Analysis Report for the WeRateDogs Project

by Ifeoma Ugbomah

Date: September 07, 2022

Introduction:

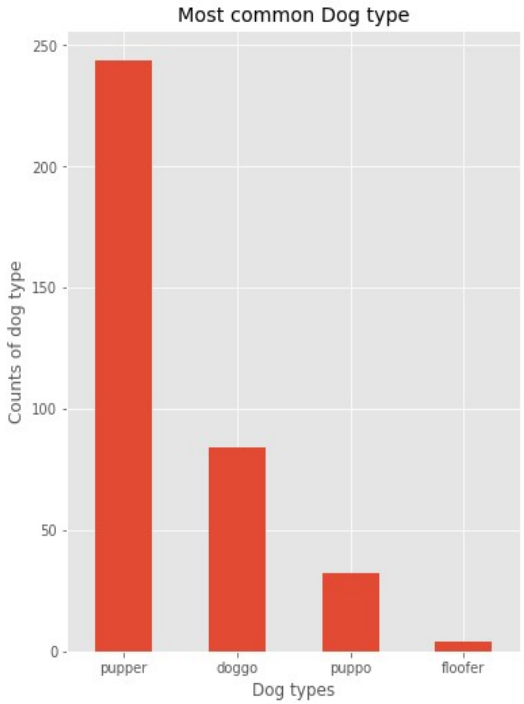
After having gathered, assessed and cleaned WeRateDogs twitter data I conducted some analysis on the dataset to establish some insights. The final twitter archives master dataset had features like favourite count, retweet count, tweet text, dog type, image url, prediction for the images, etc.

Analysis and Visualization:

Out of all the features in the dataset the features of interest to me were source of the tweet, common dog names. I found out that most of the tweet were from the Iphone device which suggested that majority were iphone users, and also there seems to be common dogs names like Charlie, Oliver, Lucy, Cooper, Penny. In terms of dog types, floofers and puppos had higher average ratings as compared to doggos and puppers. Though this can also be attributed to the fact that the number of dogs classified as floofers and puppos was a lot less than the number of dogs classified as doggos and puppers.

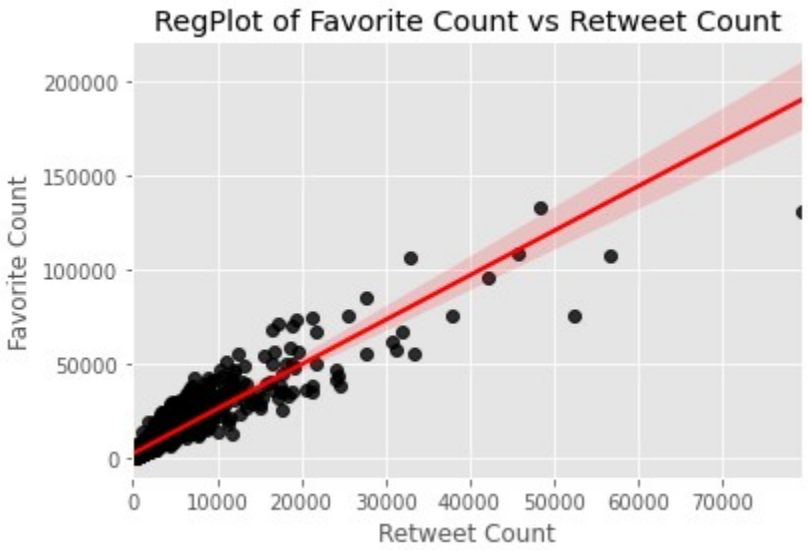
Let's move on to some visualization where i observe that the popular dog type is pupper with the value 264, doggo is 93, puppo is 37, and floofer been the least with value 4

Bar chart of most common dogs



There is a positive correlation between favorite (“like”) counts, and how much a post was retweeted. This correlation is important for the owner of the WeRateDogs twitter account to understand when determining method to increase users’ traffic on the page. A data analysis team could recommend previous posts with either a high retweet counts or high favorite count so that page owner could model future posts off historically popular posts.

Scatter plot of Favorite vs Retweet Counts



Conclusion:

The write up offers a straight look at the data wrangling process. There is so much more that can be done with this data set.