

## **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 5 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**

We can see that the dataset spans over about 2 years. During this time, we can investigate possible patterns and trends that occurred and use those for other predictions and forecasts.

#### **Insight #2: Some interesting statistics were found...**

- The average **\*\*rating numerator\*\*** is 12.2! Lots of good pups out there
- The average **\*\*rating denominator\*\*** is 10.4. This is due to some entries of data including 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 either stability needed in the rating system, or simply for inconsistent values to be taken into account.

**Insight #3: The top 5 dog breeds predicted were:**

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

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

**Insight #4: There seems to be a direct correlation between the top predicted breeds and the favorite and retweet counts.**

We can see the top 4 predicted dog breeds also hold the top 4 spots for top favorites and retweets. This demonstrates some correlation which 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.*

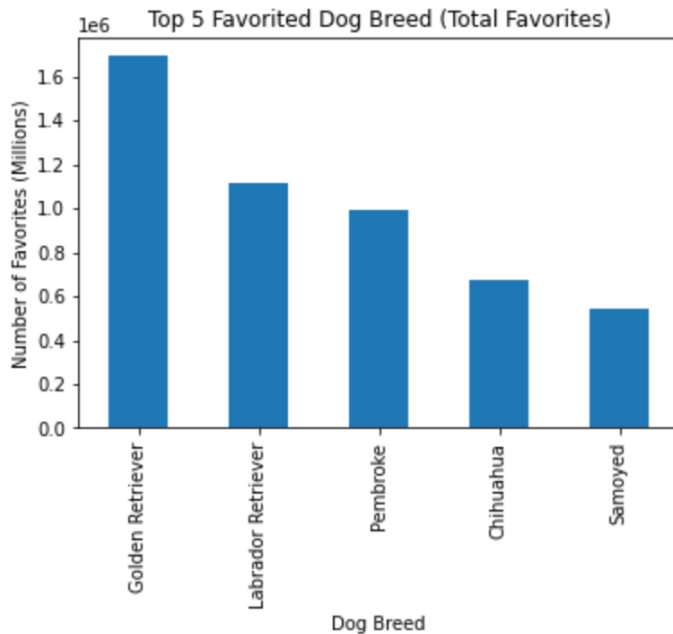
Another insight gained here is the fact that Labrador Retrievers and Chihuahuas were more consistently at the top for total favorites/retweets and highest favorites/retweets for an individual post. It could be interesting if these breeds were also rated highest according to the WeRateDogs rating system.

**Visualizations:**

Let's see how those counts look for the Top 5 predicted dog breeds using a bar chart:

\*Source for using matplotlib:

<https://www.python-graph-gallery.com/191-custom-axis-on-matplotlib-chart>



Let's also take a look at the Top 5 predicted breeds that had the highest favorites for an individual post:

