Wrangle and Analyze Data @WeRateDogs

Analysis Report

By

Qusay Elewy

November, 2021

Introduction

WeRateDogs is a Twitter account where people can post photos of their dogs allowing the account followers to rate them. Tweet data as well as supporting data have been provided in this project for us to wrangle (clean and transform) and prepare for analysis.

Once our data has been wrangled, we can use it to get some insights. In this document, we demonstrate how we can use our wrangled data to produce some useful insights.

Analysis

First off, let's check the correlation between the number of likes (favorite) a tweet receives, and the number of retweets. From the scatter plot below, we can see that there is a positive correlation between these two variables. That is, people tend to retweet the tweets they like.

With further analysis, our data shows that tweet #348 is the most liked tweet, which has received 132,810 likes. On the other hand, tweet #1116 has received the lowest number of likes accounting for 52 likes only.

Our analysis also shows that tweet #247 has not been retweeted at all, followed by tweet #1116 which has been retweeted only twice.

Correlation between favorite count and retweet count
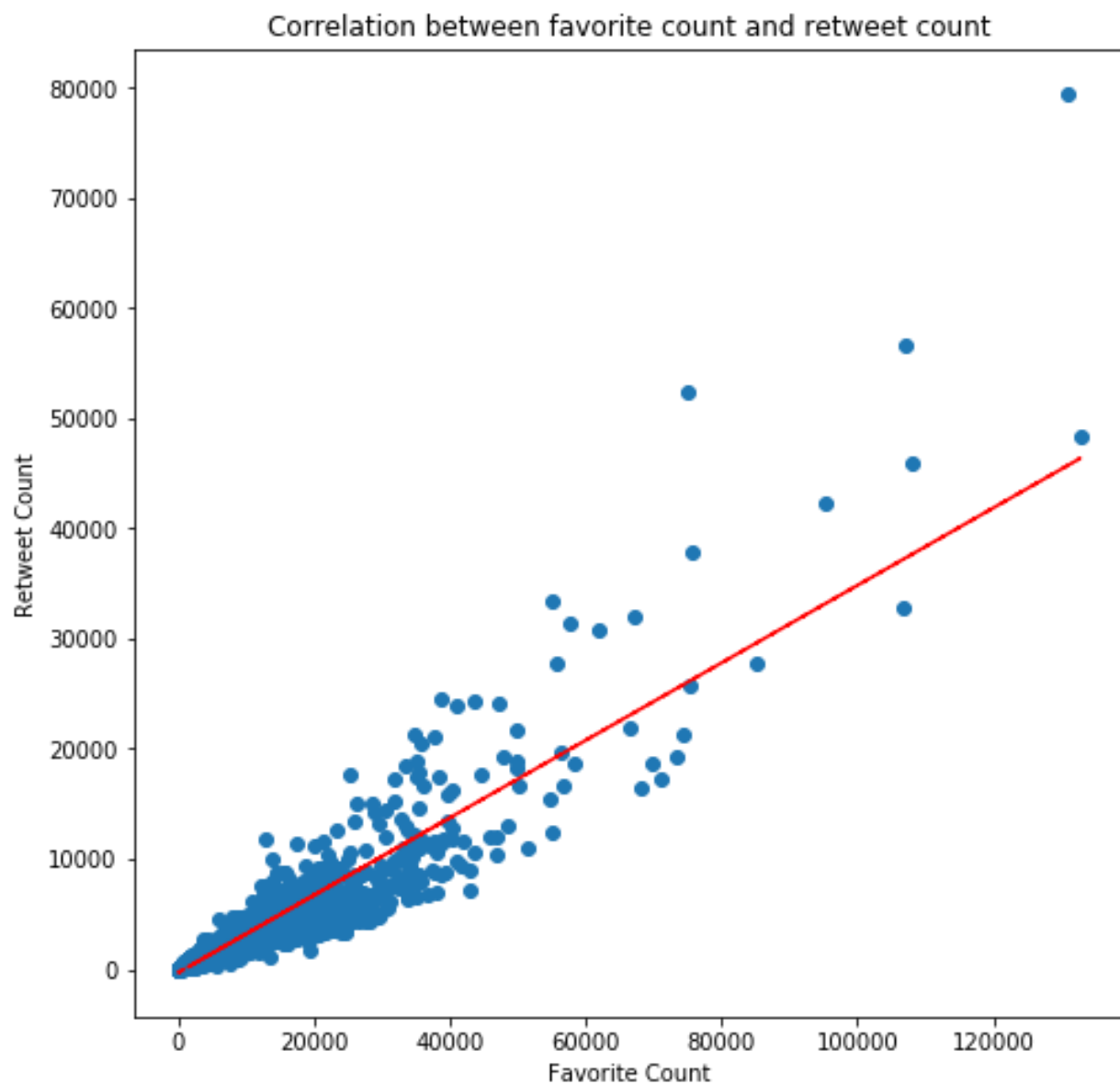
The horizontal bar chart below shows that Golden Retriever is the most tweeted dog breed in our dataset accounting for 139 tweets, followed by Labrador Retrievers accounting for 95 tweets.

Top 10 Dog Breeds
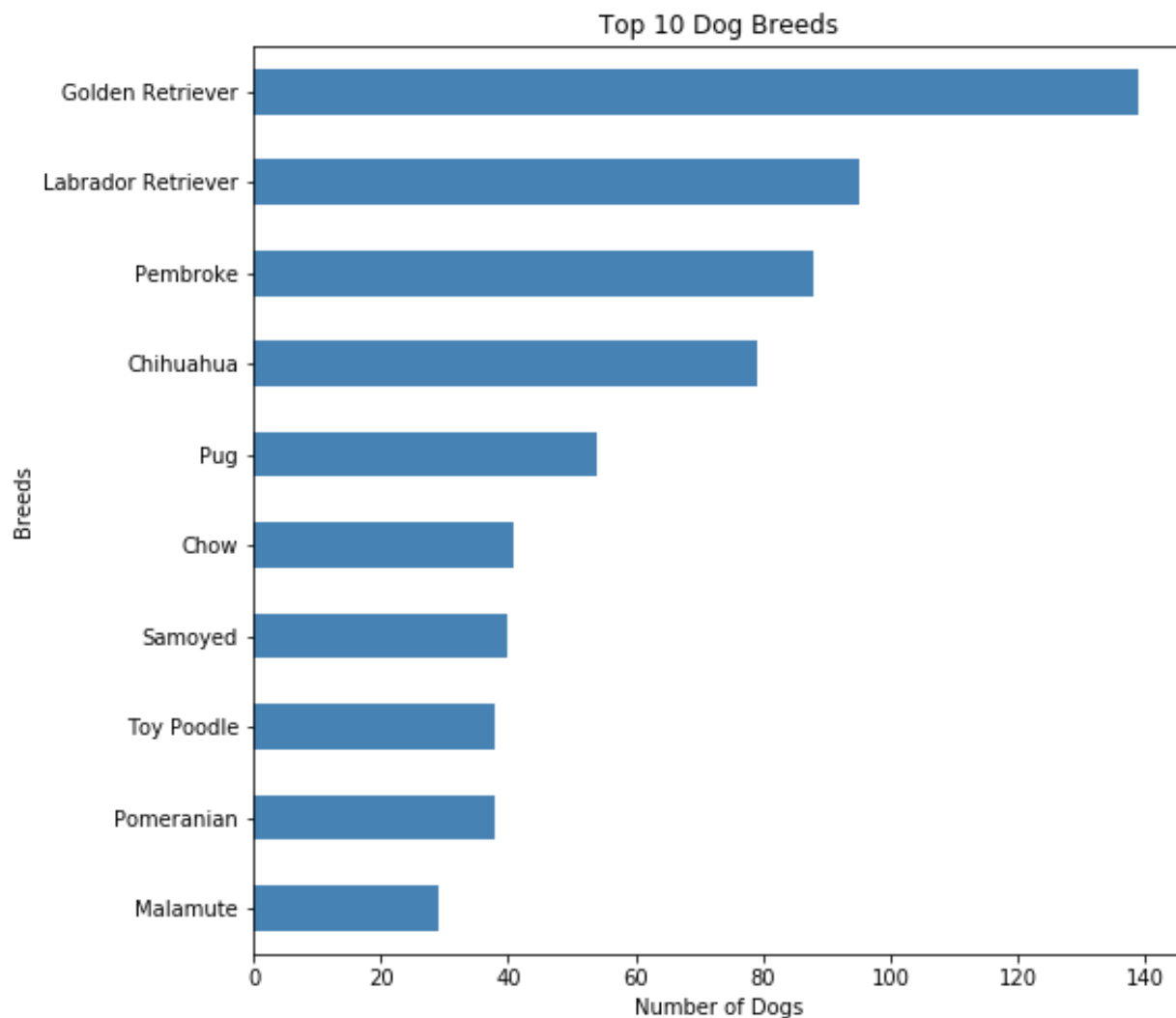
The data also shows that Puppers are the most dog stage tweeted accounting for 224 tweets in our dataset.

The data also shows that both Lucy (for female dogs), and Charlie (for male dogs) are the most used name dog names, each accounting for 11 tweets.

Percentages of Dog Images vs. Other Images
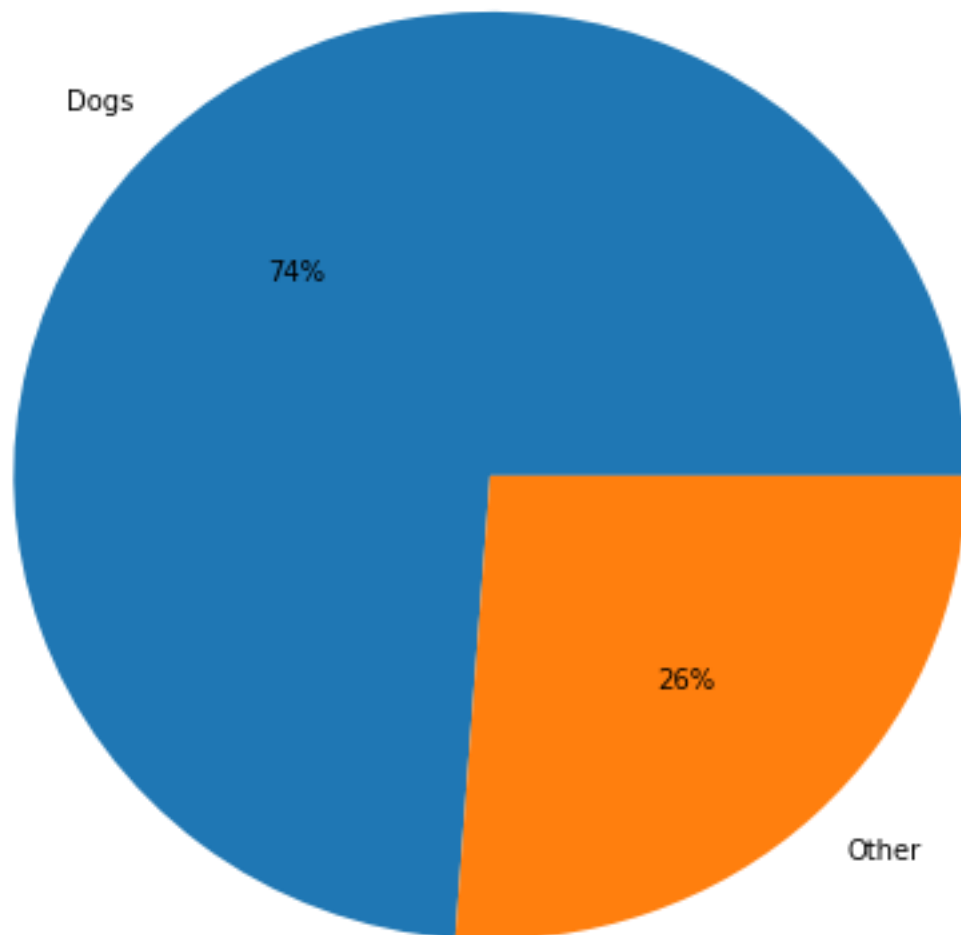
And finally, joining archive table with prediction table allows us to get an idea about the quality of our prediction algorithm. For example, the pie chart above suggests that only 74% of the images in our dataset are for dogs. However, when we checked some of the remaining 26% records, it turned out that the algorithm presented some dog images as other objects. For example, the classifier considered a dog in a shopping cart as a shopping cart, and not a dog. With this information, we can see that the used image classifier algorithm may need some improvement.