

WeRateDogs Project: Insights Report

Data Analyst Nanodegree Program

Mário Damhur

March 31, 2020

1. Introduction

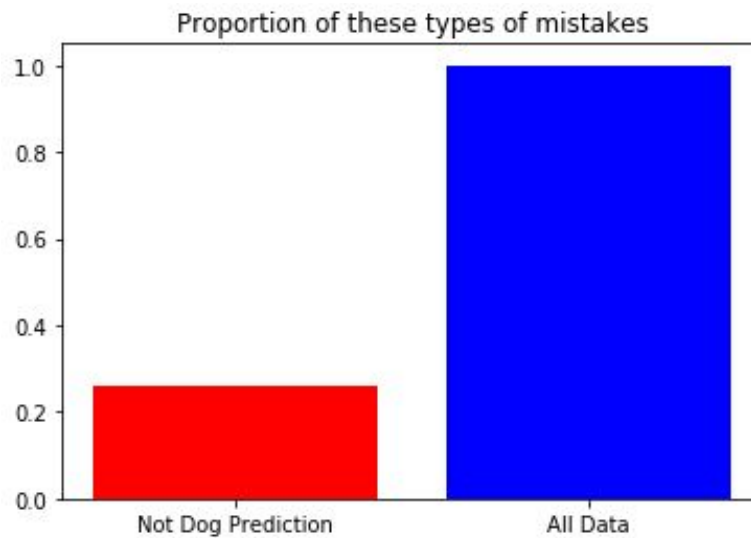
I will discuss some insights that I had for understanding the dataset in this project, starting with why the model made mistakes, followed by the correlation between numeric features and finishing with the top 10 dog most popular dog names. Let's dive in!

Insight 1

Let's check the mistakes when the top 1 prediction of the model didn't consider the image as a dog. For this I will set:

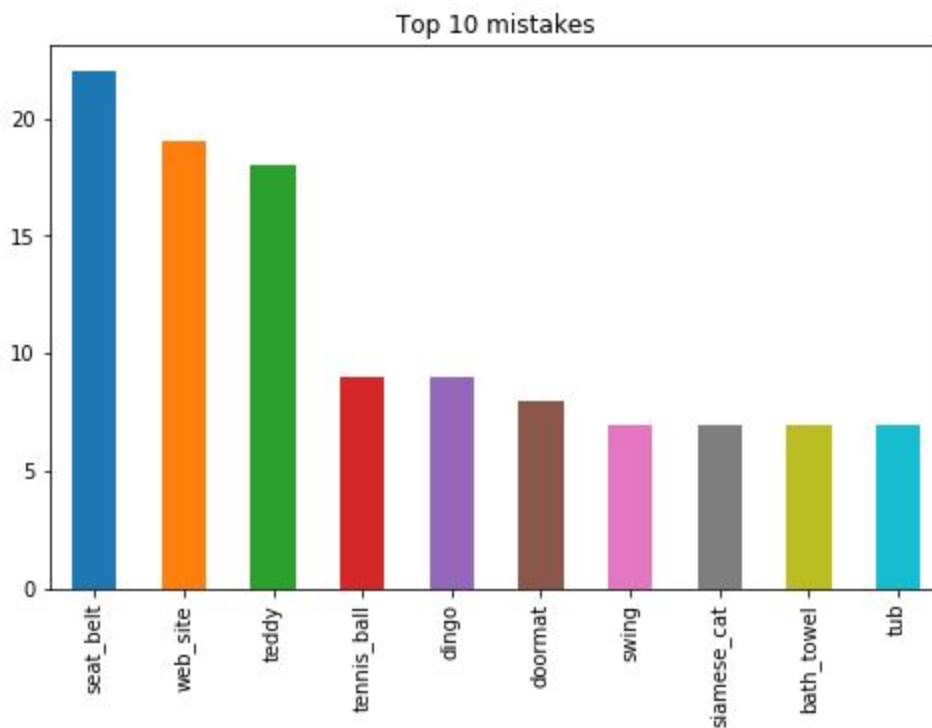
- **False** to prediction_1_isdog column

Why didn't the model consider the image as a dog?

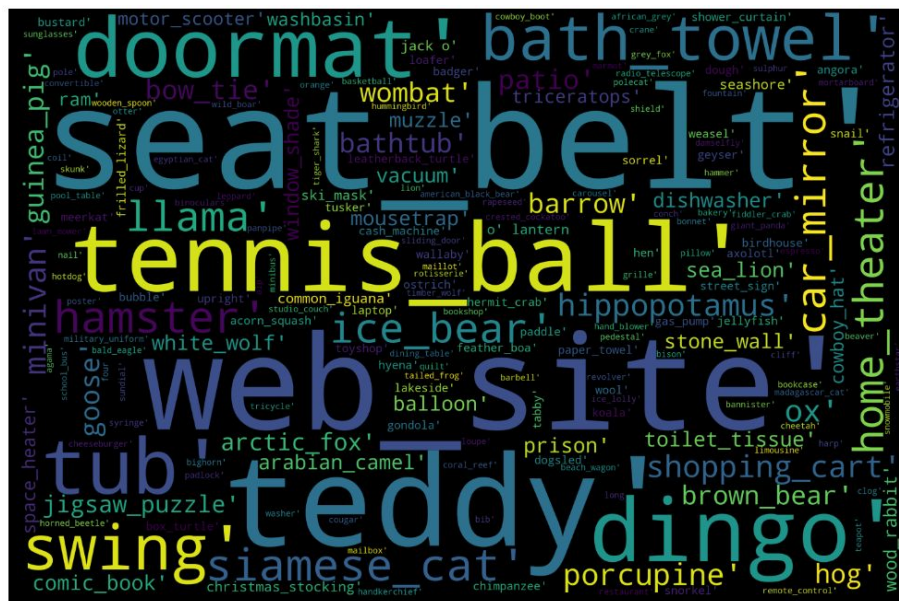


The proportion of the mistakes made by the model that considers the images as not dogs is about 1/4 (0.2614) of all the dataset. Since we don't have information about the model, we can't analyze deeply but we can analyze the images of those mistakes.

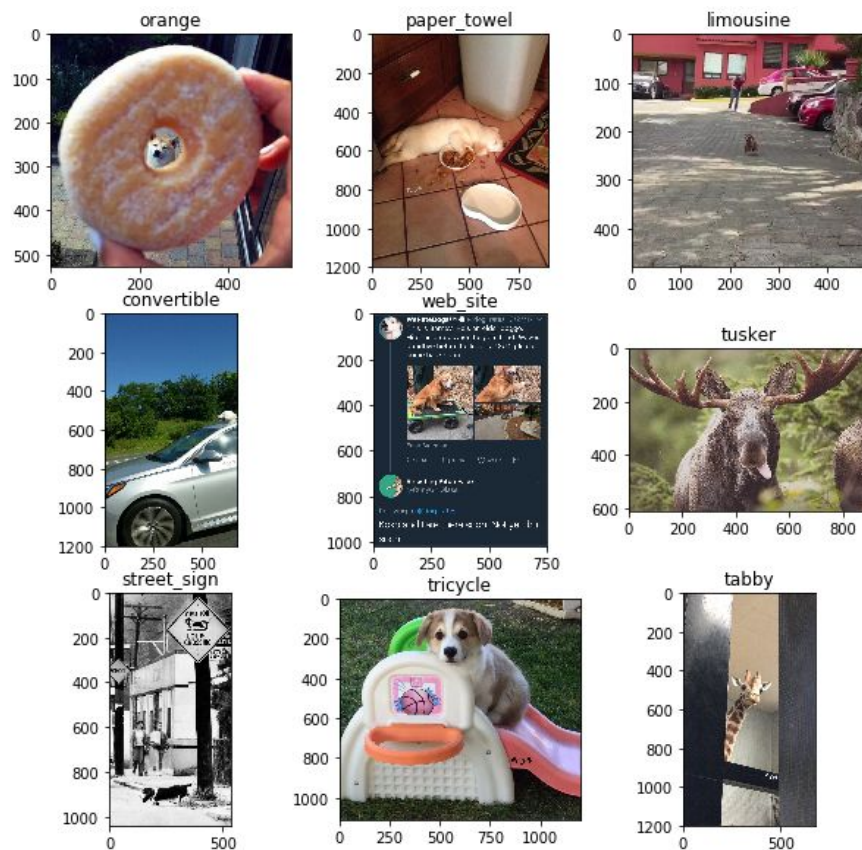
Let's check the top 10 prediction mistakes made by the model and their frequency.



Allright, now let's check the word clouds of the prediction mistakes.



Now we can check some images that the model predicts incorrectly.



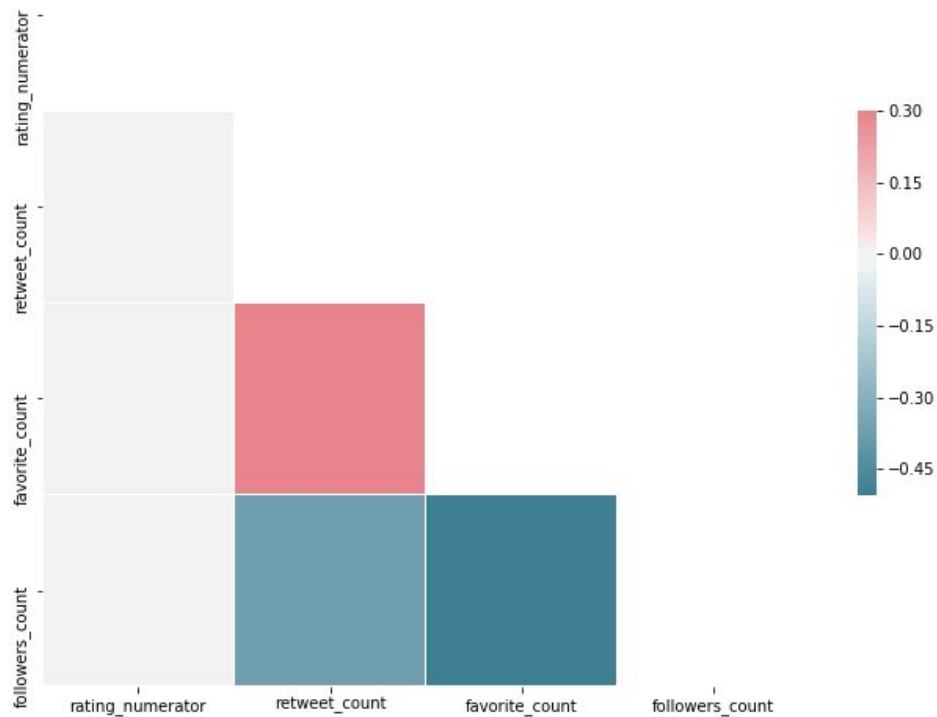
With those images, it is clear that some images are very hard for the model to predict correctly. The object predicted is part of the image and the model looked only at this object in the all image. Some of them are outliers, that is, images not related with dogs, but other animals.

Insight 2

Let's check the correlation of numeric features. For this we will consider the columns:

- rating_numerator (Since the ranking is done by this variable in principle)
- retweet_count
- favorite_count
- followers_count

There is a correlation between the numeric features of our dataset?

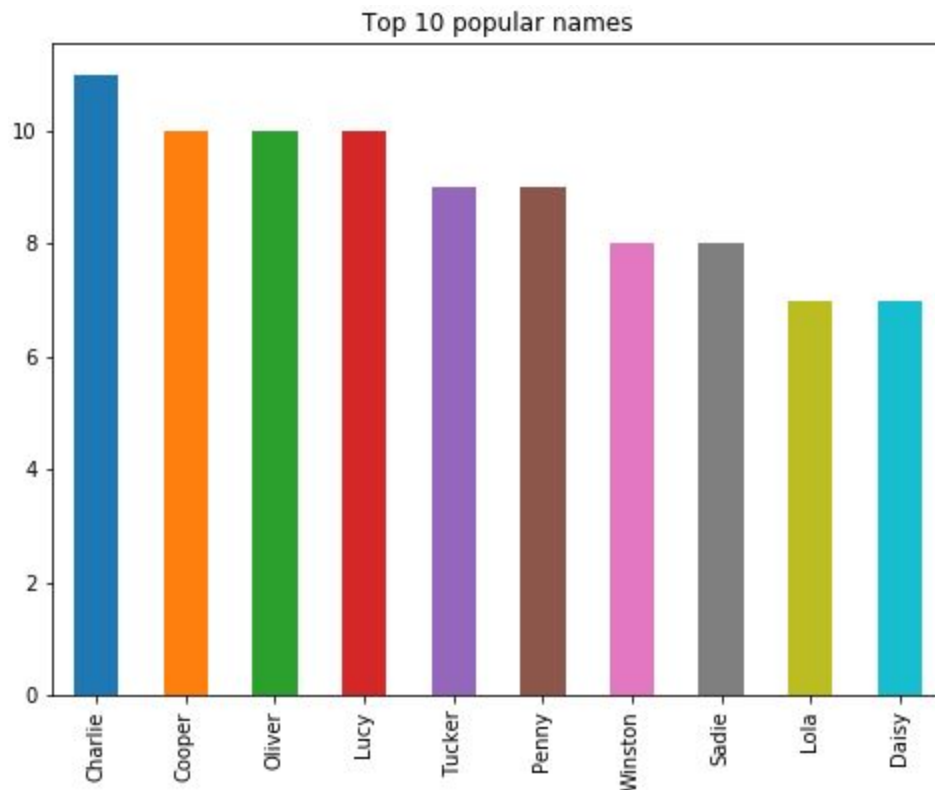


We can observe that favorite_count feature have positive moderate correlation with retweet_count. The favorite_count have negative moderate correlation with followers_count and retweet_count have negative weak correlation with followers_count. None of them have correlation with rating_numerator.

Insight 3

Let's check the most popular dog names.

What are the most popular dog names?



The most popular dog name is Charlie, followed by Oliver, Cooper, Lucy and so on. Other names like Sadie, Penny, Lola are also popular.