Data Wrangling Project - Analysis

Udacity Data Analysis NanoDegree

WeRateDogs Twitter Archive

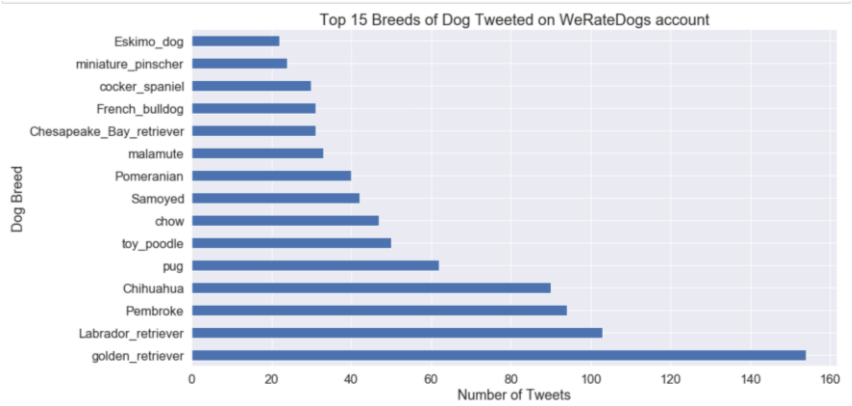
By Lovelina Richter

WeRateDogs Data

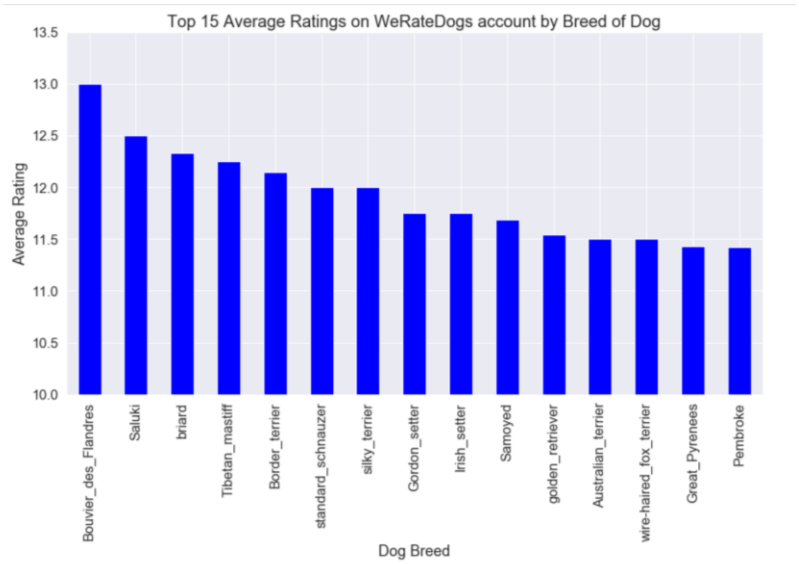
The WeRateDogs Twitter archive contains basic tweet data for all 5000+ of the original tweets. Combining the other data sets for image predictions, favorites, and retweet counts provided a better source of data for analysis. In this analysis I looked into the most tweeted by dog breed. I analyzed if there is a correlation or relationship between the dog breed and score given. As part of analysis, I removed the tweets with no dog breed data, because it will not provide me any insight for confidence level and rating based on dog's breed. Although, the project only requires one data visualization, I prepared graphs of number of likes or number of times the tweet was loved (favorites), ratings, and number of times the original tweet was retweeted, so that I can have additional information or insight to my analysis.s

Data Visualization

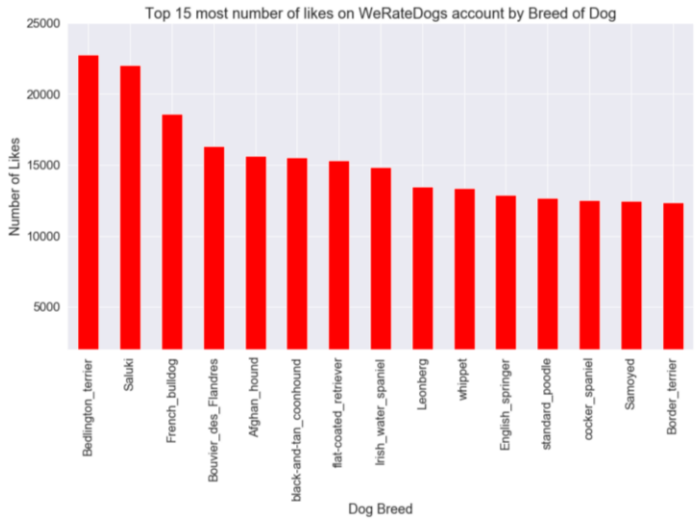
1. Top 15 dog breed originally tweeted



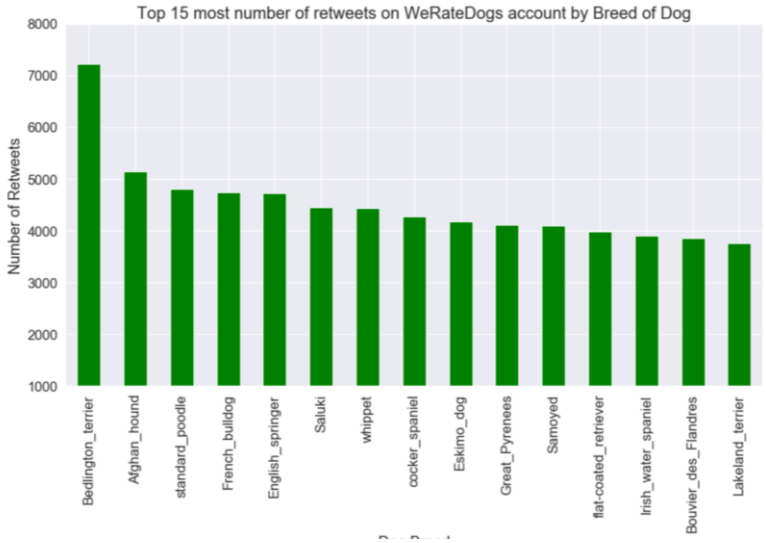
2. Top 15 average ratings by dog breed



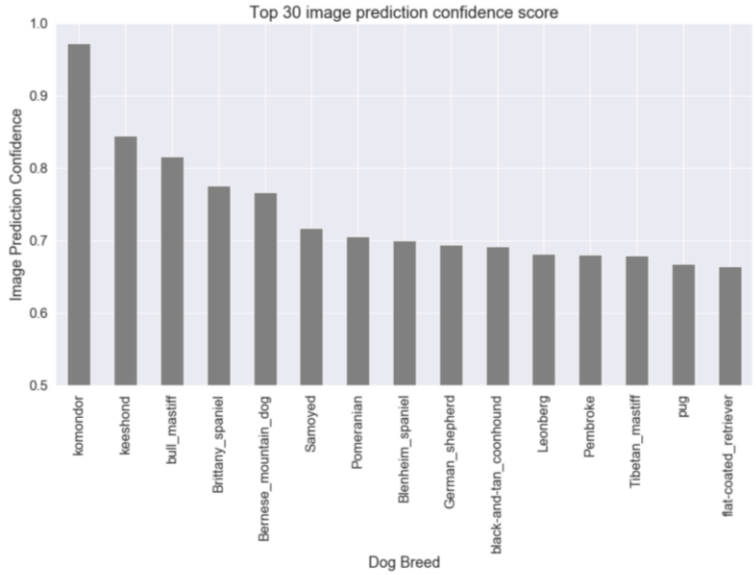
3. Top 15 most number of likes by dog breed



4. Top 15 most number of retweets by dog breed



5. Top 15 confidence rate by dog breed



Analysis

The data shows that the most tweeted dog breed is Golden Retriever. Personally, I find this result quite interesting, because I myself post pictures of my furry best friend golden retriever. However, three different breeds shows on top for every attribute. Bouvier des Flandres has the highest rating, Bedlington Terrier has the greatest number of likes, Bedlington Terrier again has the most number of retweets, and Komondor has the most confident image prediction rate. Among these 5 data measures, the dog breeds that appeared at least three times are Pembroke, Samoyed, and Flat coated retriever. Finally, Bedlington Terrier appeared to be the favorite among twitter users, based on number of likes and retweets.