

We Rate Dogs Tweets Analysis and Visualization

- **Introduction**

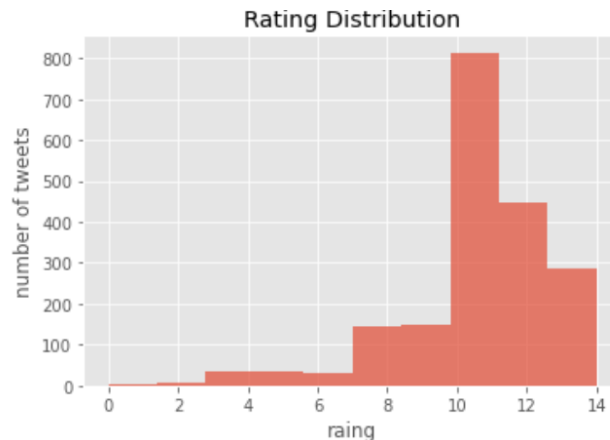
- The dataset utilized to process is from a tweet archive of Twitter user [@dog_rates](#), also known as [WeRateDogs](#). WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.
- WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for you to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

- **Data Wrangling**

- Before analysis, however, this dataset has some quality and tidiness issues. By using both visual and programmatic assessment, I found 12 quality issues as well as 4 tidiness issues. After addressing these problems, I analyzed and visualized the clean data and got some insights and summarized as below.

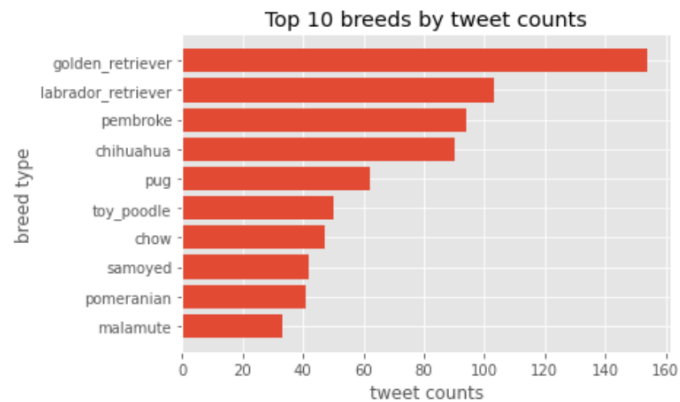
- **Insights**

1. **Rating distribution for the tweets**



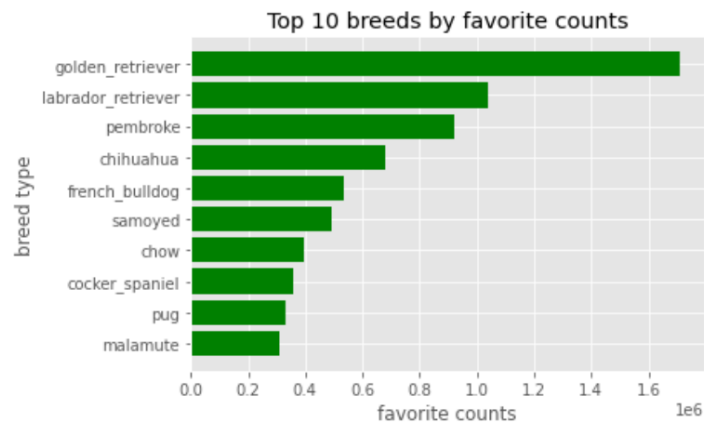
- It is clearly that most tweets got very high ratings (bigger than 10) and the percentage up to nearly 80%. People really like those dogs shown in the tweets!

2. What are the most common dog breeds found in the tweets?



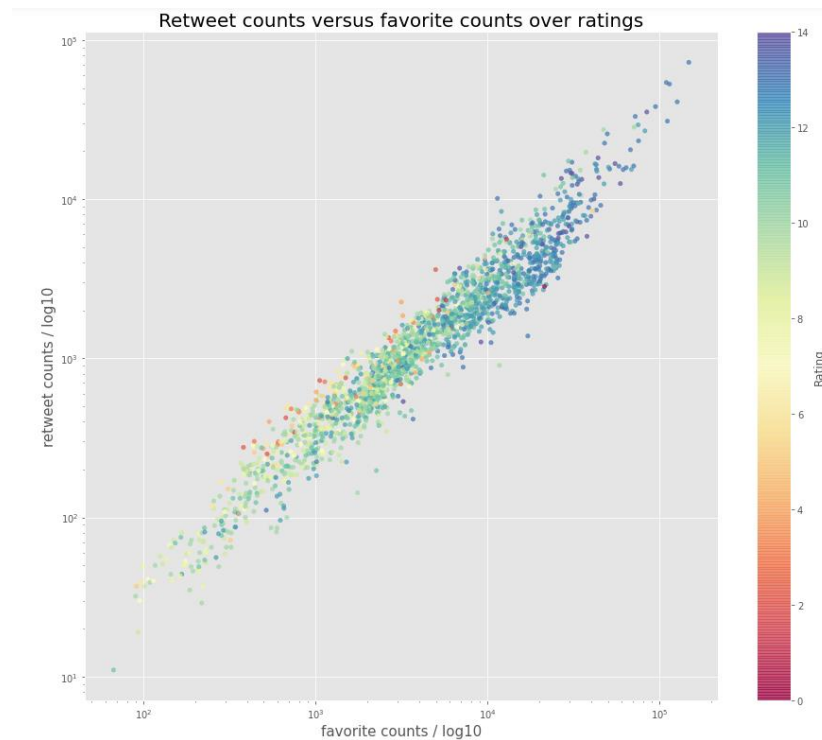
- It seems that [golden retriever](#) is the most popular breed, and it is up to 21%.

3. Which dog breed get the most favorite counts?



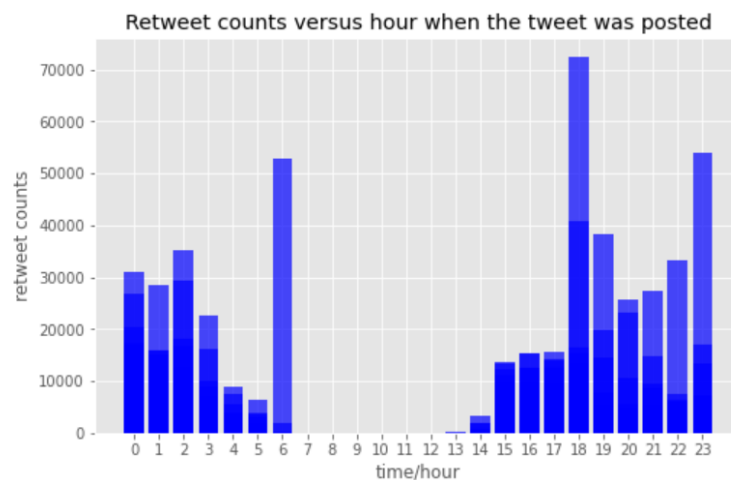
- Unsurprisingly, it is still the [golden retriever](#) got the most favorite counts, up to 1710974 times.
- By comparing above two graphs, we also find that top 4 breeds by tweet counts and by favorite counts are in the same order which makes a lot of sense.

4. Relationship between retweet counts and favorite counts



- As we can imagine, the retweet counts are strongly correlated with the favorite counts. In other words, if a user likes a tweet, it is more likely that the user will retweet it.

5. When people are more likely to retweet?



- From the analysis, we notice that if a tweet was posted on afternoon 6 pm, it has a higher chance of being retweeted. Instead, if it was posted between 7 am to 12 am, it barely being retweeted. So next time before tweeting, choosing a suitable time is necessary 😊