Onesmas Ngugi

11/20/2017

`WeRateDogs` Twitter Archive Insights

An analysis of the data from the `WeRateDogs` twitter archive gave a few useful insights that are outlined below. The data included about 3000 tweets with information such as the favorite and retweet counts, dog ratings for each tweet and dog breed information obtained from running the twitter archive through a [neural network](#) that can classify breeds of dogs.

After cleaning and combining the data sets, some records were eliminated as they did not have complete information.
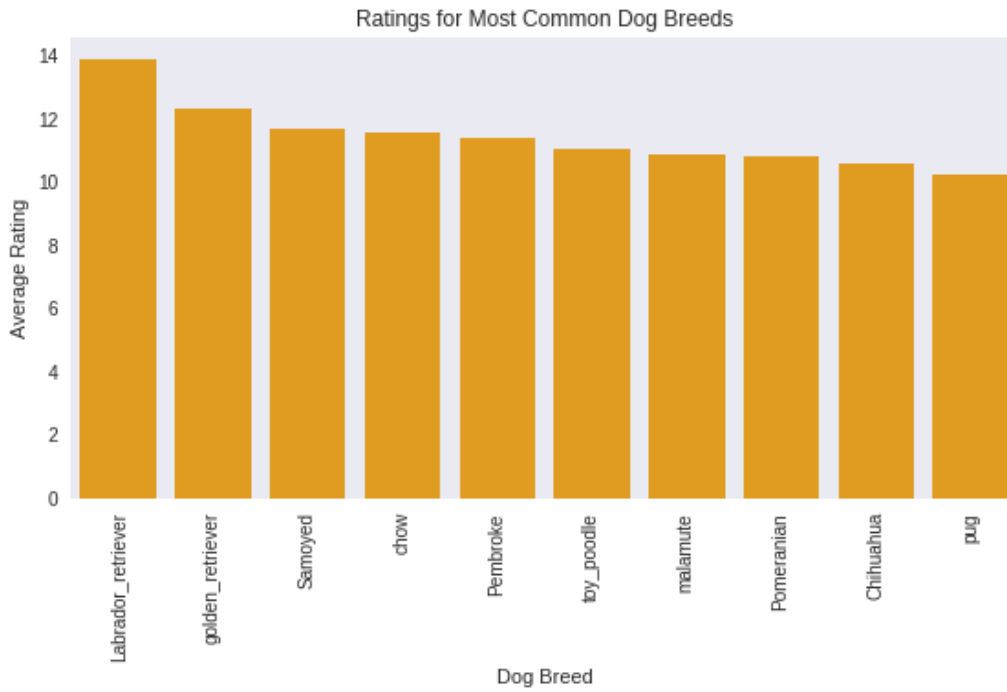
The first step in the analysis was to group the data. I used 'p1' as the grouping variable which was the most likely breed of dog from the image predictions. Grouping by dog breed allowed me to compare dog breeds by aggregating other columns.

To get top ten most common dog breeds, I used `count` as the aggregating function, then sorted the values in descending order and limited the output to ten. Below are the ten most common dog breeds from the archive.

1. Golden retriever
2. Labrador retriever
3. Pembroke
4. Chihuahua
5. Pug
6. Chow
7. Samoyed
8. Toy Poodle
9. Pomeranian
10. Cocker Spaniel

Of the top 10 most common dog breeds, the Labrador retriever has the highest average rating of 13.90 and the Pug has the lowest rating of 10.24.

The Samoyed has both the highest average favorite and retweet counts among the ten most common dog breeds.

Ratings for Most Common Dog Breeds

Above is a bar graph of average ratings for the ten most common dog breeds.