Report

Wrangling and Analyze Data

From Michael Eibner

With Twitter Data from WeRateDogs

A Udacity Project

Insights and Visualizations

Insights and Visualizations 1

Atticus:

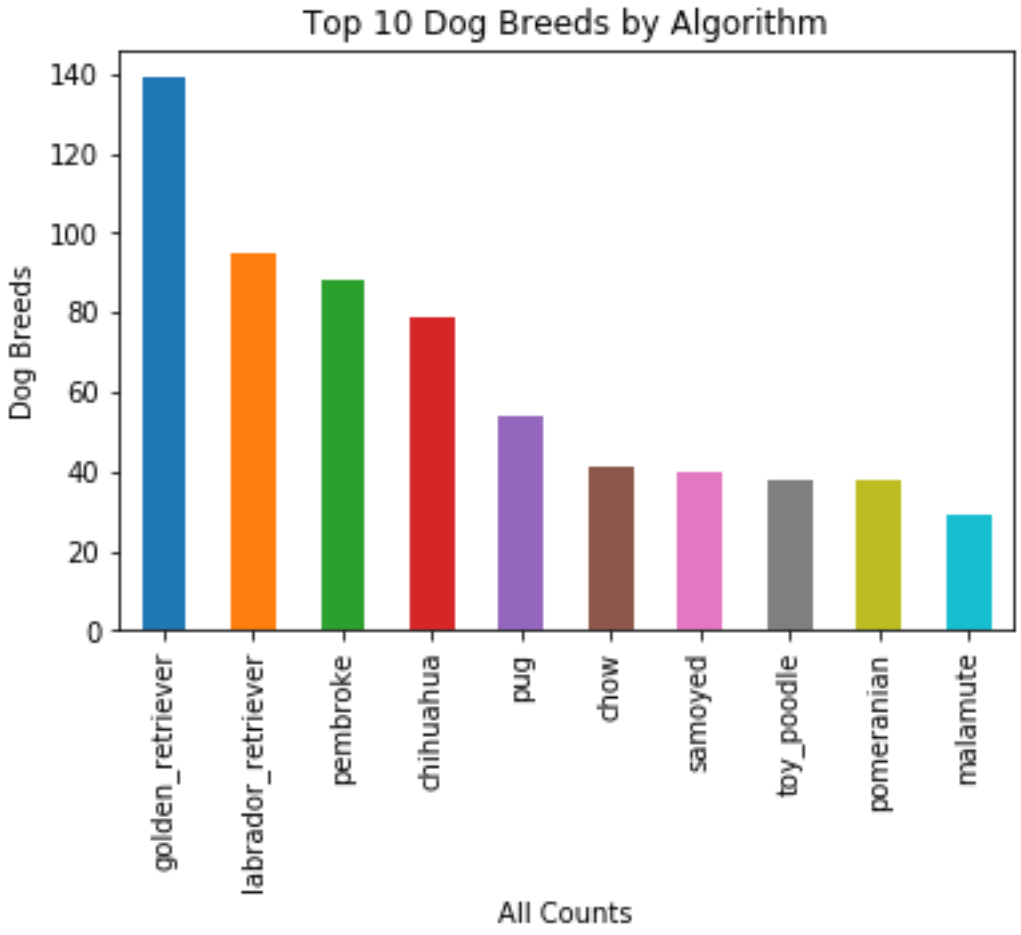
After I cleaned the rating system of the data, I was curious what the Top 10 of the ranked dogs will be and I found it out. It's Atticus the American dog! *"He's quite simply America af…"*

To get this information, I ranked all ratings from highest to lowest and I set the limit to 10, so I got the Top 10 List. After that I searched for the dog with the highest ranking. With distance, it's Atticus (rating of 177.6).

Insights and Visualization 2

Predicted Dog Breeds from the Algorithm:

The Algorithm was predicting the Twitter images. I wanted to know what the most common predictions are.

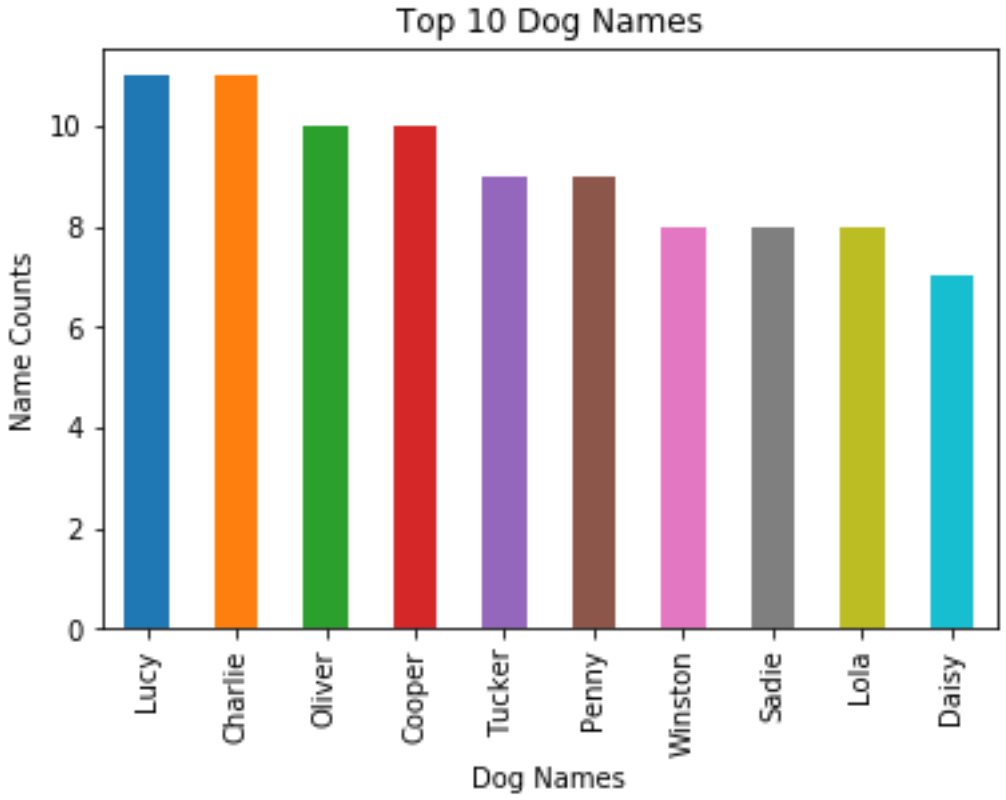


You can see, the Golden Retriever is, with distance, the most common predicted dog breed in the data.

Insights and Visualization 3

Dog Names:

Every Dog has its special and wonderful name. But not every dog has an unique name. So I found out what the most common dog names are.

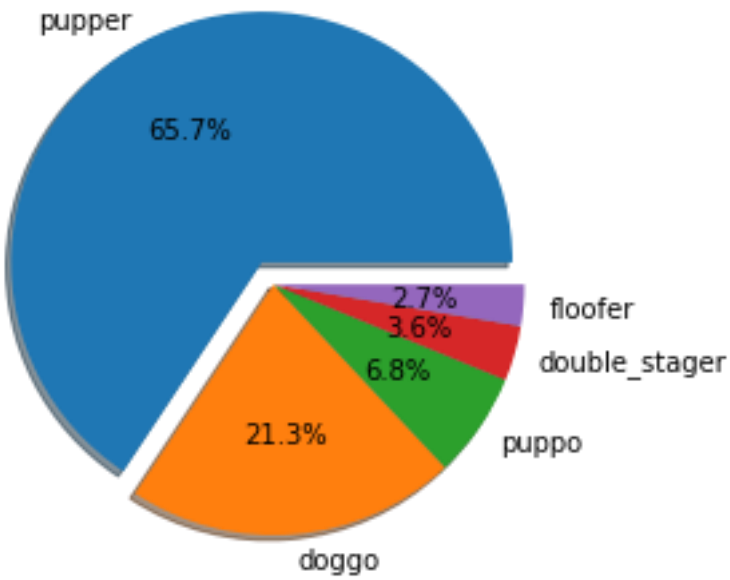


The surprise is the names are well distributed. Maybe the owner of the dogs are more creative than I thought. Charlie is the most common name, closely followed by Oliver, Lucy, and Cooper.

Insights and Visualization 4

Dog Stages:

Some dogs in the data are called “doggo”, “floofer”, “Pupper”, “Puppo” or even more than only one of it. I want to know, how distributed the dog stages are.

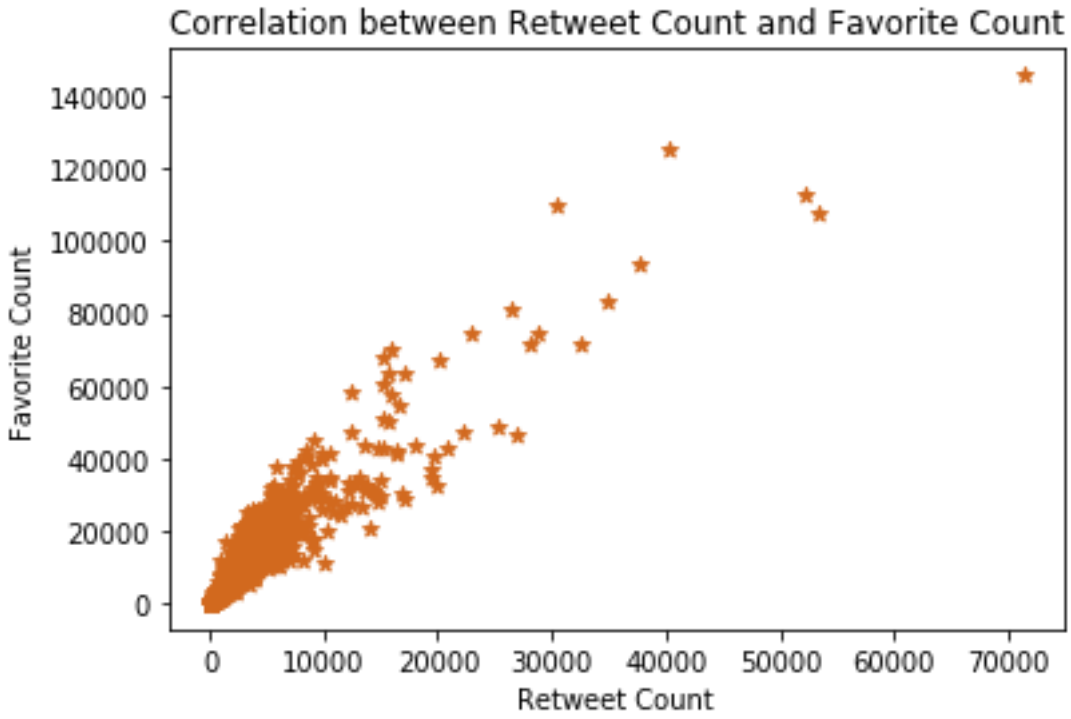


As we can see, the dogs are called “Pupper” more often than the other ones combined. The “Double Stager” are even more often called than the “Floofer”. Are there so less “floofy” dogs?

Insights and Visualization 5

Correlation between Retweet Count and Favorite Count:

I wanted to know if there is a correlation between the Retweets and the Likes of the dogs.



And as we can see, if there is no Favorite/Like, there are anyway Retweets. But you also can see a slightly correlation between both.