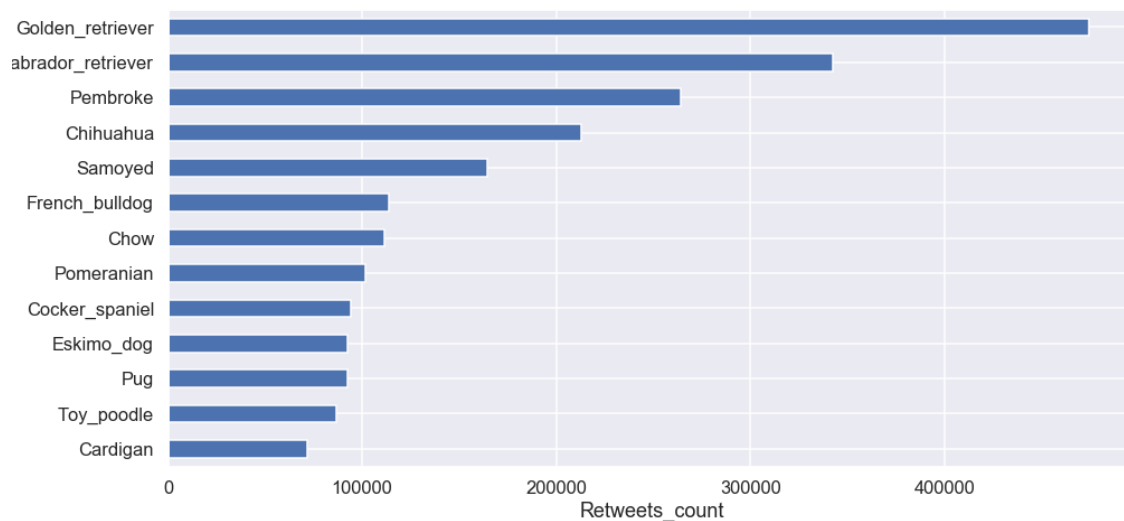
c) The dogs breed also has its impacts on the engagement; The most frequent dog breed featured in the tweets are as follows:
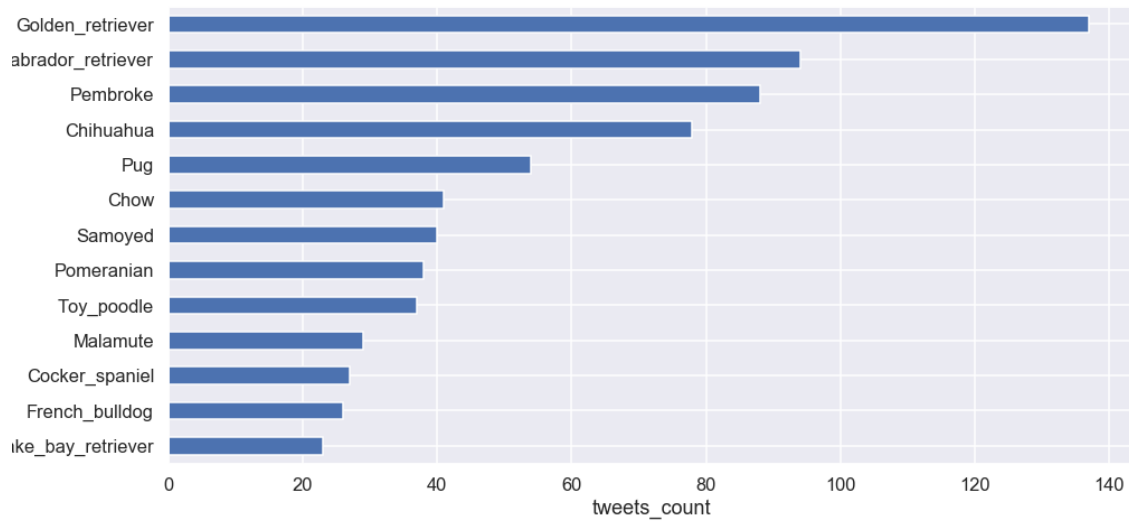
d) The dog breeds ranking in terms of generating favorites is as follows:

e) While the dog breeds ranking as per the retweets they had is as follows

Golden retrievers are the most widely loved and retweeted about as illustrated in the above charts.

Final words:

The dataset is really rich and full of insights that deserves much more time to glean after massaging the dataset and digging it up more and more. But for the sake of the project, that's enough for now.