

WeRateDogs Data Analysis Report

Background

As part of the Udacity Data Analysis Nanodegree Program, I am required to wrangle and analyze the Twitter Archive of 5000+ tweets about the WeRateDogs twitter account from November 15, 2017 to August 1, 2017. 2356 tweets were successfully extracted, about ratings of different dogs posted within the period.

A separate dataset containing predictions of posted dog images and another containing retweets counts and favorite counts was also documented to become part of the dataset.

This report dwells on the findings of the final data.

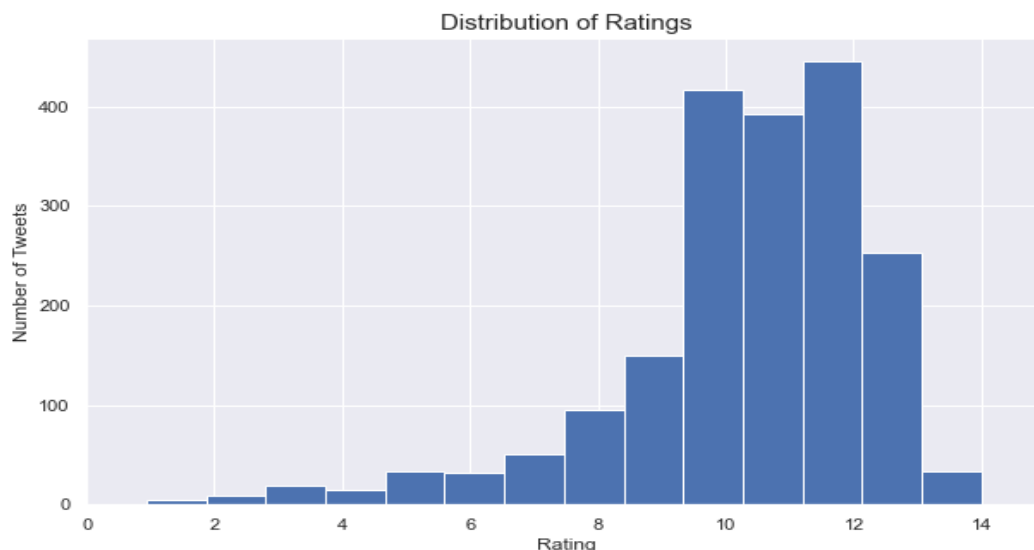
Data Wrangling:

The final data set is a product of the merging of the three datasets based on certain conditions and as such there were virtually non-existent data quality and tidiness issues prior to analysis.

Analysis:

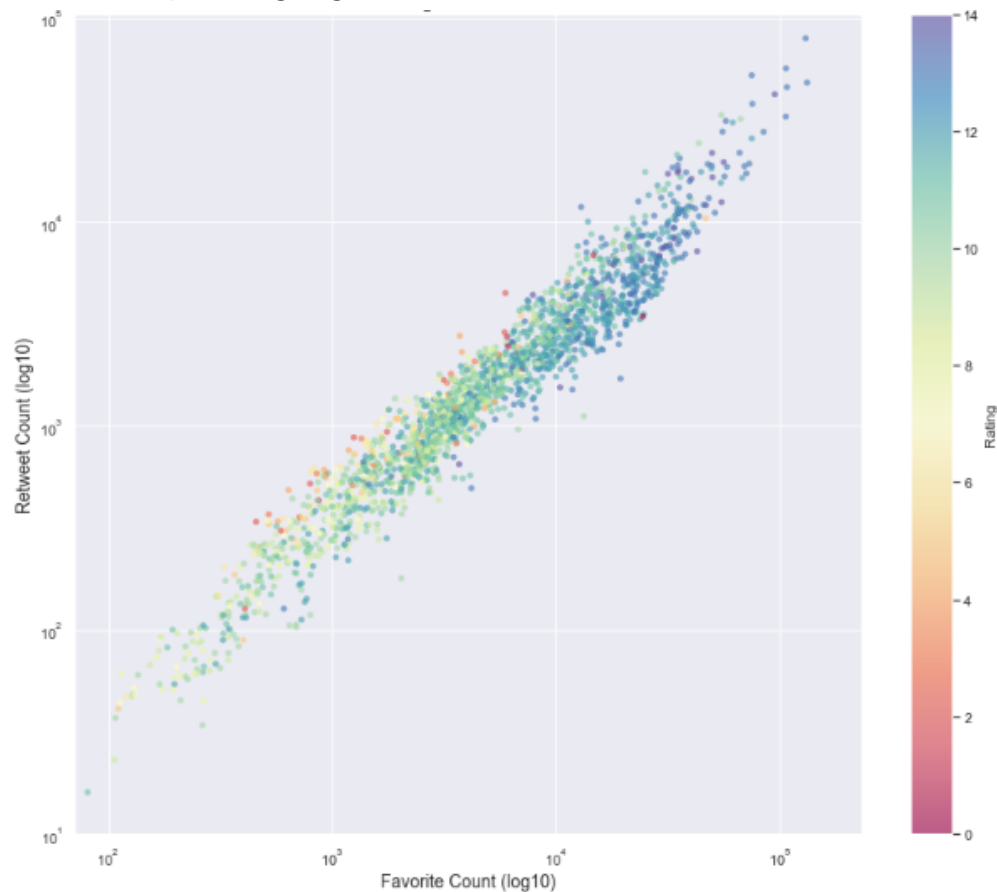
Ratings:

The distribution of ratings has a left skew which shows that most of the ratings were high. Infact 87% of the ratings were between 9 and 14.



More Retweets, More Favorites

My analysis shows a strong positive correlation between retweets and favorites. This pattern is noticed when you visually investigate the retweet counts and favorite counts. I find that tweets with disproportionately high number of retweets also have a corresponding high number of favorites.



Retrievers as Celebrities

Dog owners seem to love retrievers of different kinds according to feelers from our data. The Golden Retriever, and the Labrador Retriever occupy the top two positions of most popular dogs by number of tweets and favorite counts. The Pembroke, Chihuahua, Pug, Chow, Samoyed and Malamute are also popular among twitter dog pet owners.

