WeRateDogs Analysis Report

About

In this analysis, I'll be using data from the Twitter account @dog_rates which currently boasts of over 9 million followers.

This Twitter account has had coverage by international media for its humorous comments and ratings of people's dogs.

The ratings almost always have a denominator of 10 and a numerator greater than 10, but why you may ask? Because ***"they're good dogs Brent."***

Data Gathering

First, I downloaded a file from Udacity's servers called *twitter_archive_enhanced.csv*, then uploaded it and read it into a pandas DataFrame.

Second, I programmatically downloaded another file hosted on Udacity's servers called *image_predictions.tsv* using the request library.

Finally, we obtained Twitter data from their API using tweet IDs from the WeRateDogs *twitter_archive_enhanced.csv* file and stored each tweet's entire set of JSON data in a file called *tweet_json.txt*.

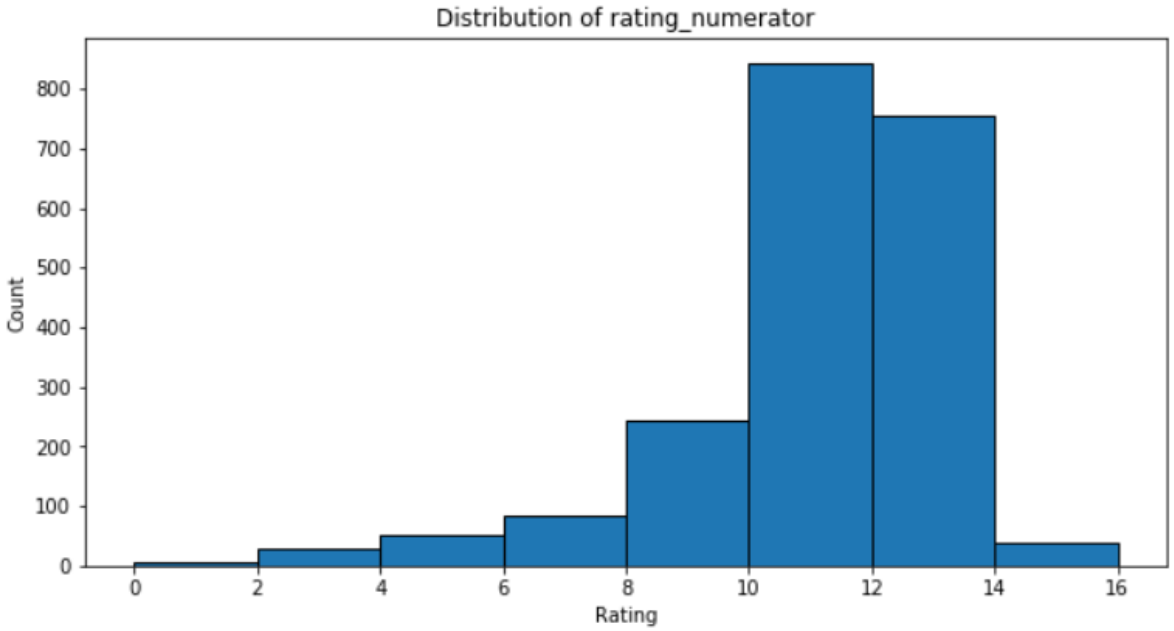
Data Wrangling.

I had to wrangle the data into shape before I could begin analyzing it. I looked at the data both visually and programmatically, checking for errors and inconsistencies. After cleaning many of the issues found during that assessment, there were about 2050 tweets with good quality and tidy data.

Analysis insight

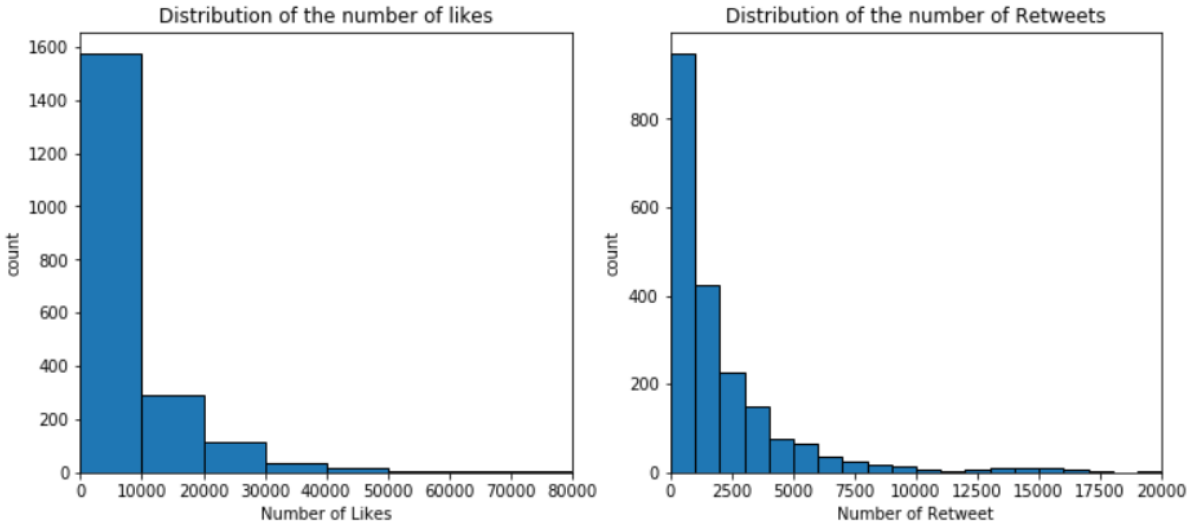
Good Ratings

The histogram below shows the distribution of rating received by the dogs. Most dogs receive a rating that is between 10 to 14.



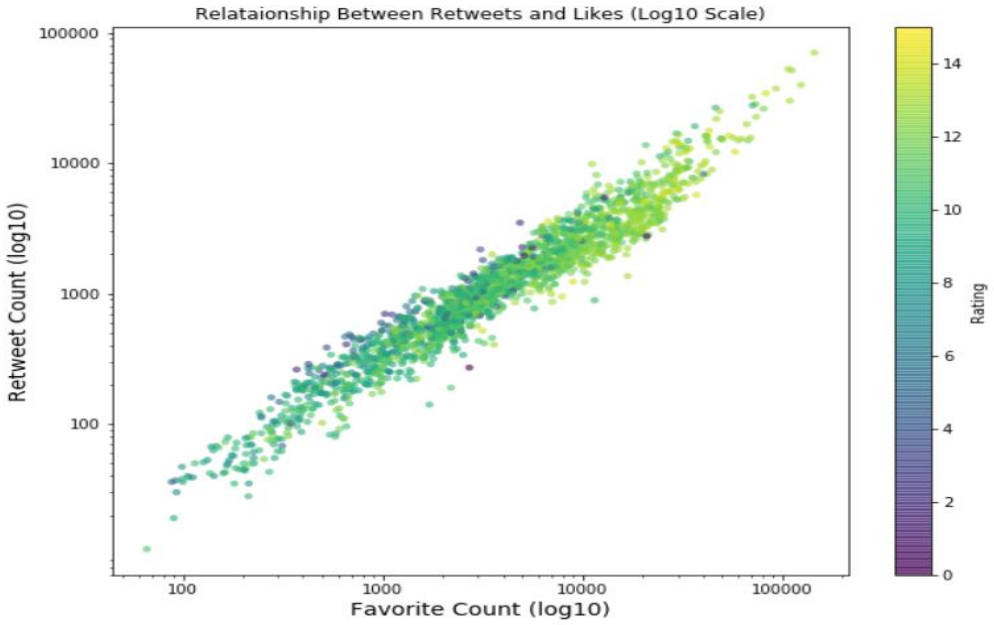
Likes and Retweets

Although the favourite count and number of retweets are both skewed to the right, you can see that most tweets tend to have higher number of likes. I would say, it's always easier to hit that heart icon than to hit the retweet button.



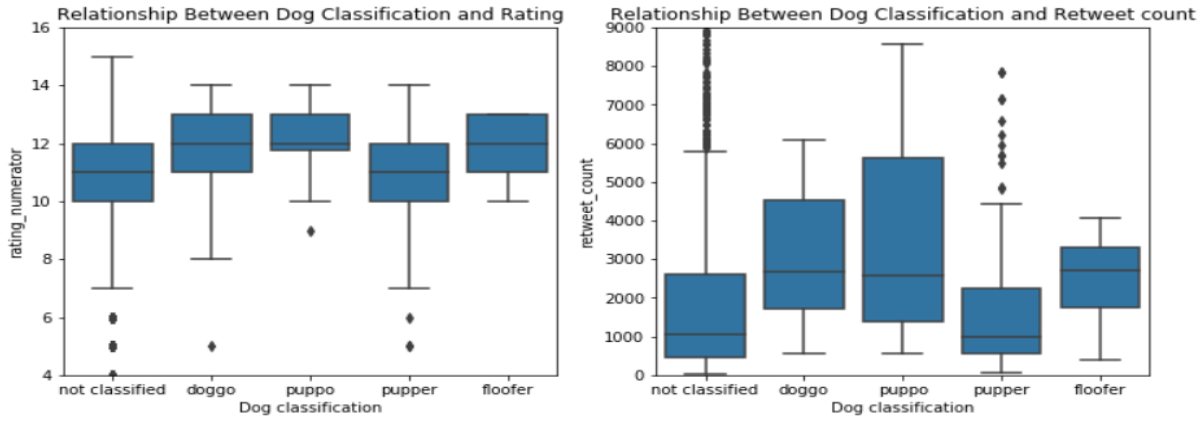
Ratings VS Retweets VS Likes

The number of retweets is strongly correlated with the number of likes a tweet would get. The correlation coefficient is 0.86, so there is a very strong positive relationship between the two measures. When you add ratings to the equation, you can clearly see, as demonstrated on the graph below, that the higher your rating, the higher the number of retweets and likes you will receive. From the graph, yellow and yellow-green represents a higher rating, and these two colors are scattered at the highest peak of the graph.



Dog Stages

I can say that doggo and puppo are loved because they received a higher rating score than the other dogs. Since a higher rating translates to more retweets as we discussed earlier, their retweets numbers are on average higher.



Conclusion

Do you have a dog? What rating do you think it would receive on Twitter? If you're curious, find out by following @dog_rates. I hope this analysis convinces you to give it a try. "They're good dogs Brent." And they will receive a good rating.