Data Analysis

Data Analysis and Visualization Documentation

By alaa elhariry

Introduction

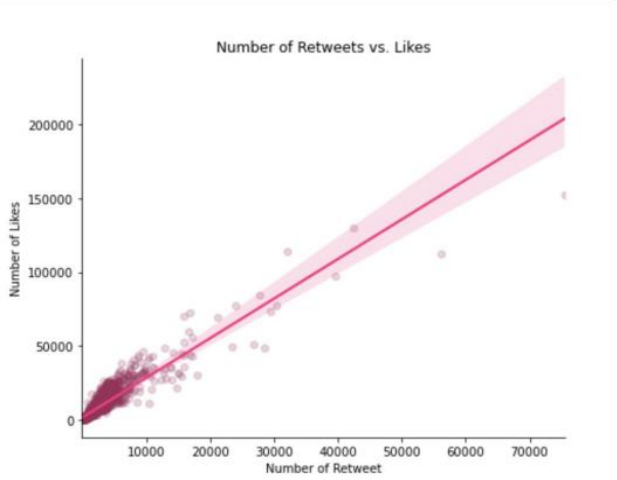
We Rate Dogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." We Rate Dogs has over 4 million followers and has received international media coverage. The goal of this part of the project was to wrangle We Rate Dogs Twitter data to create interesting and trustworthy analyses and visualizations.

The Addressed Questions

- What is the relationship between the number of retweets and likes?
- How many dog breeds the algorithm was able to predict?
- How does each variable correlates with the other?
- Where do most of the numerators lie?
- What is the most common dog stage?
- Which dog stage received the highest likes and retweets?
- What is the most common dog breeding the algorithm predicted, and how confident was it?

To answer these questions, we used different visualizations and methods.

What is the relationship between the number of retweets and likes?

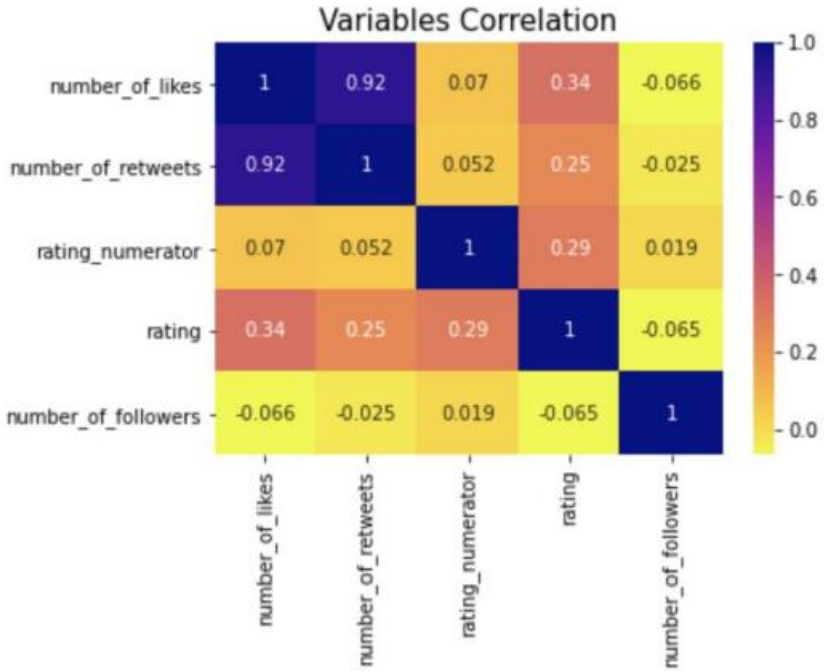


The plot above shows that there is one retweet for every four likes. The majority of the data lies below forty thousand likes and ten-thousand retweets, and as we can see, the number of retweets and likes are correlated positively.

How many dog breeds the algorithm was able to predict?

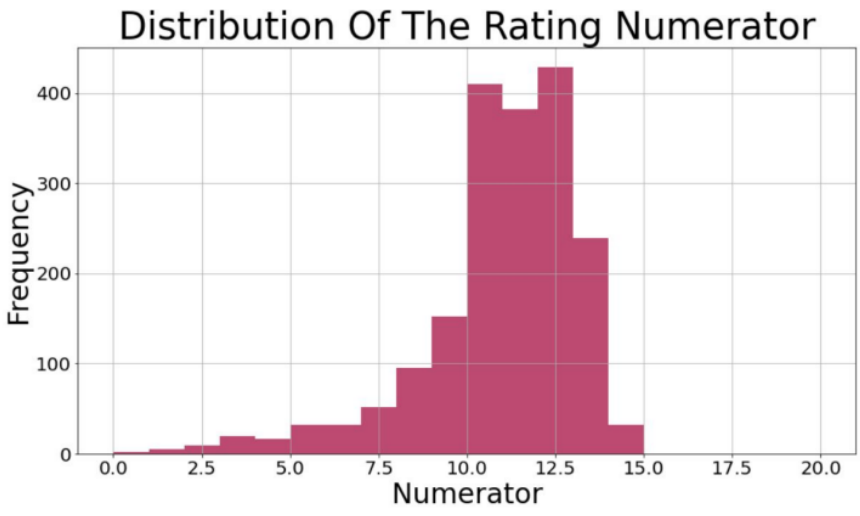
The algorithm was able to predict 1620 dog breeds and failed to do so in 300 cases.

How does each variable correlates with the other?



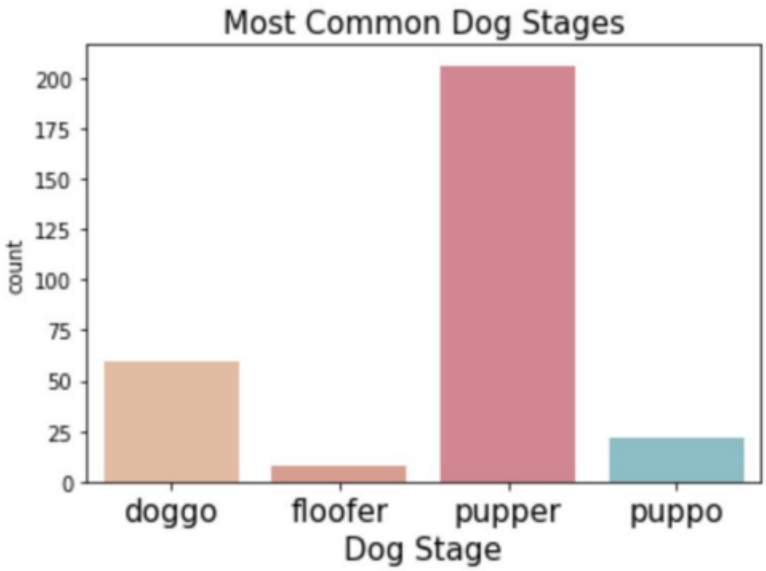
The strongest obvious correlation is between likes and retweets, which is of 0.92. While there is a negative weak correlation of -0.025 between the retweets and the followers. Also, it is normal for rating to correlates with the rating numerator since the latter actually has been used to calculate it.

Where do the majority of the numerators lie?



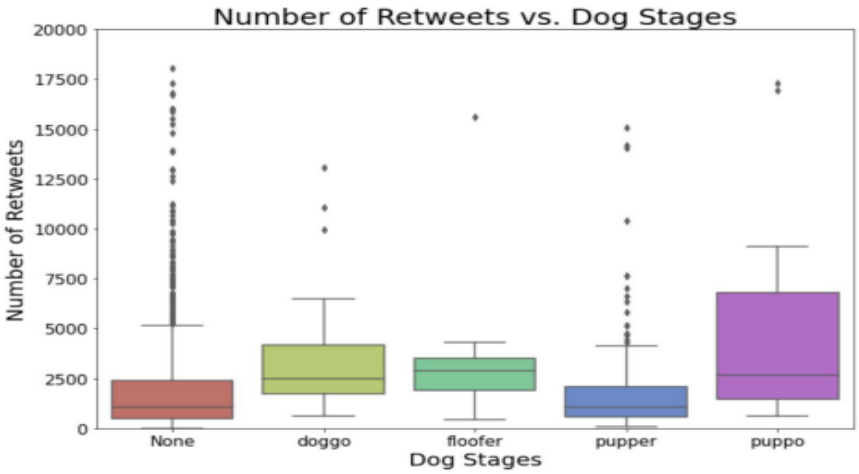
As the plot shows, the majority of the numerators lies between 10 and 13.

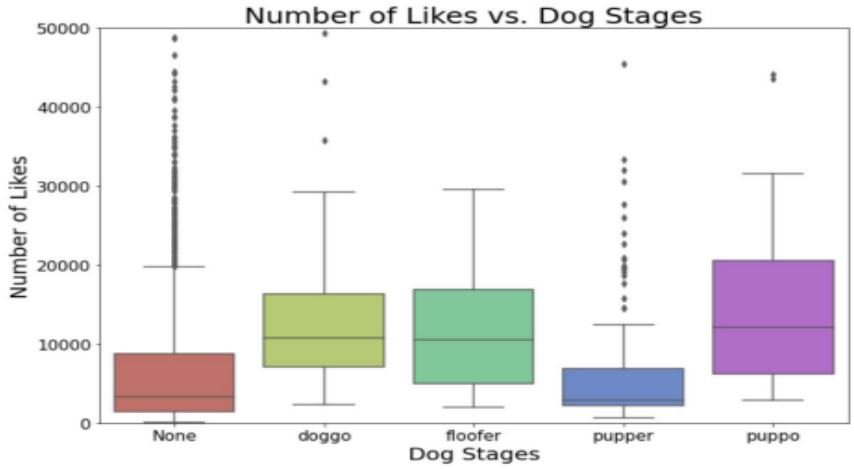
What is the most common dog stage?



As it shown above, pepper is the most common dog stage as it appeared 206 times in the dataset. Next comes doggo, which appeared 59 times

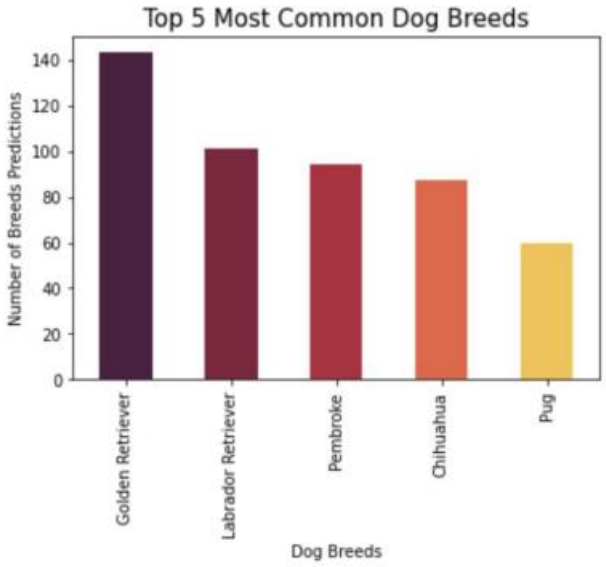
Which dog stage received the highest likes and retweets?

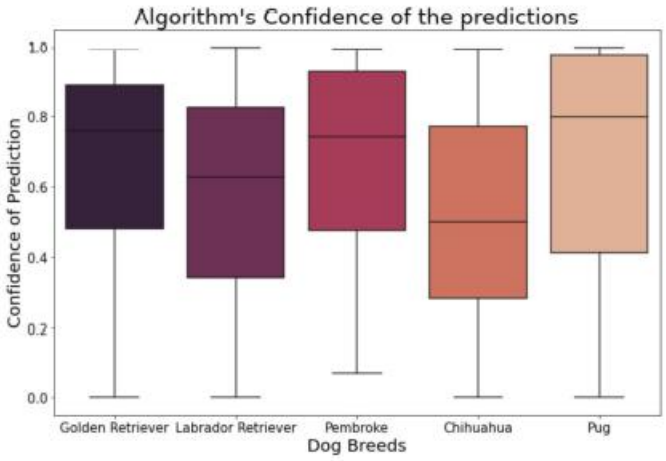




As it shown above, the dog stages with the highest number of retweets are puppy then floorer. While the puppo then doggo received the higher likes. Overall, pepper is the least in both number of likes and retweets

What are the most common dog breeds the algorithm predicted, and how confident was it?





The top five most common dog breeds are Golden Retriever, Labrador Retriever, Pembroke, Chihuahua, and Pug. Obviously, the conference of predicting each of these dogs varies from one to the other. The algorithm was almost sure of predicting the dog breed for some dogs, while in some cases it was pretty unsure, such as The Chihuahua, where its median confidence was about 50%.