

Udacity Data Wrangling Project

Action Report

Matthew Bond

bondini051@gmail.com

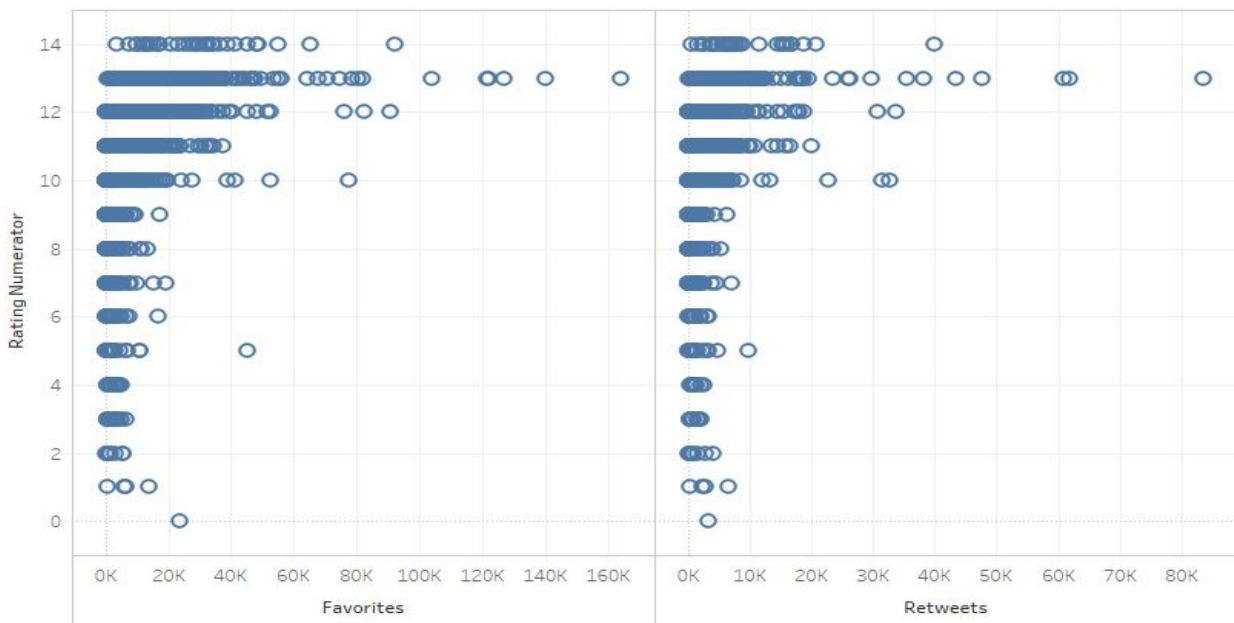
The @dog_rates, WeRateDogs twitter account, is the focus of this project. The project's goal is to gather the data from Twitter on the WeRateDogs twitter account and for the last 3000 tweets up to August 2017 to see what we can learn. At first glance going into the project I wanted to see what I could learn about the popularity of dog breeds.

The WeRateDogs twitter account posts or reposts pictures of people's dogs along with a funny caption and rating. The ratings are very subjective and even though the denominators are usually 10 but the numerators range from 0-420. In the cleaning phase, some of the outliers are removed but many ratings are still greater than the denominator of 10.

The bulk of the analysis is focused on favorites and retweets. We will also look at the most common names and the breeds associated with those names. The breeds are found using a neural network and had to be checked using the image from the specific tweet.

In analysing the number of favorites, it was found in general as the ratings increase the number of favorites and retweets for each tweet increased. In the visualization below, each circle is a tweet with the ratings on the y-axis and the favorites and retweets are on the x-axis. It shows that overall as ratings increase there is a higher number of favorites and retweets. There is also a greater number of tweets as the rating increases.

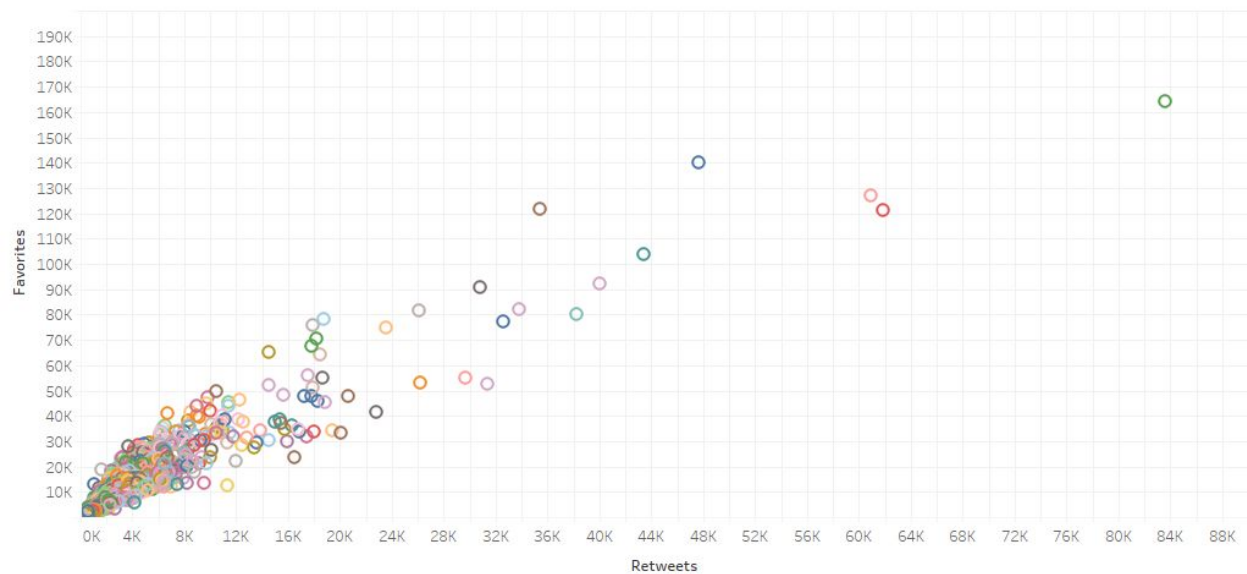
Favorites & Retweets as Ratings Increase



Favorites and Retweets vs. Rating Numerator. The view is filtered on Rating Numerator, which ranges from 0 to 14.

Below, the chart shows that overall as favorites increase retweets increase as well. There is not a one to one correlation but there is a loose positive correlation between favorites and retweets.

Retweets & Favorites



Retweets vs. Favorites. Color shows details about Tweet Id.

In doing analysis of the dog names there was a four way tie for the most popular dog names within the data we gathered. Lucy, Cooper, Charlie and Oliver were the most popular names. Within the 40 dogs that had those names the most common dog breed was Golden Retriever at 6 dogs and Labrador Retriever is a close second at 5. According to [The American Kennel Club](#), one of the biggest authorities on dog breeding and dog shows, since 1991 the Labrador Retriever is the most popular dog breed. From 2015-2017 the Golden Retriever was #3 on their list of most popular breeds. Unfortunately, the dog with the most amount of favorites and retweets was unnamed. However, the dog was a Labrador Retriever, the most popular dog according to the AKC as stated above.