Wrangle Report

In this project we wanted to Wrangle and Analyze data from the twitter's account WeRateDogs (@dog_rates). WeRateDogs account received international media attention for its singular way to rate other people dog's photos and because it was suspended for Twitter for breaking social media copyright law. The first step was to gather data from Udacity and Twitter API. Then I cleaned and analyzed it using Python libraries.

Favorite x retweet

There is a positive correlation between favorite count and retweet count.

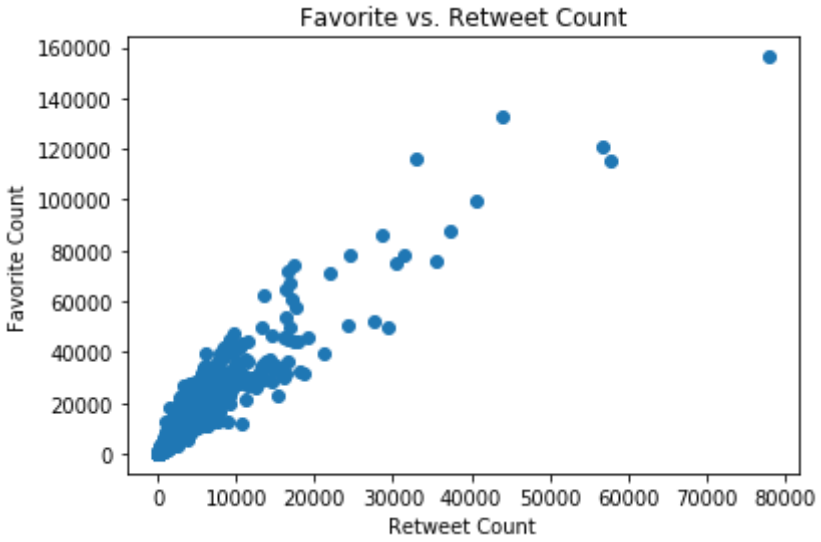

Figure 1: relationship between favorite count and retweet for twitter account @dog_rates.

To study this I used the previous lesson and did the fit from these two variables. I found that 86.3% of the variability in retweet count can be explained by the favorite count for WeRateDogs account and The p-value of 0.00 associated with favorite count suggests that it is statistically significant in providing information in predicting the retweet counts.

Dog's names

The top 8 names from dogs that has its photos posted in WeRateDogs are Cooper, Oliver, Charlie, Penny, Lucy, Tucker, Winston and Sadie. The figure 2 show this names and the amount of times they repeated.

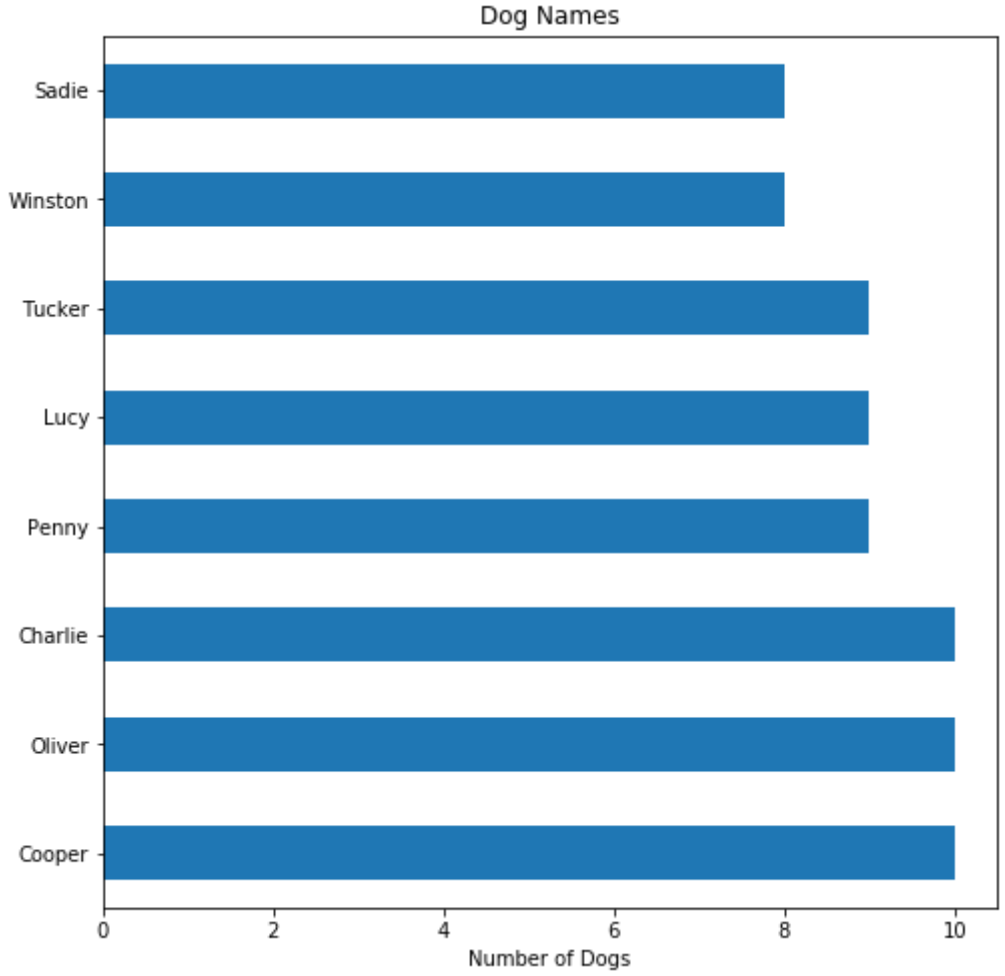

Figure 2: Most common dog's names in WeRateDogs.

Dog's brand

The most predict brand of dog in first prediction was golden retriever followed by the labrador retriever. The prediction tables were a little weird because there was some values that were not dog's brand so I did not take it so serious.

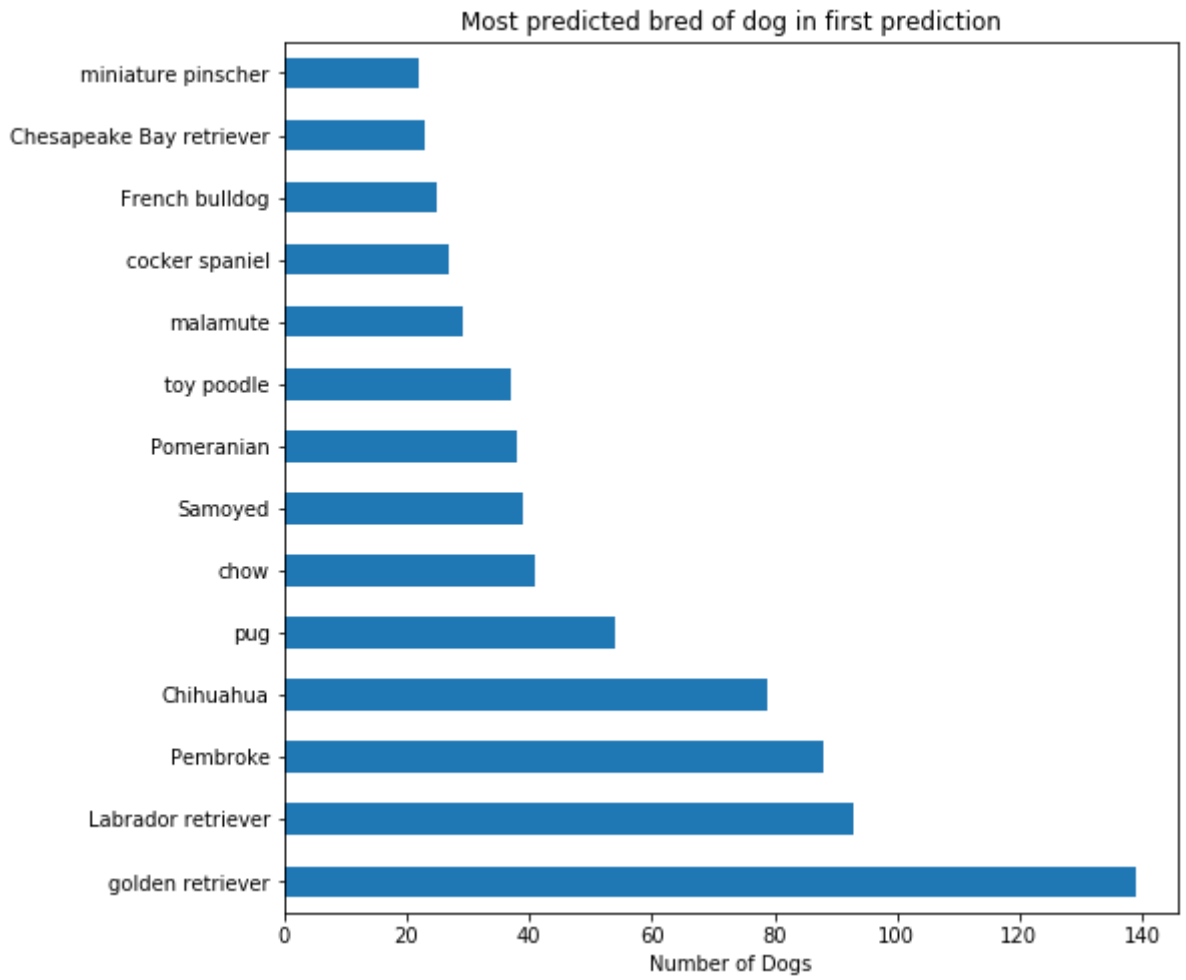

Figure 3: Most common dog's brand in first prediction.