Ainur Artay
Data Analytics NanoDegree
January 2020

ANALYSIS OF WERATEDOGS TWITTER ARCHIVE DATA

OVERVIEW

This Project is wrangling, analysis and visualisation of a dataset from the 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 dogs. The ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10! 11/10, 12/10, 13/10, etc. WeRateDogs has over 8.6 million followers, has received international media coverage and produces its own merchandise.

WeRateDogs Twitter archive in this Project contains basic tweet data (tweet ID, timestamp, text, etc.) for 5K+ tweets from November 15, 2015 to August 1, 2017.

WRANGLING

The data was "cleaned" to make it easier to analyze and draw valid, accurate insights. The cleaning was aimed to achieve tidy, complete, valid, accurate and consistent data. The final, cleaned, dataset consists of about 2K tweets.

INSIGHTS

- People "favorited" the tweets more than they have retweeted them -- this is consistent with all tweets. Favoriting is a simpler action than retweeting, especially for people who don't like "flooding";
- Almost 80% of all tweets had ratings of 10 or more -- people love dogs!
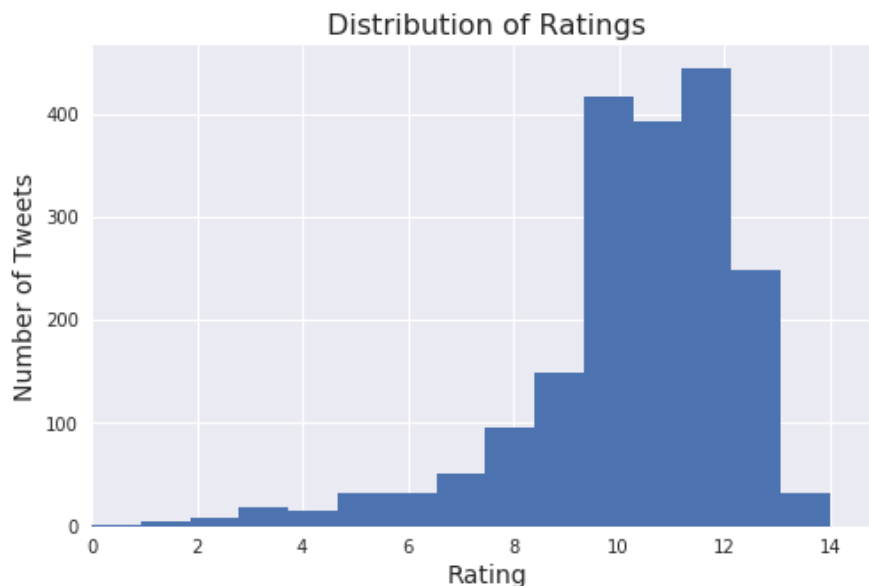
- Tweets with Retrievers were liked and retweeted the most

Top 10 Dog Breeds by Tweet Count

Top 10 Dog Breeds by Aggregate Favorite Count
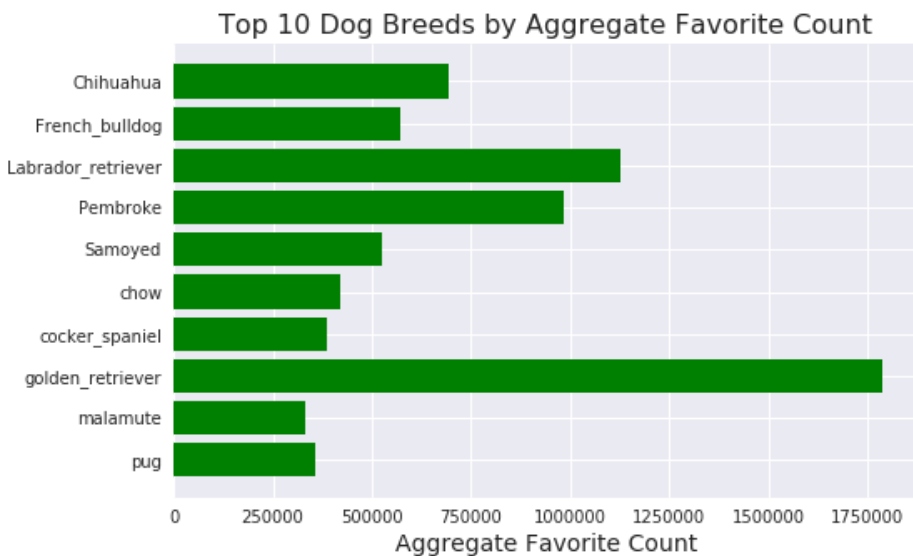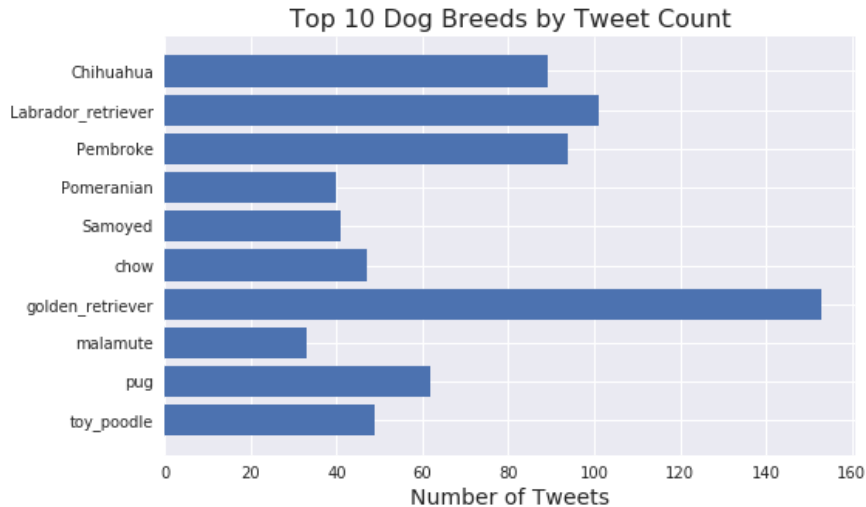
- Retweet counts and Favorite counts have a strong correlation -- which is an expected outcome

CONCLUSION

There are a lot of people who love, really love dogs (or dogs' pictures, at least).