Analysis and Insights in WeRateDogs Data
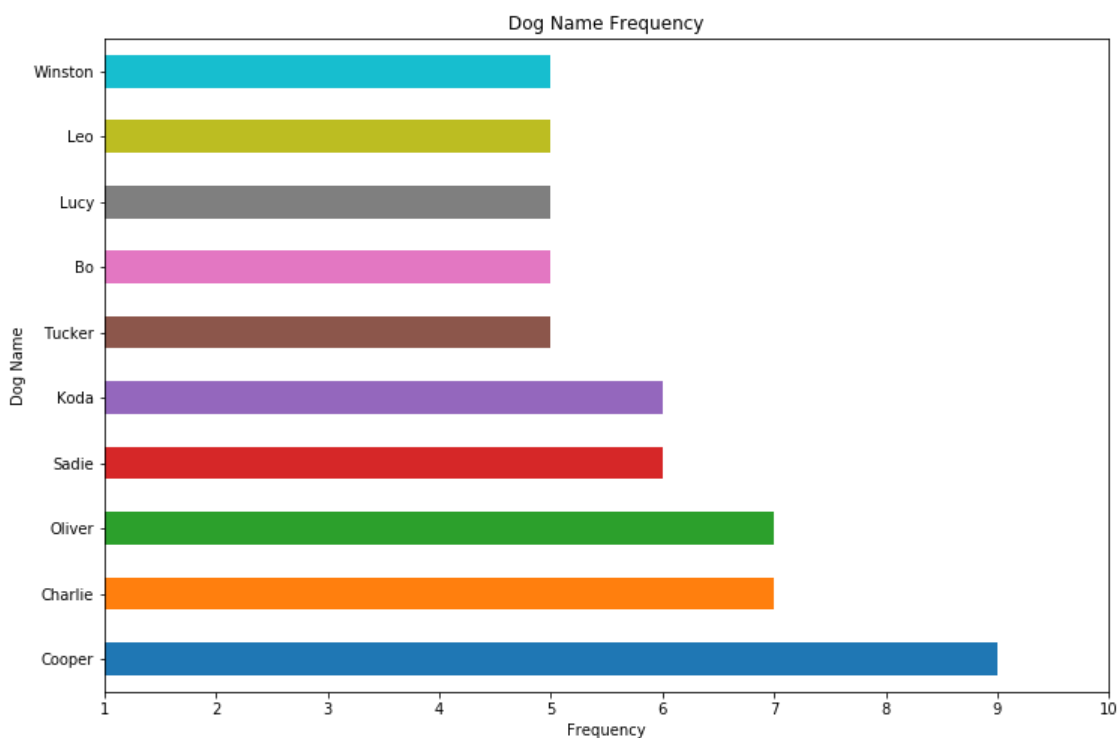*Folakemi Orekoya*

Introduction

The dataset that is been wrangled is the tweet archive of 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 almost always greater than 10. E.g. 11/10, 12/10, 13/10, etc. It has over 4 million followers. Twitter API was used for the retrieval of dog tweet data. This project works through the data wrangling process, focusing on the gathering, assessing, cleaning and combining of data to provide insights into how the data is correlated.

Common Dog Name: Cooper

The cleaned combined data consists of nearly 850 tweets that contain, among other data, the dogs names. A visualization of the data shows that the most popular dog name contained within the tweet data is Cooper with nine dogs sharing the moniker.
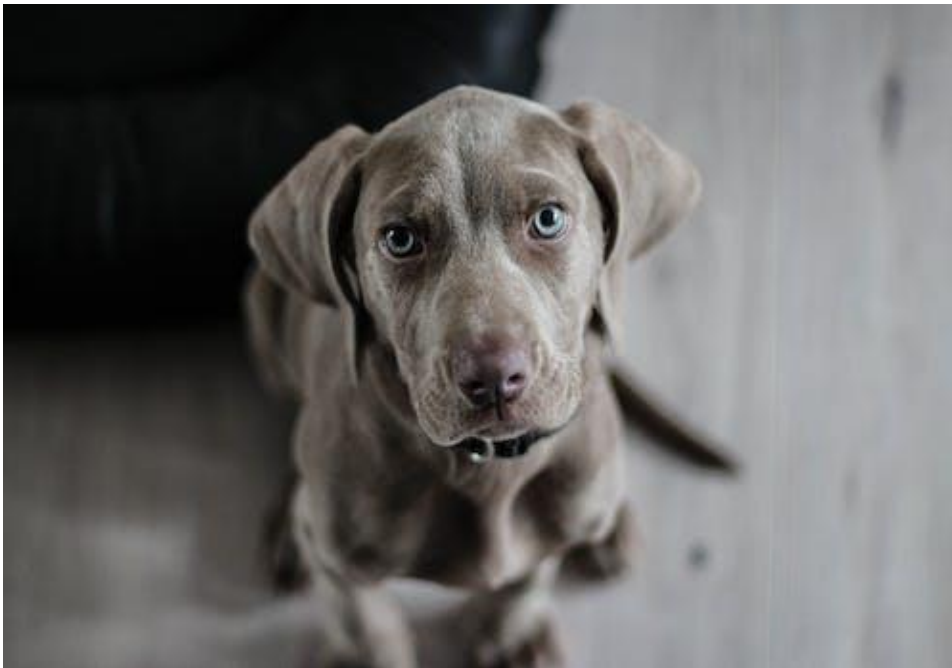

Dog ratings

Most of the dogs that are rated on the WeRateDogs twitter account have rating numerators larger than their denominator of 10. The dog breed with the highest mean rating numerator is the clumber with a mean numerator of 27!

Clumber

Breed predictions

The data also contains predictions of the various dog breeds. The prediction were generated by running the tweets through a neural network that can classify breeds of dogs The most frequent dog breed first prediction for a given dog is golden retriever follower by Labrador retriever in the second and third predictions. The prediction algorithm has a 63% confidence for the first prediction followed by 14% for the second prediction and only 6% for the third prediction.

Golden Retriever