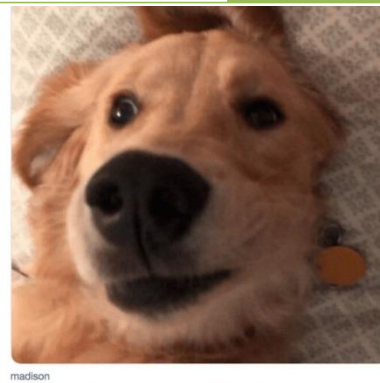


'WeRateDogs' Data Analysis



Imesha Kuruppu

Udacity

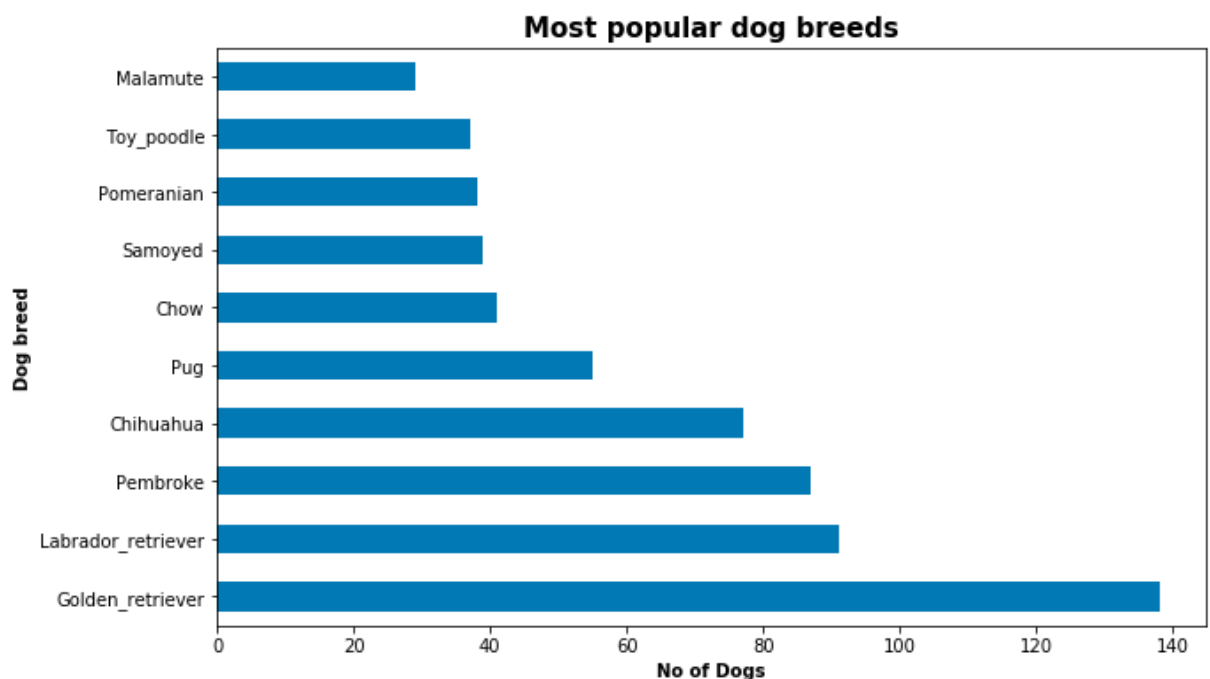
4/28/2021

Introduction

'WeRateDogs' 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. But, the numerator is always greater than 10. 11/10, 12/10, 13/10, etc. This rating system is unique to the 'WeRateDogs'. This unique rating system is a big part of the popularity of WeRateDogs.

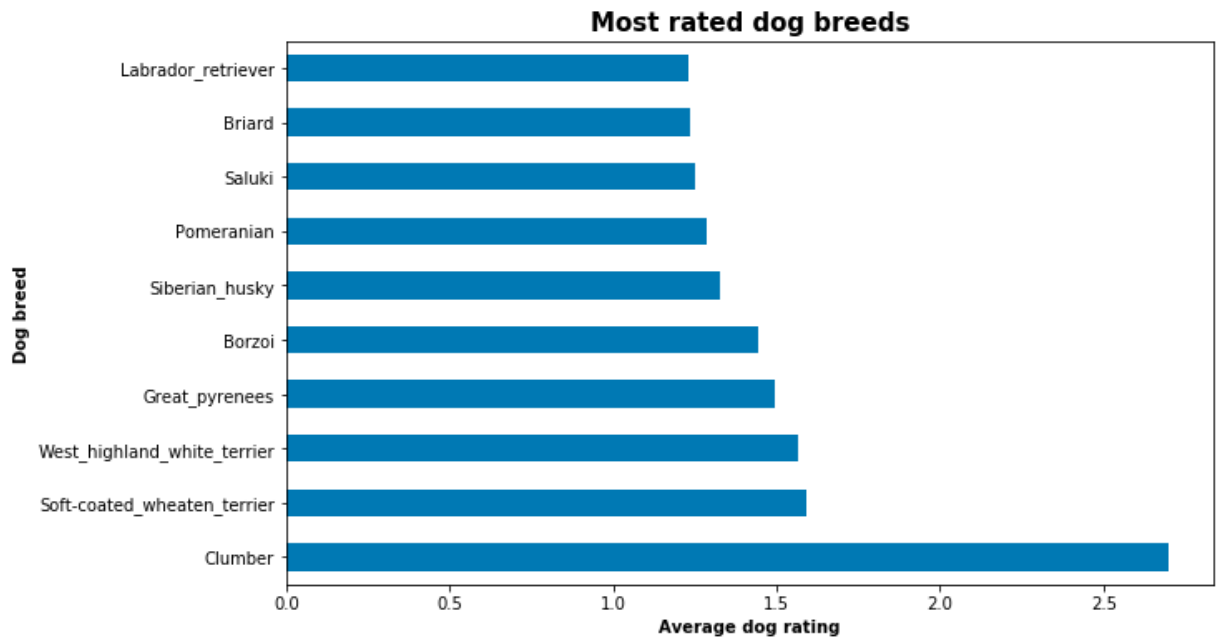
The main goal of this report is to analyze the gathered and wrangled WeRateDogs Twitter account data to find interesting and trustworthy insights. This dataset includes 5961 rows and 12 columns. One row contains the data related to a tweet. That is, tweet_id, timestamp, dog_name, dog_stage, image URL, image prediction details, image favorite & retweet counts, and dog rating. This analysis is based on the correct first prediction of the dog breed.

1) What are the most popular dog breeds?



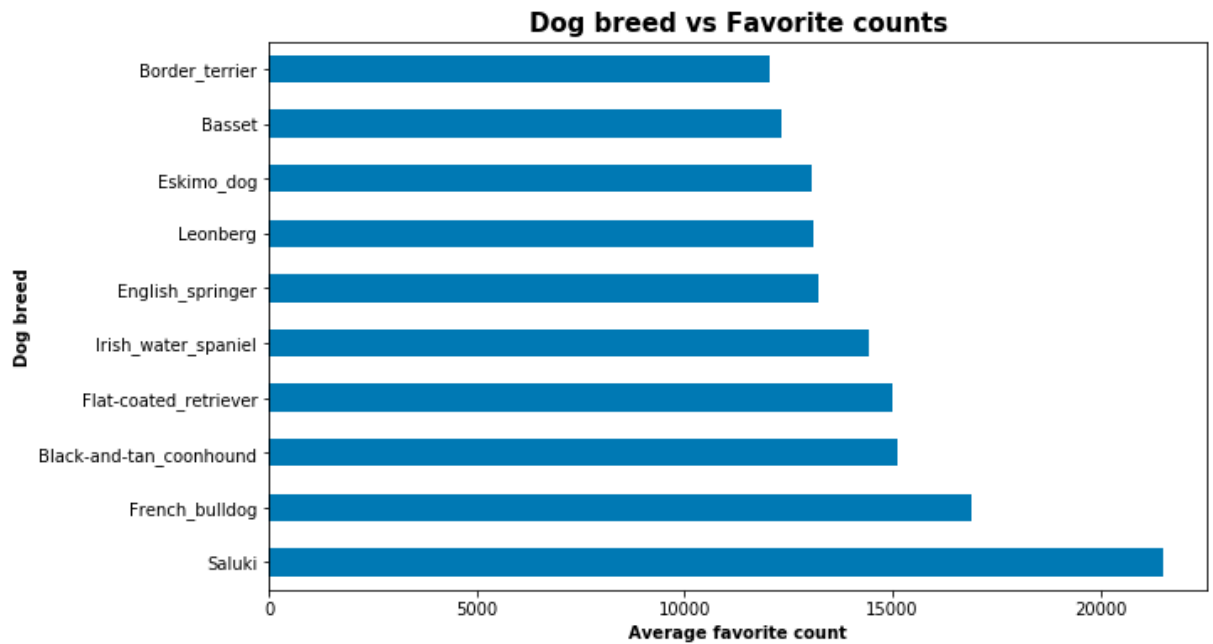
According to the above graph, the most popular dog breed is Golden_retriever. Other popular dog breeds respectively are Labrador_retriever, Pembroke, Chihuahua, Pug, Chow, Samoyed, Pomeranian, Toy_poodle, and Malamute.

2) Which dog breed has the highest ratings?



The highest dog rating was gained by the 'Clumber' dog breed. Other highest rated dog breeds are soft_coated_wheaten_terrier, West_highland_white_terrier, Great_pyrenees, Borzoi, Siberian_husky, Pomeranian, Saluki, Briard, Labrador_retriever.

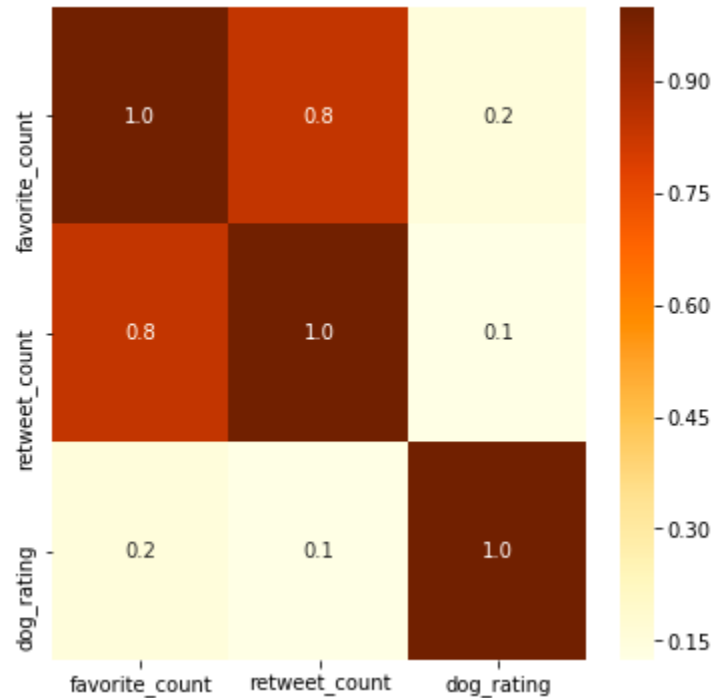
3) Which dog breed has the highest favorite counts?



The highest favorite count was gained by the 'Saluki' dog breed. Other dog breeds that gained the highest dog breed are French_bulldog, Black_and_tan_coonhound, Flat_coated_retriever, Irish_water_spaniel, English_springer, Leonberg, Eskimo_dog, Basset, and Border_terrier.

4) Is there a relationship between favorite counts, retweets, and dog rating?

Correlation coefficient between favorite count, retweet count and dog ratings



There is a positive correlation between the favorite count and retweet count as the correlation coefficient between them is 0.84(value close to 1). That means when the favorite count is increasing for an image retweet count also increases. There is no relationship between dog_rating and the other two variables.

5) Most popular dog names

None	1821
Oliver	30
Cooper	30
Charlie	30
Tucker	27
Penny	27
Lucy	27
Sadie	24
Winston	24
Toby	21

In the 1821 rows, the dog name was not defined. Except for those data most popular dog names are Oliver, Cooper, Charlie, Tucker, Penny, Lucy, Sadie, Winston, and Toby.

Conclusion

According to the above analysis, we can see that the most popular dog breed (Golden retriever), highest rated dog breed (Clumber) and dog breed (Saluki) that received the highest favorite counts are different from each other. From the result of the heatmap, we can conclude that there is a strong positive relationship between favorite count and retweet count variables. But there is no relationship between dog ratings and the favorite count or retweet count. In this data set, the most popular dog names are Cooper, Charlie, and Oliver.