

WeRateDogs – Insights into the @dog_rates Twitter page

Introduction and Background

Real-world data rarely comes clean. The dataset wrangled for this project is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

Here's an example:



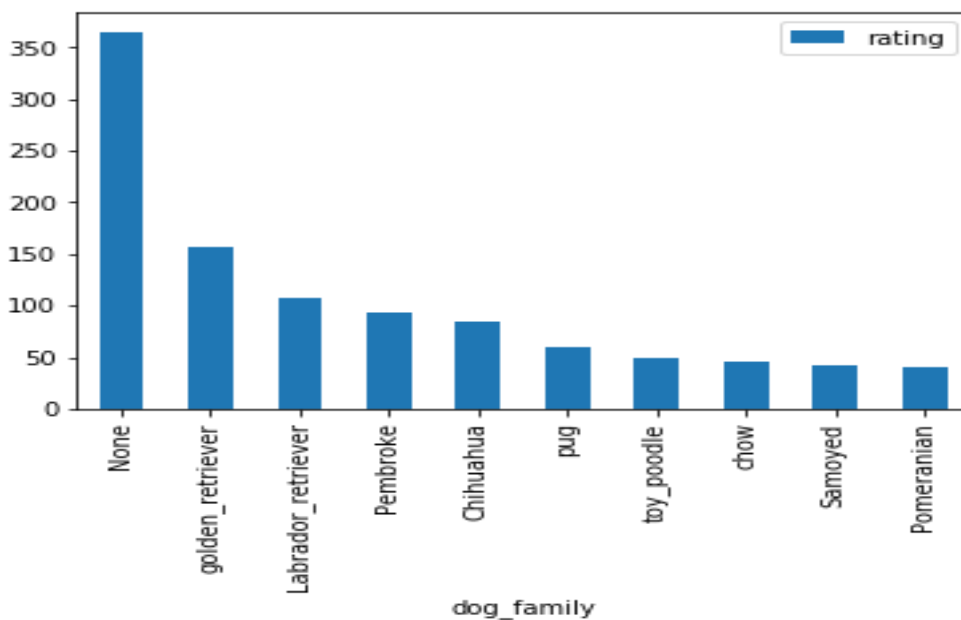
-
- - WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for you to use in this project.
 - This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.
- The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file was provided to Udacity students.
- Twitter API and Python's Tweepy library to gather each tweet's retweet count and favorite ("like") count.
-

Analysis and Visualization

Dog Family Popularity

I chose to analyze and present on four different pieces of the WeRateDogs data set.

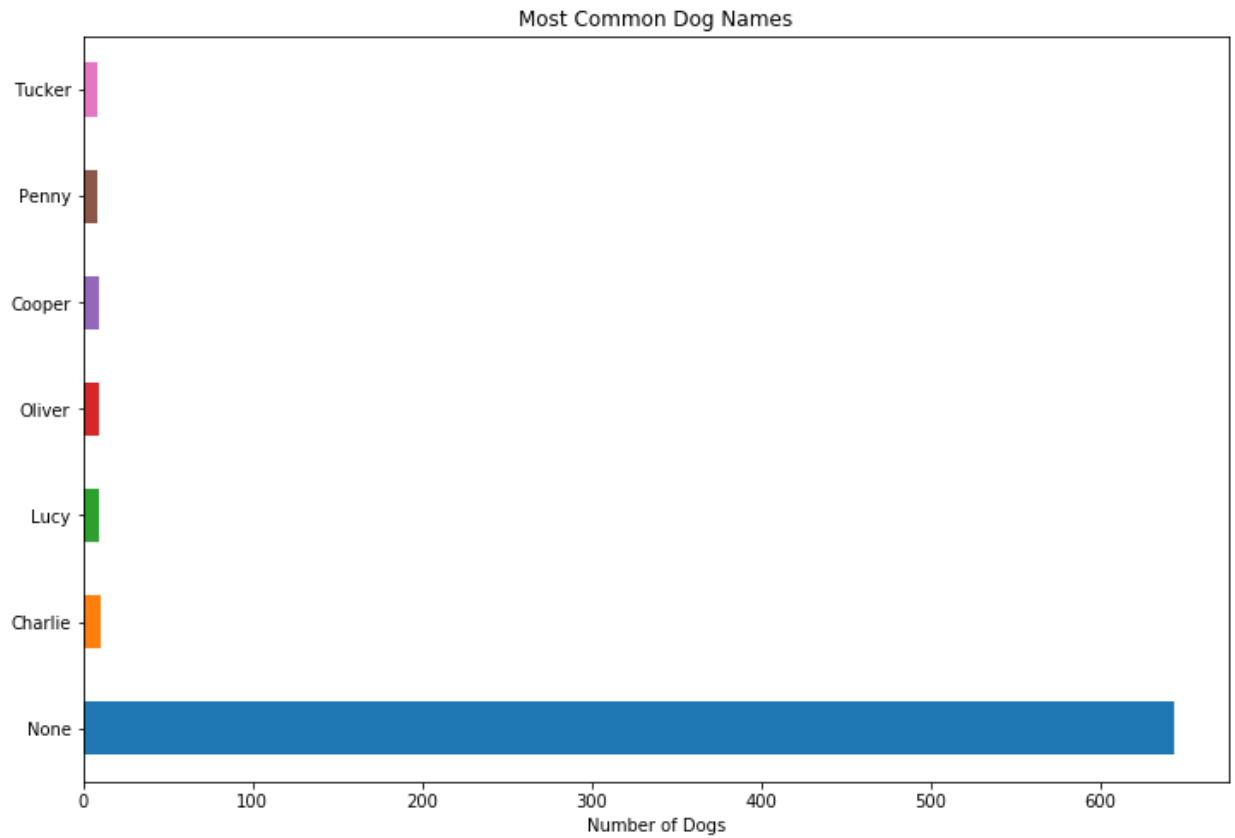
The most popular dog family is a golden retriever (ignoring the None label), with a Labrador Retriever coming in as the second most popular breed. Pembroke isn't far behind. The page owner could use this information to create targeted marketing efforts for certain breeds that aren't as popular to increase their popularity, but also utilize the breeds that are proven to be popular to drive user traffic to the page.



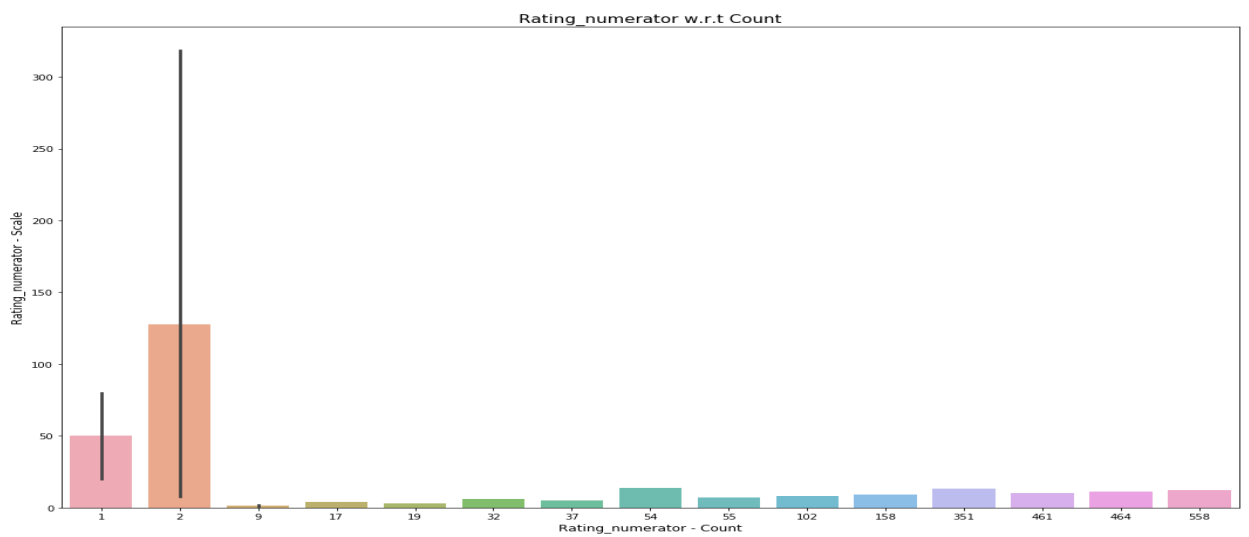
The most popular dog breed is a golden retriever (ignoring the None label), with a Labrador Retriever coming in as the second most popular breed. Pembroke isn't far behind. The page owner could use this information to create targeted marketing efforts for certain breeds that aren't as popular to increase their popularity, but also utilize the breeds that are proven to be popular to drive user traffic to the page.

Dog Name Commonality

Names are important, especially for dogs! The four most popular dog names are Oliver, Cooper, Charlie, and Lucy. But majority of the names are unknown or "None"



- Rating numerators have majority of the ratings around 10



This graph is important for the owner of the WeRateDogs Twitter account to understand when determining methods to increase user traffic on the page. A data analyst team could recommend previous posts with either a high rating count so the page owner could model future posts off historically popular posts.

Conclusion

This write-up offers a straightforward look at the data wrangling process. There is so much more that can be done with this data set, but I encourage aspiring data analysts to dive deep into this data set and see what else you can find!