

act_report
September 24, 2023
Marcus Thompson

After completing the data wrangling process, we can now investigate the data to see what insights can be discovered.

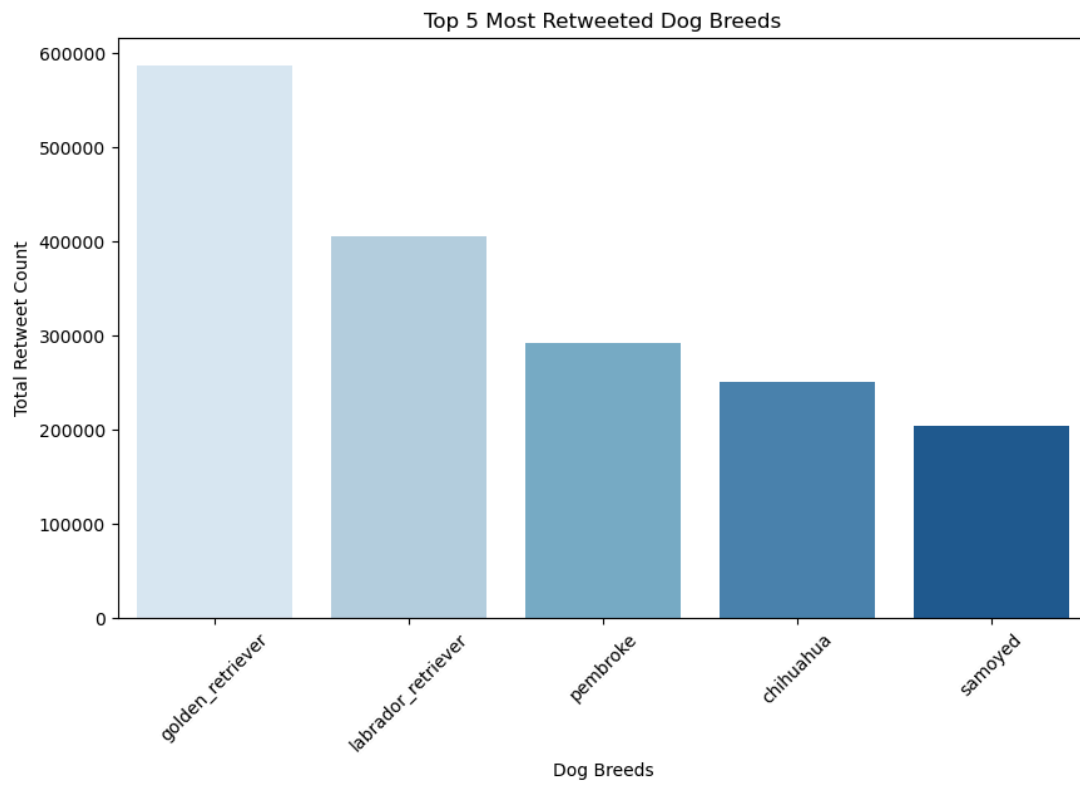
WeRateDogs comments on images of dogs and gives ratings scores on a scale of 1-10, with many dogs earning scores that exceed a 'perfect' score of 10/10. This is in no way 'scientific', and done purely for the love of dogs. Nevertheless, some interesting insights can be drawn from analyzing the data available to us. Let's explore:

Insight #1:

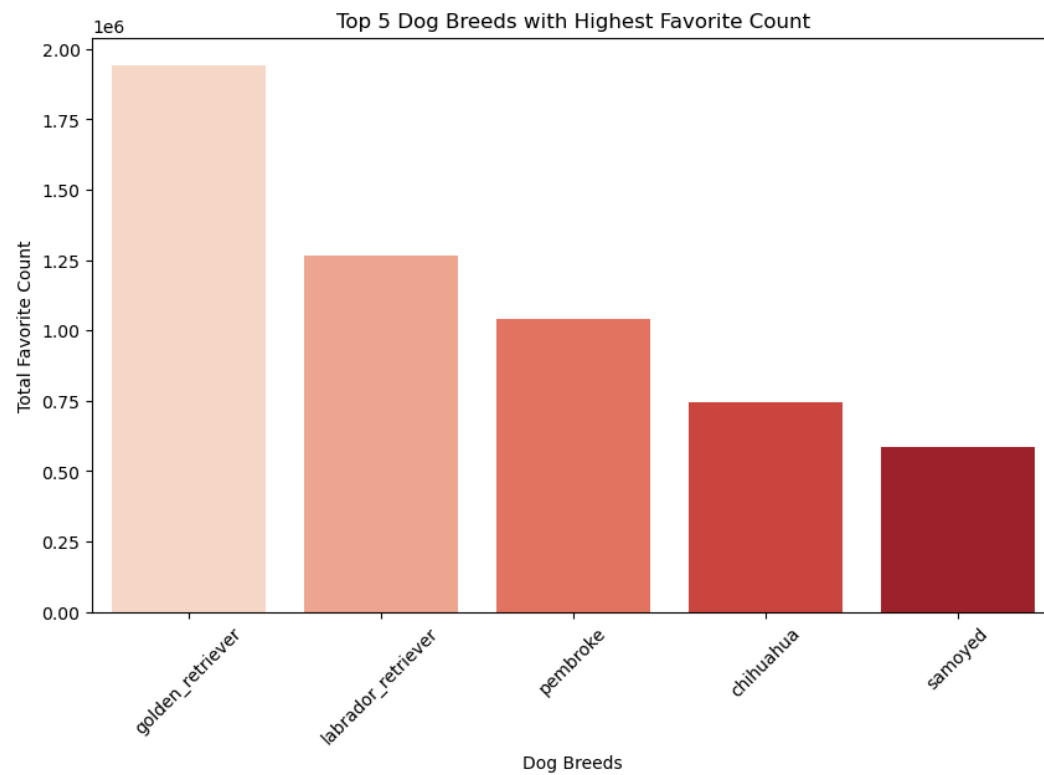
What are the top 10 most common dog breeds to appear in our data? How many times does these dog breeds appear?

| Dog Breed | Value Counts |
|--------------------|--------------|
| Golden Retriever | 156 |
| Labrador Retriever | 106 |
| Pembroke | 94 |
| Chihuahua | 90 |
| Pug | 62 |
| Toy Poodle | 50 |
| Chow | 48 |
| Samoyed | 42 |
| Pomeranian | 41 |
| Malamute | 33 |

How does this compare to the five most retweeted dog breeds?



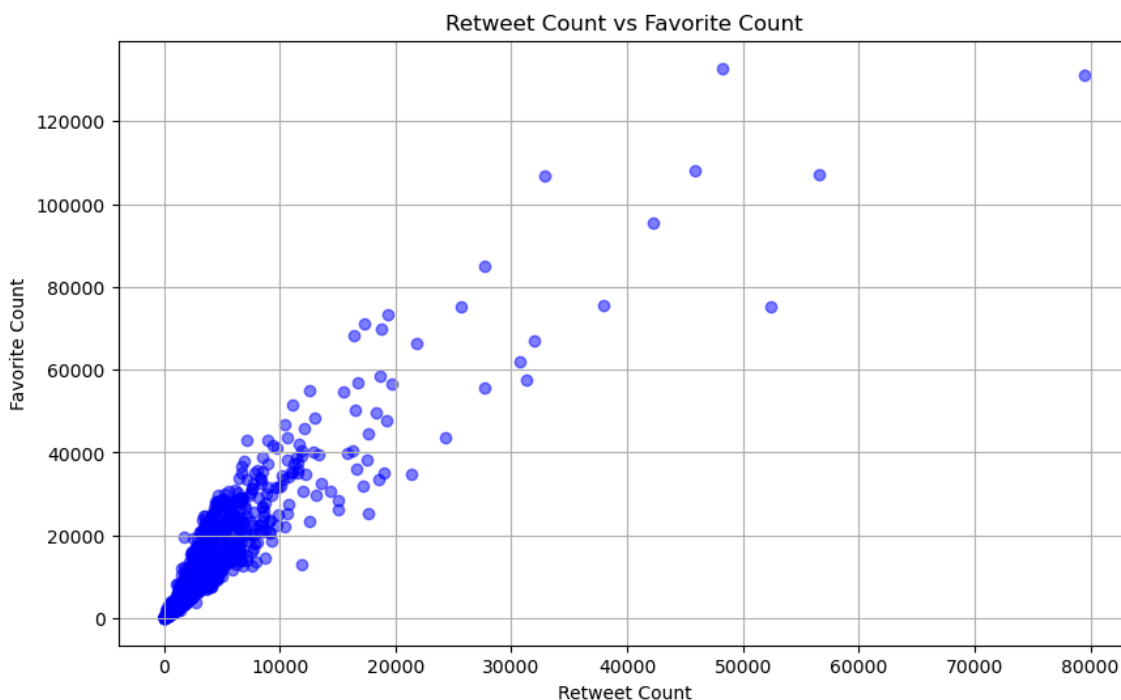
What about the five dog breeds with the most favorite counts?



- The top five breeds with the most retweets and favorite counts are the same.
- The top four breeds are also the top four most common breeds to appear in our DataFrame.
- However, the breed with the fifth most retweets and favorite counts, the samoyed, is actually the eighth most common breed to appear in our DataFrame. This suggests that the most common breeds do not necessarily receive the most favorite counts or retweets.

Insight #2:

The two bar plots above (retweet count and favorite counts) are similar. Perhaps retweet counts and favorite counts are correlated? Let's see the scatterplot and interpret the results:



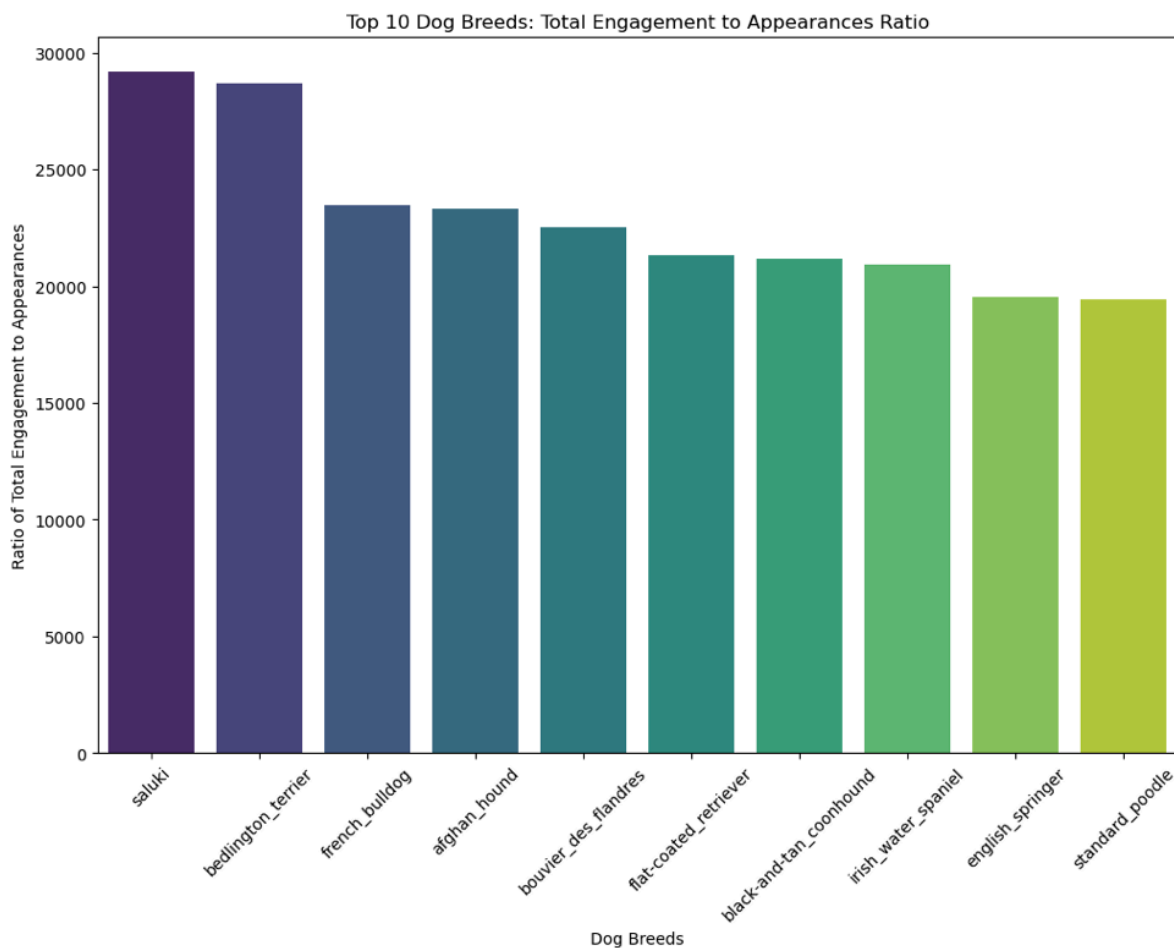
It does indeed appear that retweet counts and favorite counts share a positive correlation.

Insight #3:

What are the top 10 breeds with the highest engagement ratios?

engagement ratio = (retweet counts + favorite counts) / total appearances of breed

Let's explore our visualization and see how it compares to our bar plots from Insight #1:



We see a striking difference between this visualization and the bar plots of breeds with the most retweets and favorite counts.

We can see that golden retrievers are the most common dog breed that appears in the DataFrame, and also holds the most retweet counts and favorite counts. We can contrast this with the saluki, which only appears four times (compared to 156 times for the golden retriever), but has almost double the engagement ratio. In fact, none of the dogs with the top 10 highest engagement ratios appear in our charts of breeds with the most favorites or retweets.

The data shows that there are far less common dogs that have a much higher engagement ratio than the top five dog breeds that have the highest retweet and favorite counts.

Final Thoughts:

Some limitations exist with this project. We did not consider mixed dog breeds, and the image prediction algorithm can not be trusted to be 100% accurate. WeRateDogs is not a serious, scientific account, as it features ratings that are highly subjective, with many dogs receiving ratings of 10+/10, or even 1776/10. There were also many missing values for the 'doggo', 'puppo', 'floofer' and 'pupper' categories.

Nevertheless, this was a fun, light-hearted project, yet still offered valuable experience in the data wrangling process.