

Act Report

Analyzing and Visualizing Data

Questions answered using code:

1. What is the time frame for this data?
2. What are some general statistics for the dataset?
3. What were the top five most commonly predicted dog breeds (first_prediction)?
4. Which breeds received the highest favorite count? Retweet count?

Insight #1: The starting date of the dataset is 2015-11-15, and the ending date is 2017-08-01

The dataset spans over two years. During this time, we can investigate possible patterns and trends that occurred and use those for other predictions and forecasts.

Insight #2: Statistics Found About the Data

- The **average rating numerator** is 12.2. There are lots of good pups out there.
- The **average rating denominator** is 10.4. This average denominator is due to some data entries which include denominators greater than or less than 10 (min=2, max=170). Upon inspecting the 'text' column earlier in the data, this may have been intentional but could be further inspected.
- The average favorite count is 8980.9, the minimum favorite count is 81, and the maximum is 132,810.
- The average retweet count is 2841.7, the minimum retweet count is 16, and the maximum is 79,515.

These statistics indicate that the rating system used by the users is inconsistent, prompting stability needed in either the rating system, or simply for varying values to be taken into account.

Insight #3: The Top Five Dog Breeds Predicted

1. Golden Retriever (137)
2. Labrador Retriever (94)
3. Pembroke (88)
4. Chihuahua (78)
5. Pug (54)

When inspecting the complete list of values for the first breed prediction column, I found that many first predictions were not dogs. This should be further inspected by comparing the pictures to the resulting prediction to improve the prediction algorithm.

Insight #4: There is a strong correlation between the top predicted breeds and the favorite and retweet counts.

The top four predicted dog breeds also hold the top four spots for top favorites and retweets. This demonstrates a correlation between these variables and could be further inspected.

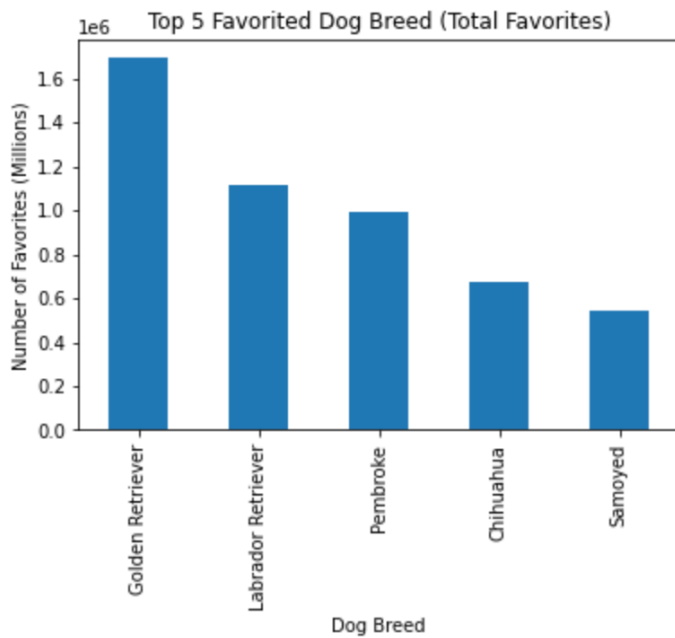
Something to note is that when investigating the **mean** favorite and retweet counts, the top counts were for images not predicted as dogs.

Another insight gained here is that Labrador Retrievers and Chihuahuas were more consistently at the top for total favorites and retweets, as well as the highest favorites and retweets for an individual post. It could be interesting if these breeds were also rated highest according to the WeRateDogs rating system.

Visualizations:

Below are the counts for the top five predicted dog breeds:

[*Source for using matplotlib:](#)



Below are the top five predicted breeds that had the highest favorites for an individual post:

