Data Analysis Project: Insights from WeRateDogs Twitter Archive
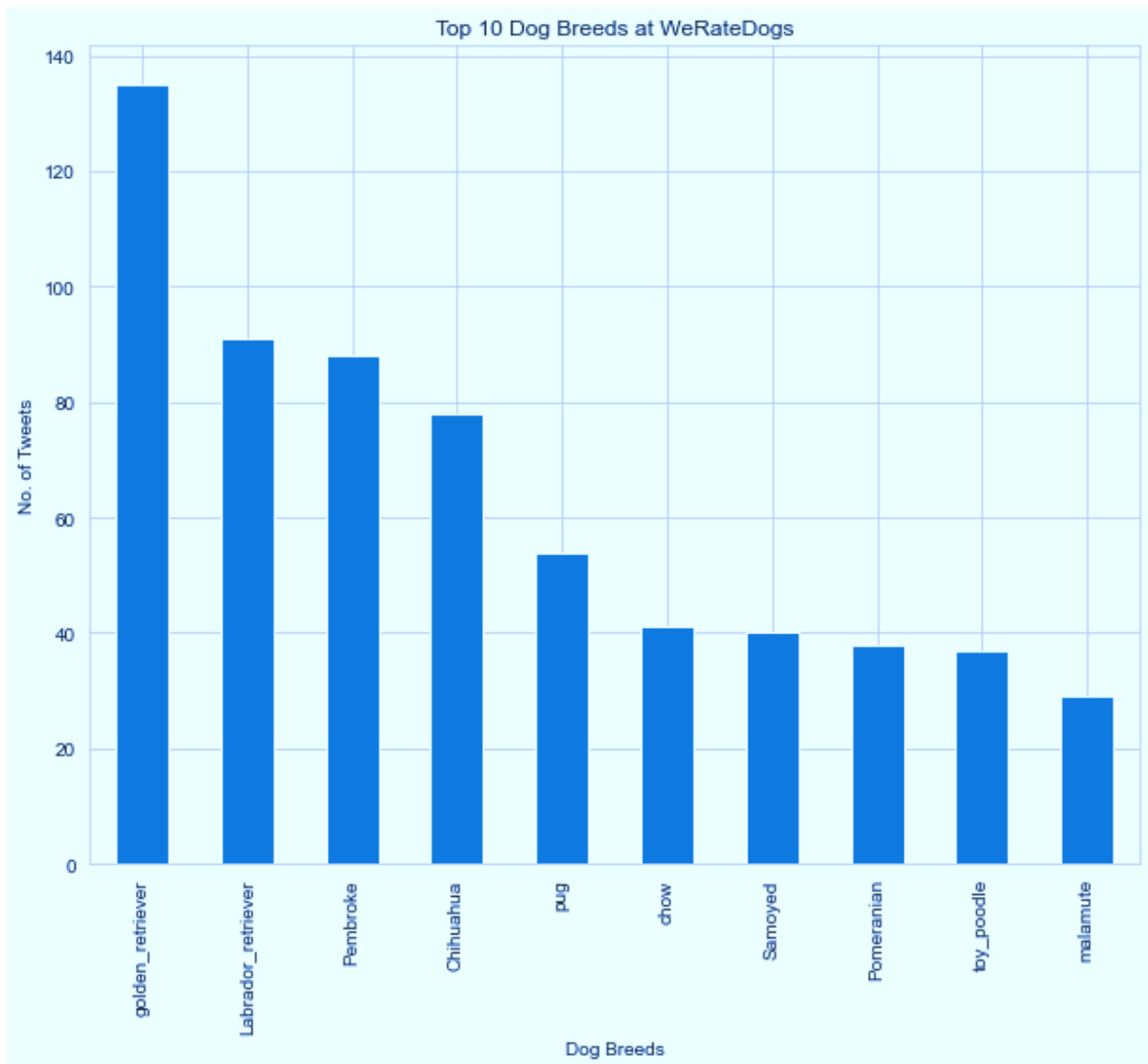
Author: Judith Agara

Thanks to ALX-T, I am currently enrolled in the Data Analyst Nanodegree program on Udacity, and thus far, it has been a fantastic experience. The project-based, walk you through the entire process, and then break it down step-by-step manner of instruction that Udacity offers is what I appreciate about it the most. It is most effective because we learn best by doing. I recently finished my second project, and I want to share some of the insights I got.
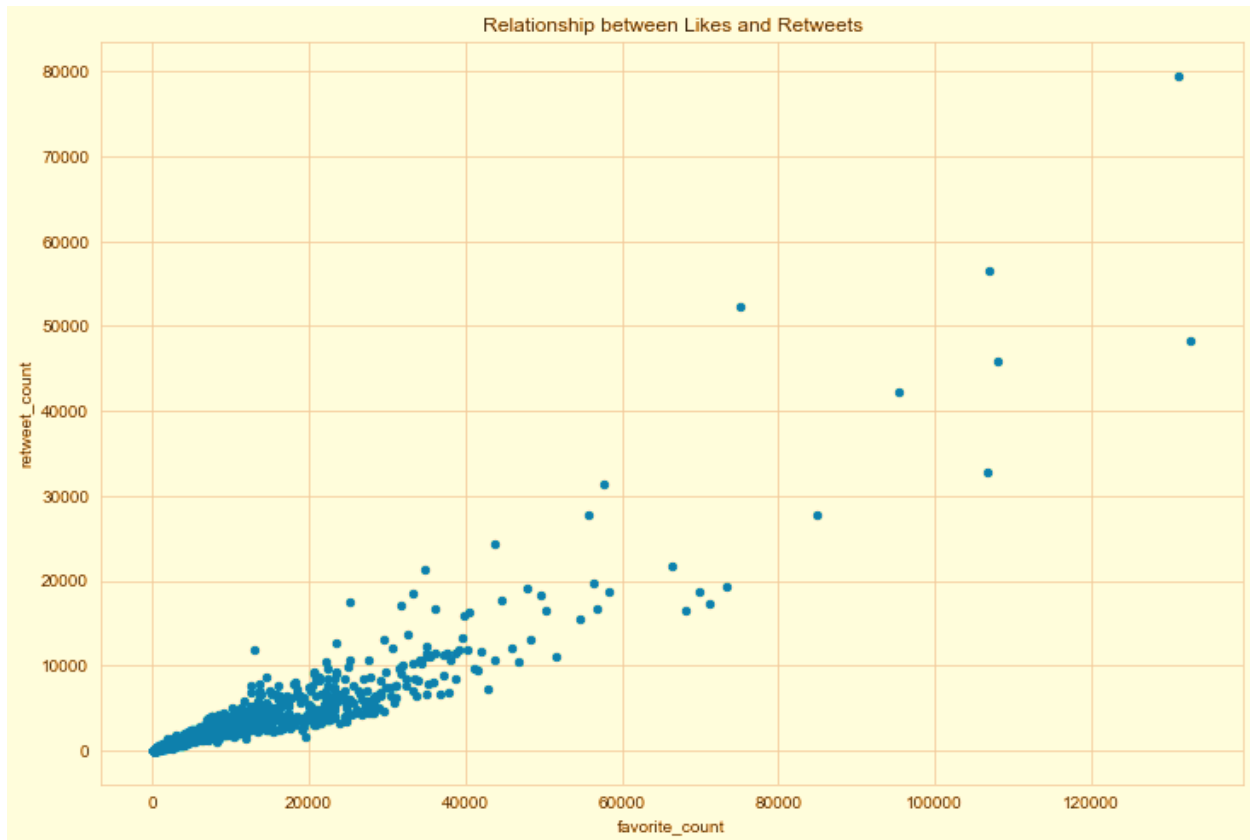
The tweet archive of Twitter user **@dog_rates**, also known as WeRateDogs, served as the dataset for this project which was designed to develop data wrangling abilities. **WeRateDogs** is a popular Twitter account that rates people's dogs with a lighthearted comment about the dog. It is famous for sharing adorable pictures of dogs, witty captions and dog ratings that consistently exceed a perfect 10 because "they're good dogs Brent". The account was started in 2015 by college student Matt Nelson to simply share dog pictures and make people laugh but has grown into a business, selling merchandise and posting branded content, which generates six figures annually. We Rate Dogs also aids and shares fundraisers for dogs in need with its millions of followers across social media platforms. In 2020, the account raised $1.3 million for more than 170 dogs.

What's to see in this dataset? Well, after combining the twitter archive with additional data gotten via the Twitter API and image prediction results obtained from running each image through a neural network for classifying dog breeds, I found the following:

1. The golden retriever is the most popular dog breed shared with and by Twitter user WeRateDogs.

Top 10 Dog Breeds at WeRateDogs

2. There is a strong correlation between retweet count and favorite count. For most cases, more likes meant more retweets. This would mean that most people retweet because they like the tweet. There are a few exceptions of course.

3. The Saluki breed appears in the top five breeds whether we rank by rating, retweet count or favorite count. It also has the highest average rating.

4. Golden retrievers and Labrador retrievers do not make the top 10 lists for ratings, favorite counts and retweet counts even though they are the most popular dogs on the block.

   These two insights can be observed by looking at the plots below:

Top 10 Dog Breeds by Rating

Top 10 Dog Breeds by Number of Likes

Top 10 Dog Breeds by Number of Retweets