

Data Wrangling Project Insights - WeRateDogs

By Olamide Emida

Introduction

Udacity: Wrangle and Analyze data

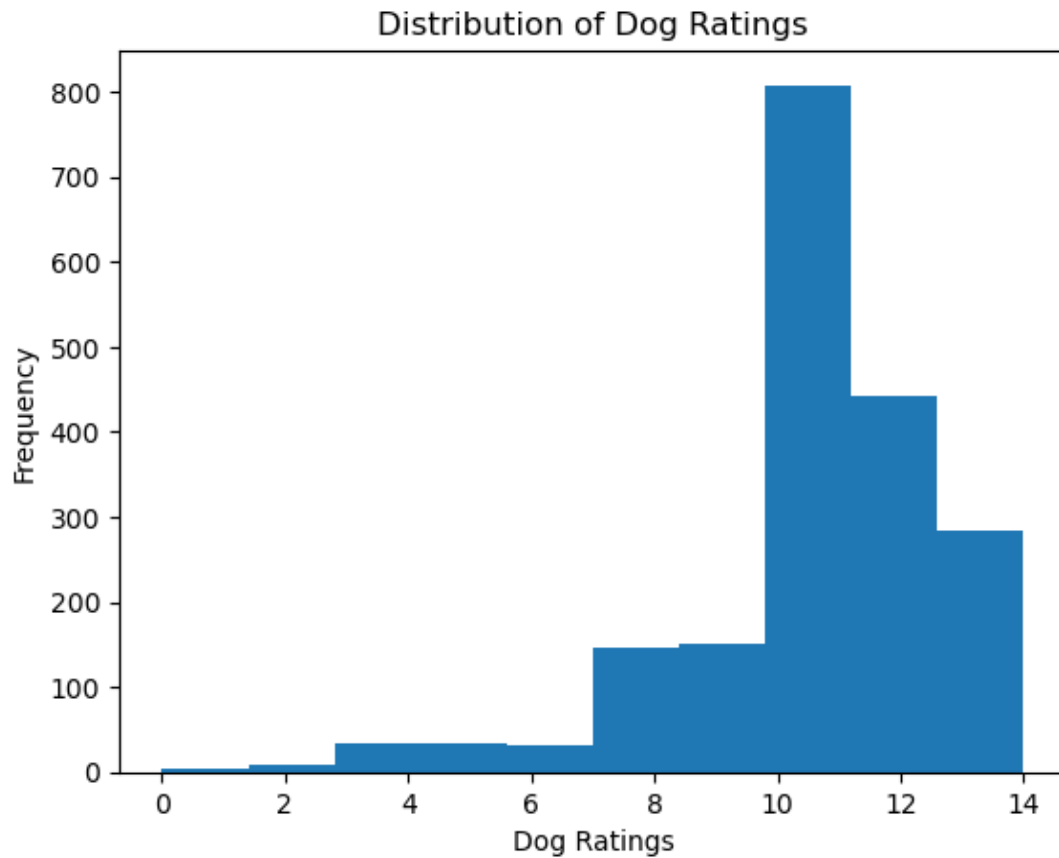
The goal is to provide the visualizations and insights from after the data wrangling of WeRateDogs data. I wrangled WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning was required for worthy analyses and visualizations.

Visualization and Insights

Having done the gathering, assessing, cleaning and storing of the cleaned data. I used the cleaned master data stored in *twitter_archive_master.csv* file to carry my analysis and visualization. I derived the following visualizations and insights:

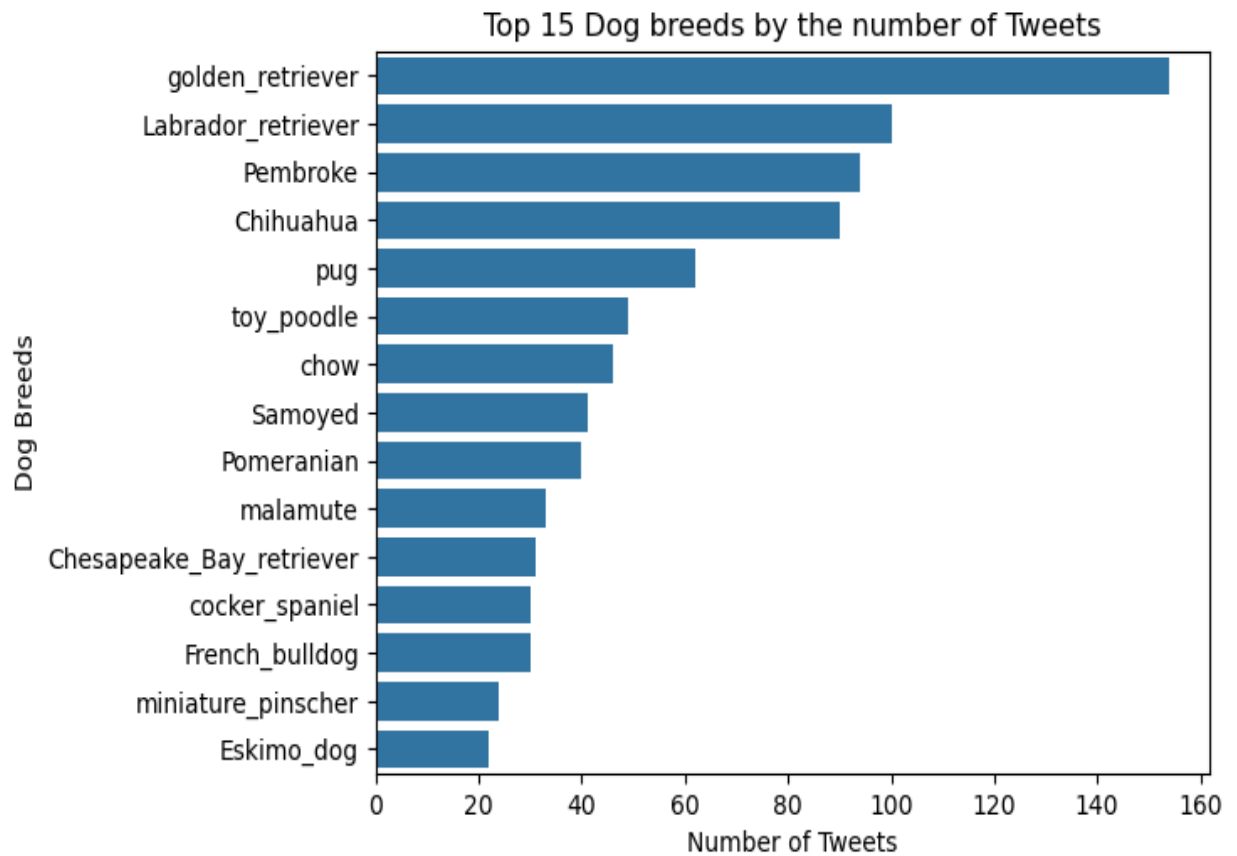
Visualizations

1. What is the distribution of the numerator ratings?



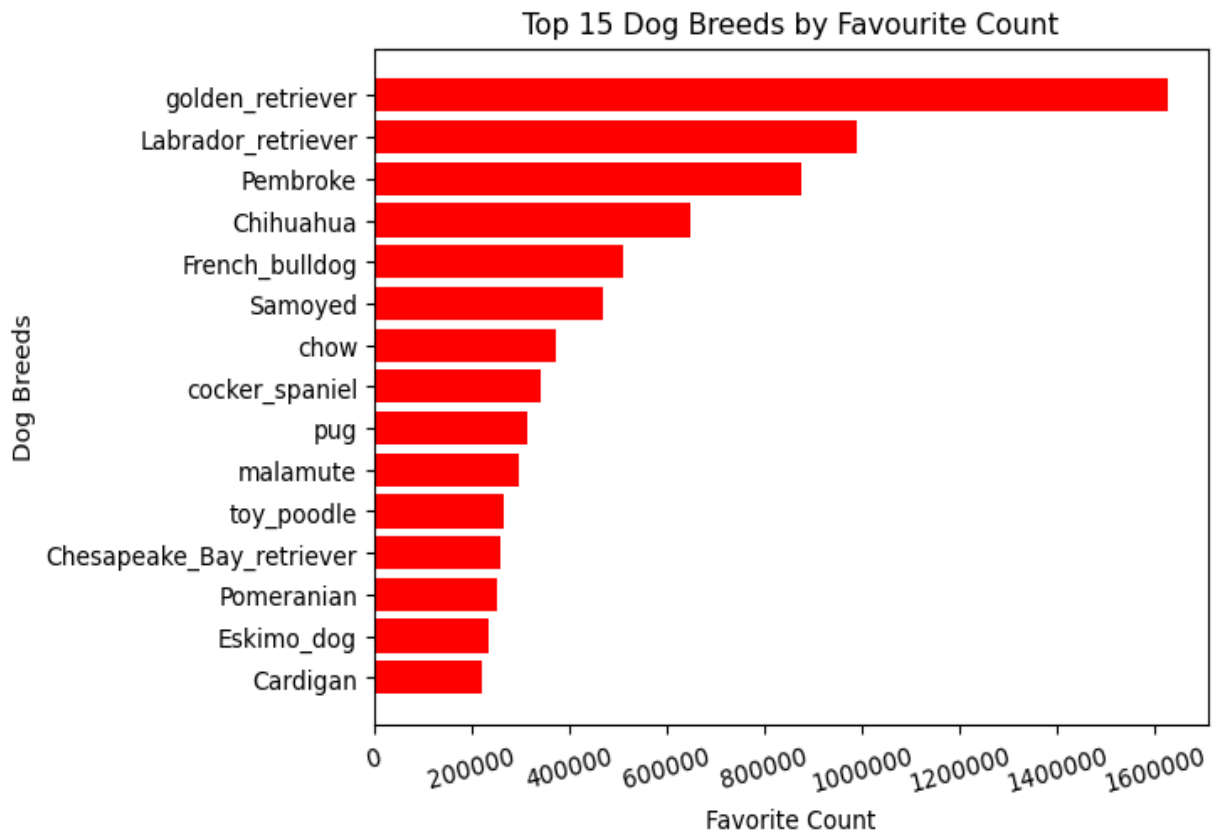
The dog ratings has a left-tailed distribution. WeRateDog gives out rates between 0 and 14 for the dogs. The rating with the highest frequency is around 11.

2. Which dog breed is mostly tweeted about?



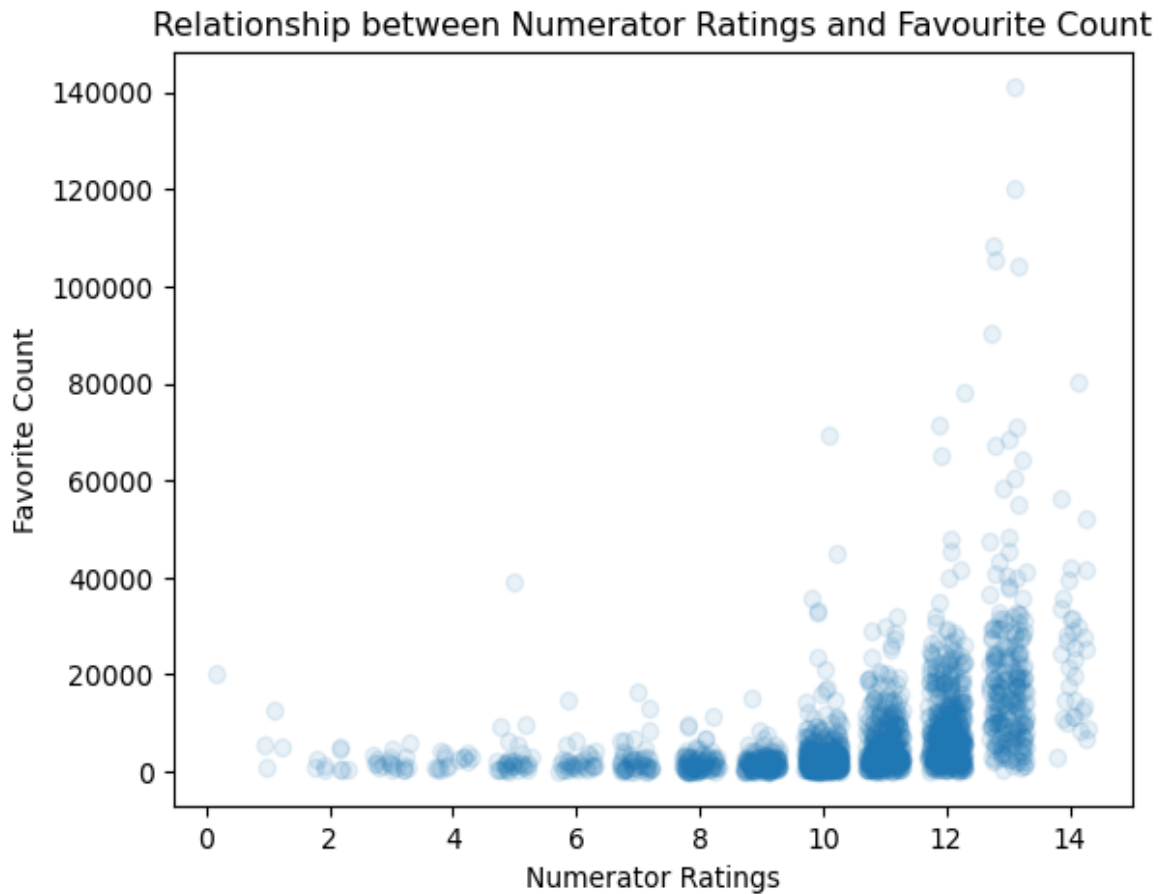
The dog breed mostly tweeted about by WeRateDog is the Golden retriever followed by the Labrador retriever.

3. Which dog breed has the highest favorite counts?



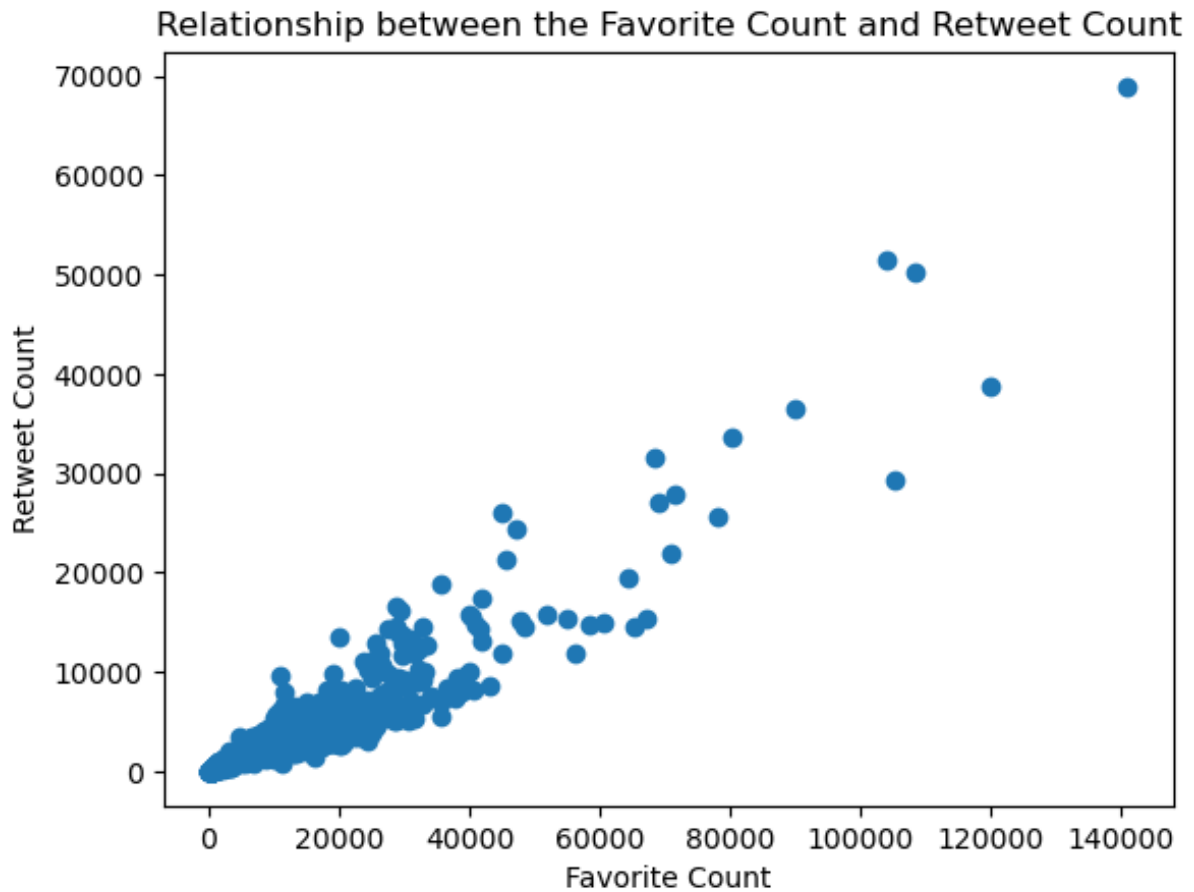
The top dog breed by favorite count is the Golden retriever followed by the Labrador retriever.

4. Does WeRateDog numerator ratings influence the favourite count?



Some high ratings have more favorite counts compared to the lower ratings. So there seems to be a positive relationship between numerator ratings and favourite count. As the numerator rating increases, the favorite count also increases.

5. Is there any relationship between favourite counts and retweet counts?



There is a strong positive correlation between favorite count and retweet count as expected. As the favorite count increases, the retweet count also increases.

Insights

In summary, the following insights were gathered:

1. The numerical rating distribution is left-tailed and the highest rating is around 11.
2. Golden Retriever is the dog breed mostly tweeted about by WeRateDog and it has the highest favorite count.
3. WeRateDog rating seems to influence the favorite count for dog breeds and
4. Favorite count has a strong positive correlation with the retweet count.