

## Act Report

After the cleaning process, I was ready to visualize and get a deeper understanding of what could I conclude from the data set.

In the beginning, I wanted to understand throughout the specific date for the data of WeRateDogs, which dog got the highest retweets? And is it the same dog who got the highest favorites as well?

I was able to get the answer using 'max()', and it appears that the highest retweets were 79515 retweets and it went to a dog standing in a pool. It is worth to mention that the image predictor predicted that this dog belongs to Labrador Retriever breed of dogs with a confidence of 82%. Figure (1) shows a picture of the dog.

```
In [13]: Image(url= highest_retweets.jpg_url.values[0],width=150)
```

Out [13]:



Figure 1: The dog with highest retweets

After that, I wanted to understand if the same dog got the highest favorites but, actually, I found that the dog with the highest favorites got 132810 favorites and it went to a supportive dog participating in an event. The image predictor predicted that this dog belongs to the Lakeland Terrier breed of dogs with a confidence of 19.6%.

```
In [20]: Image(url= highest_favorites.jpg_url.values[0],width=200)
```

Out [20]:



Figure 2: The dog with highest favorites

Secondly, I was curious in understanding which numerator rate that WeRateDogs gave most of the time to dogs, and by knowing that most of their ratings are above '10' how to understand whether or not a specific rating considered as balanced or good or too good for a dog in comparison to another rating numerator?

As shown in figure (3) below, I found that the most frequent rating numerator that WeRateDogs account use the most in their ratings is 12 in the first place, 10 in the second place, 11 in the third place, and finally 13 at the fourth place. As WeRateDogs account, having 12 at the top in comparison to 11 and 13 means that 12 rate represents a balanced rate.

```
twitter_archive_cleaned.rating_numerator.value_counts()
```

12	486
10	436
11	413
13	287

*Figure 3: Most frequent numerators*

My third thought for the data set, I wanted to understand the times that each prediction number algorithm is whether or not a breed of dog was true. The results have shown that the length of the truth first algorithm predictors was 1463 times, and the length of the truth second algorithm predictors was 1480 times, and the length of the truth third algorithm predictors was 1431 times. In comparison to the total number of rows for the data set, which is 2097 rows, the truth times of each predictor exceed half of the number of rows. This means that the predictions process is good.

Finally, to get an interesting visualization of the portions of each dog stage based on their retweet count, I decided to visualize the dog stage by drawing a pie chart as shown in figure (4) below:

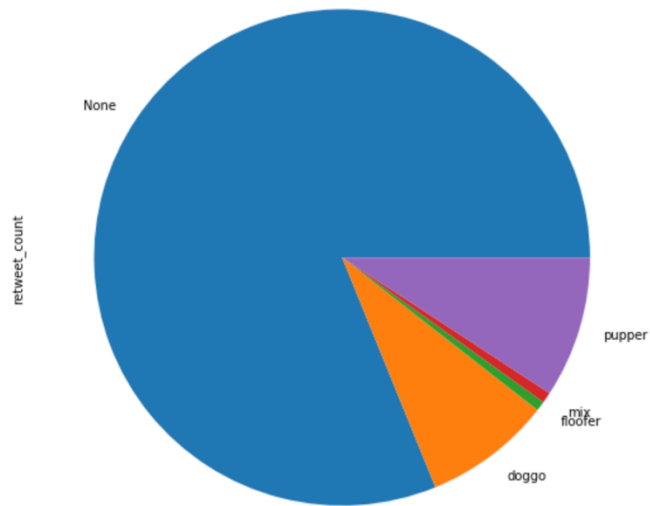


Figure 4: Dog stage by retweet counts

The above pie chart result has shown that based on the number of retweets, the highest dog stage that got the highest retweets is not known, the second-highest number of retweets went for the pupper dog stage, and doggo dog stage got the third place in the highest number of retweets.