Act Report for Project 2: Wrangle and Analyze Data

In this report I will detail and share my insights and analysis I made from the wrangled dataset. The wrangled dataset is a table of WeRateDogs's tweets with tweet_id, ratings, dog names, dog stages, breed names, favorite and retweet count.

Insights

The first thing that popped in my mind what was the most common dog name from this dataset. The result is that Charlie and Lucy both tie for the most common dog names with a count of 11 while Cooper and Oliver tie for second place with a count of 10.

Since the dataset had predicted breed names from a neural names and confidence level for each prediction, the next question I had was in the most retweeted tweet what was the breed of the dog? The findings were *tweet_id* **744234799360020481** was retweeted **79,515** times with the text *"Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad)"* sadly since the tweet contains a video, I can't show it on this report but here is a thumbnail of the video. The neural network predicted the dog's breed as Labrador Retriever with a confidence score of 82%.


Since I checked the most retweeted tweet, I was also curious of most favorited tweet. So the next question I had "What was the dog breed in the most favorited tweet?" The findings were that *tweet_id* **822872901745569793** was retweeted **132,810** times with the text *"Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10"*. Below is the picture of the dog. The neural network predicted it as an Irish Terrier with a confidence level of 70%.

Visualizations

I wanted to know if there was a trend where tweets featuring dogs in a particular stage were more likely to be retweeted or favorited.

From the chart we can see that Pupper had the most favorites and retweets with 43% while Doggo coming up second with around 37%.
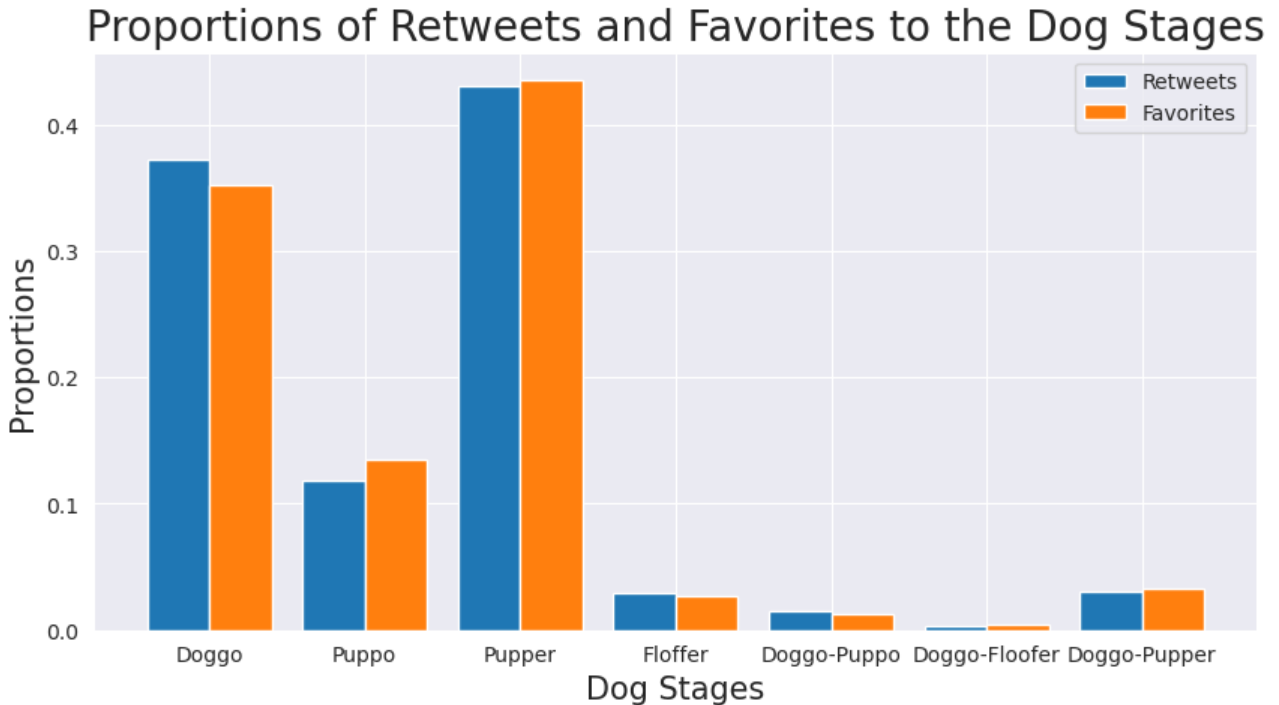
I also wanted to know if there was a relationship between the number of retweets and the dog ratings. Suprising enough the highest rating in the dataset 1776/10 had a meager **2,772** retweets while a rating 13/10 had a max retweet of **79,515**. The difference could be that the score 1776/10 was only featured in a single tweet while 13/10 was used **299** times! The most used score was 12/10 with it being used 492 times, it had a max retweet of **33,421**.