

WeRateDogs Act Report

The Data

The "enhanced" WeRateDogs Twitter archive contains tweet data for 2,356 of the 5,000+ tweets that were tweeted between November 2015 and August 2017. Data includes dog rating, dog name, and dog stage. What's the "dog stage"? Well, whether the dog is considered a doggo, floofer, pupper, or puppo, of course!

Sadly, the retweet count and favorite count of each tweet was missing from the original data. That's not very useful.... Luckily, we were able to use Twitter's API to retrieve that data!

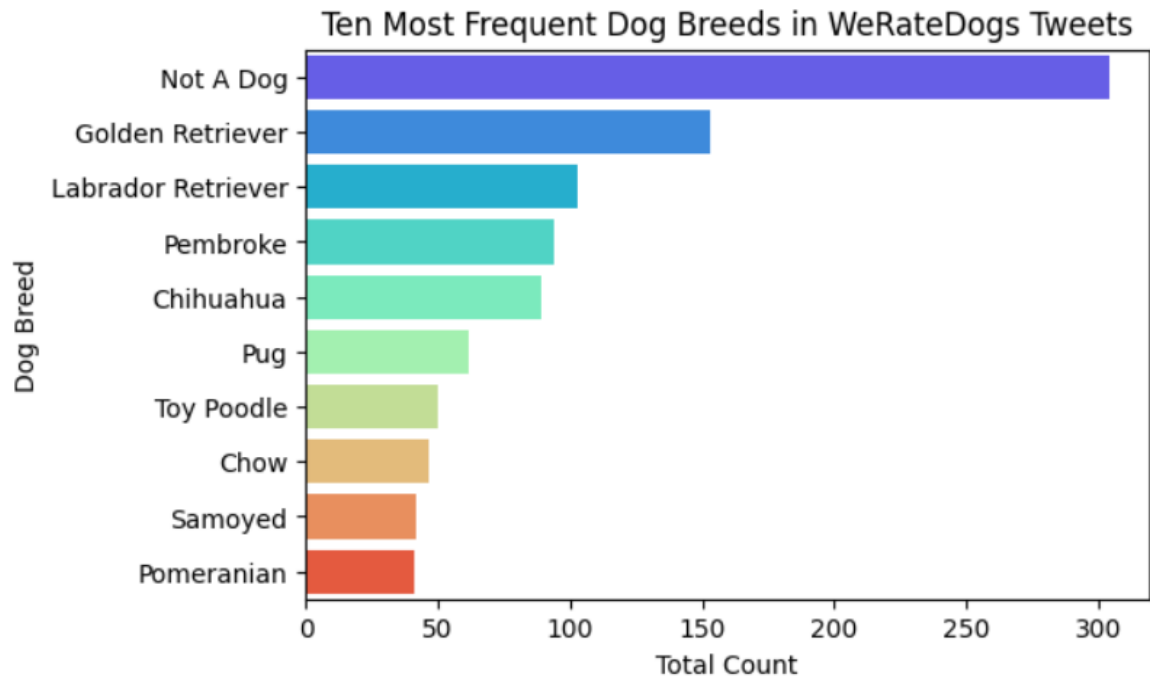
In addition, I was able to access a file containing dog breed predictions obtained through a neural network. The predictions aren't perfect and often think a dog is a laptop or an orange, but it's pretty cool to be able to use those predictions in this analysis!

After wrangling and cleaning the data, we were able to gain some valuable insights about WeRateDogs.

The Insights

Which dog breeds appear the most often in the dataset?

First, I looked at which dog breeds were most frequently tweeted about in the dataset. I wanted to see how good that neural network's predictions were.



"Not A Dog" had the highest count. We did not explore the reason for this, but it could be because the neural network "saw" a laptop instead of a dog sitting behind a laptop.

After "Not A Dog", the breed most frequently found was the Golden Retriever! Big, floofy goofballs. I am okay with this result even though my dogs' breeds didn't even make it into the top 10. No offense to the breeds above, but I think Boxers and Beagles deserved to be in the top 3. Hmmph.

Which dog names are found most frequently?

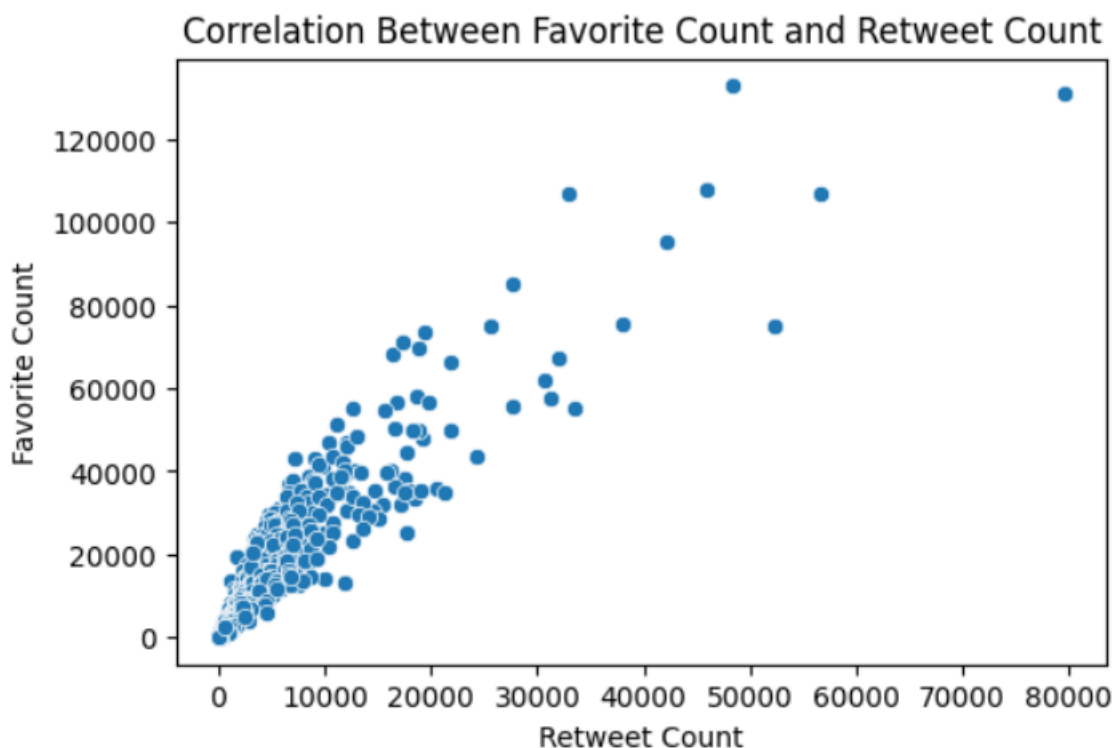
After dog breeds, I was curious which dog names were found the most frequently.

charlie	11
Lucy	10
oliver	10
Cooper	10
Tucker	9
Penny	9
Sadie	8
Winston	8
Daisy	7
Toby	7

Charlie was the number one most frequent dog name! Good boy, Charlie! This top 10 made me laugh because my human daughter's name is Penny and my cat's name is Sadie. Obviously, all Penny's and Sadie's are very good girls whether they are canine, feline, or human.

Is there any correlation between favorite count and retweet count?

Finally, I wanted to explore the relationship between favorite count and retweet count. Intuitively, it would make sense that the higher the retweet count is, the higher the favorite count will be. Let's see what I discovered:



We have a strong positive correlation between favorite count and retweet count! Intuition wins this time! There are a few very large outliers seen here, but removing them would not have significantly changed anything.

The Conclusion

Aside from having the best doggos and puppies around, WeRateDogs has some interesting data to explore. We found out that big goofball dogs and little ankle-biter dogs are both frequently mentioned in WeRateDogs' tweets, Charlie is such a good boy that many good boys share the name, and that tweets with high retweet counts also tend to have high favorite counts.

This dataset was so much fun to explore, and now I encourage you to go explore it too! Did one of the Charlie's have one of the highest ratings? Were any of the "Not A Dog" predictions actually dogs and not sarongs?

At the very least, go find a good doggo and give them a treat. They deserve it.

In []: