

Act Report

Project

2

June 2022

An analytic summary of the WeRateDogs Twitter Account
presented by Glory Edamkue



Behind The Scenes

How ratings affect tweeter reactions

The most rated dogs in the Twittersphere

The validity of Machine Learning in dog matters

**API
QUERY
ADDED!!**



INTERACTIVE STUDY WITH CONCLUSION INCLUDED

INTRODUCTION TO THE ANALYSIS

It has been an interesting project, for sure, analyzing the Twitter page of the dog rating haven [WeRateDogs](#)® (@dog_rates) on Twitter.



For starters, they have some **REALLY** good dogs and also some really interesting ratings, the most obvious being that they do not rate like every other animal rater does. With @WeRateDogs, dogs [can have ratings above ten](#) when the denominator in its ratings is 10. Standard deviation much? Let's see.

GETTING STARTED



First off, I cleaned the dataset so that it only contained original tweets, not likes or retweets. Why? Because those are the tweets that actually give value for ratings and dog ages, names, etc. What was nice to have was an image prediction result from a Convolutional Neural Network for the dataset I worked with, one of the computer vision machine learning algorithms out there.

I was curious to know which dog got the highest rating in terms of retweets and favorites. Did rating affect how people would react to a dog? Were the prediction confidence levels reflective on the accuracy of a prediction? What is the most popular dog breed? Did the most popular dog breed get the highest ratings? What age/ size of dog was the most popular? From the available data, what was the most popular dog breed?

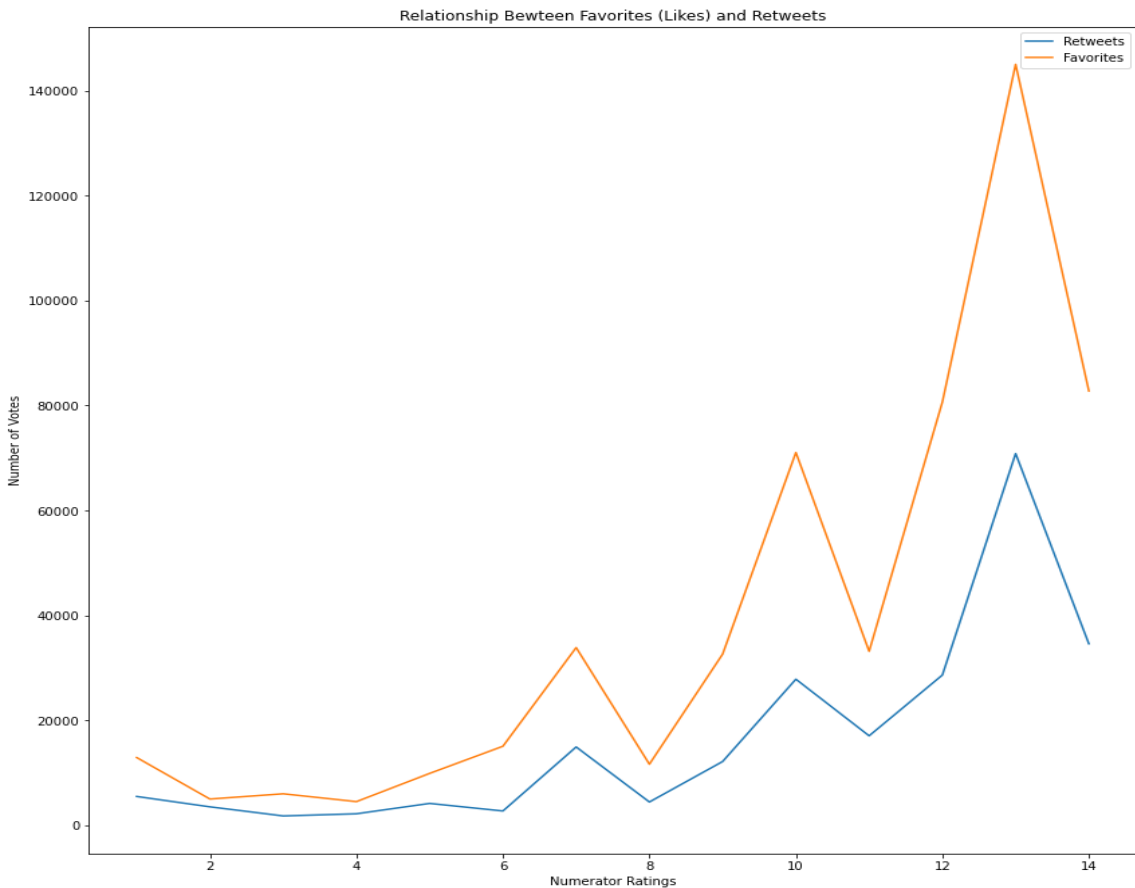
There was a lot to ask and an almost limited way of getting answers in the way I would have liked to.

I set out some questions I would like to know, and analyzed the cleaned dataset. Here are my findings.

INSIGHT INTO DOG RATINGS

Could ratings affect how tweeters react to a dog?

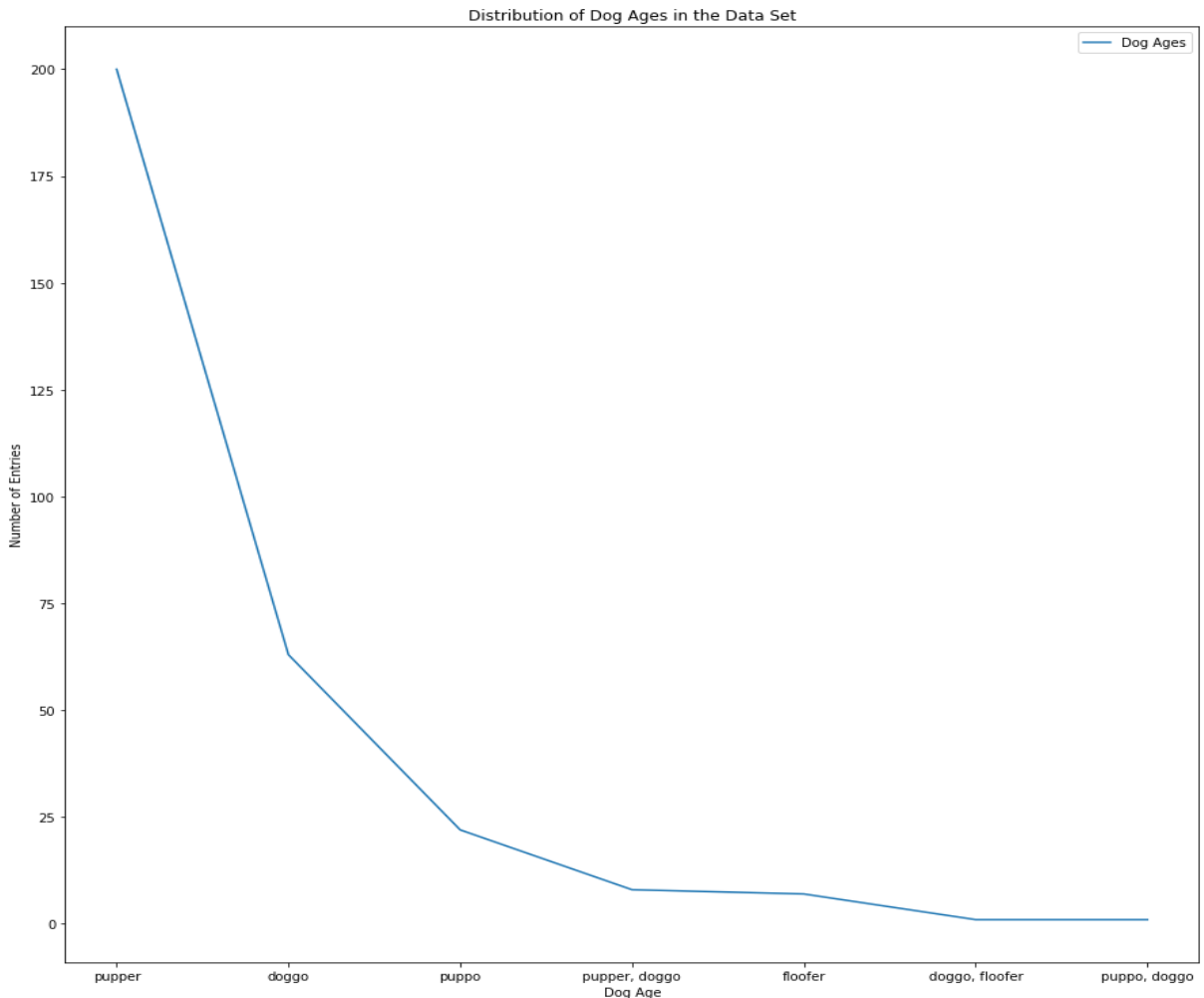
- A. Do the most retweeted images have the highest numerator ratings?
- B. Do the most liked images have the highest numerator ratings?
- C. What is the relationship between likes and dislikes?



Observation:

Yes, ratings affect how tweeters could react to dogs. It is proportional to an extent that the higher the rating, the higher the number of the retweets and likes for an original tweet.

What is the most frequent dog age category (where available)?



Observation:

Puppers were the most frequent dogs submitted for rating. The least frequent dogs submitted for ratings were floofers.

What is a floofer, and what's a pupper? Well, these are 2 of the dog stages used in classifying a dog, but there are 4 in number: a pupper is a puppy, a floofer is a fluffy dog, a puppo is somewhere between a puppy and an adult dog, and a doggo is an adult dog.

I wanted to know how dogs of different ages fared with retweets and likes. Also, I wanted to know the least favorite dog age and how it fared in terms of popularity, too.

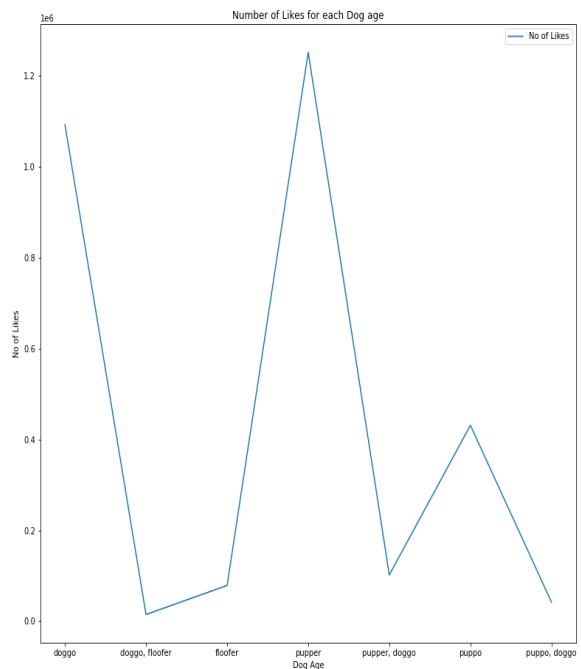
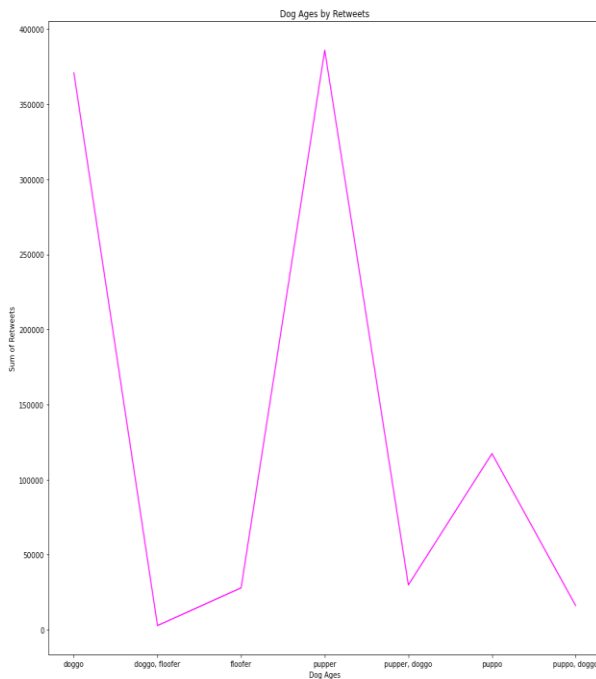
```
# To find out the sum of retweets for each known dog age
enhanced_clean.groupby('dog_ages')['retweet_count'].sum().sort_values()
```

dog_ages	
doggo, floofer	2801
puppo, doggo	16156
floofer	27951
pupper, doggo	29850
puppo	117345
doggo	370917
pupper	386043

```
# I am Looking for total Likes for each dog stage
enhanced_clean.groupby('dog_ages')['favorite_count'].sum().sort_values()
```

dog_ages	
doggo, floofer	14859
puppo, doggo	41943
floofer	78645
pupper, doggo	101815
puppo	431010
doggo	1092448
pupper	1251601

Name: favorite_count, dtype: int64



When it came to total retweets, the puppers had 386, 043 retweets, and the doggo, floofer had only 2, 801 retweets. Which is fair, because there was only 1 known doggo, floofer in the cleaned dataset.

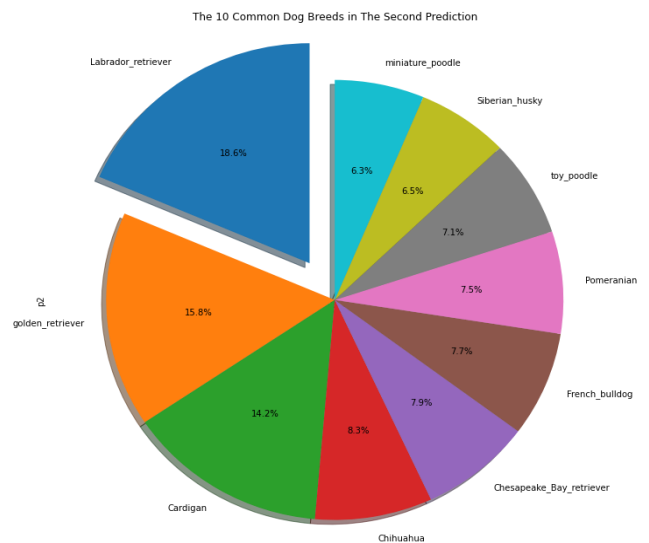
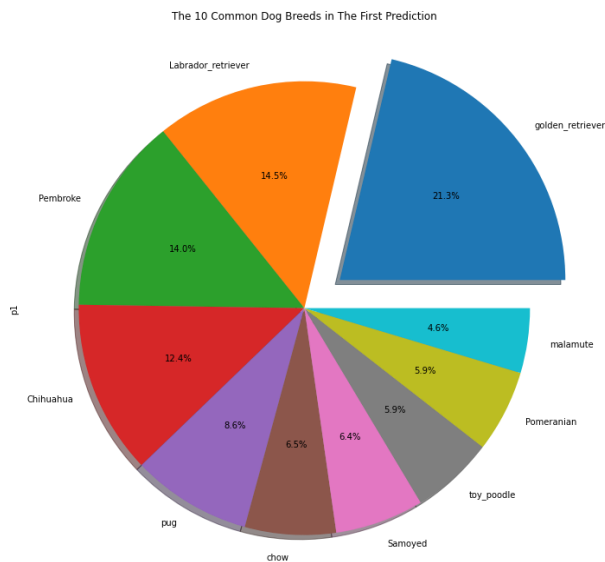
The dog age with the highest cumulative number of favorites (likes) was the pupper category with 1, 251, 601 total likes.

INSIGHT INTO THE DOG BREEDS

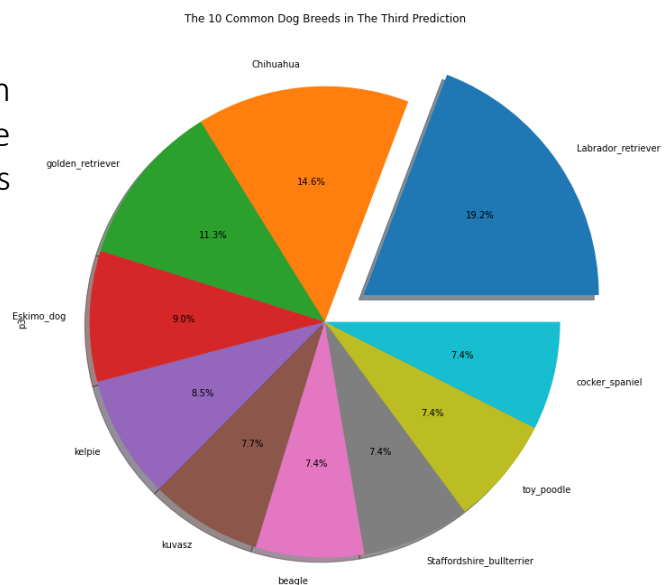
It was going to be interesting to know the most common dog breed from the predictions, and how it fared in terms of likes and retweets.

The most frequent dog in the first prediction was the golden retriever. It had 135 submissions.

In the second prediction, the most frequent dog breed was the Labrador retriever with 94 entries.



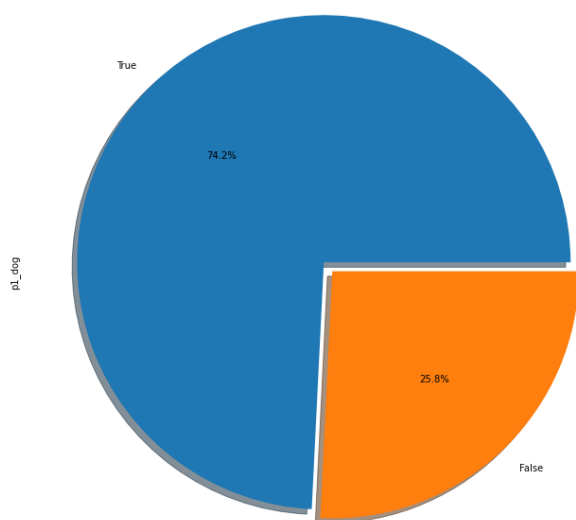
The most frequent dog breed in the third prediction was the Labrador retriever with 78 tweets of the dog breed.



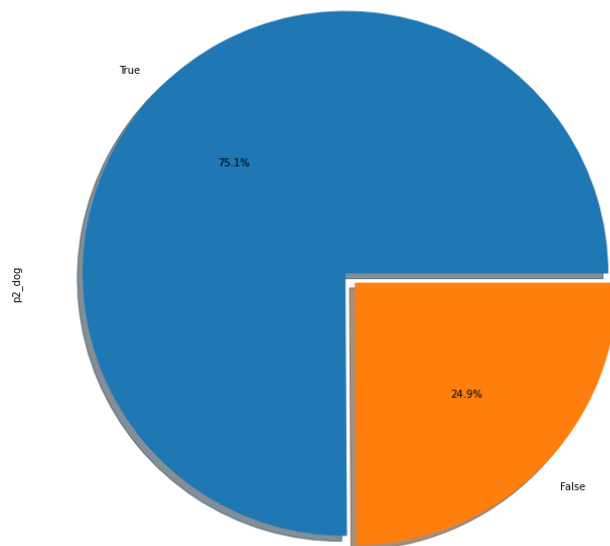
INSIGHTS INTO THE ACCURACY OF THE DOG BREED PREDICTIONS THAT HAD BEEN MADE

I wanted to know the accuracy of the of the dog breed predictions. Turns out the pie charts speak louder than any words.

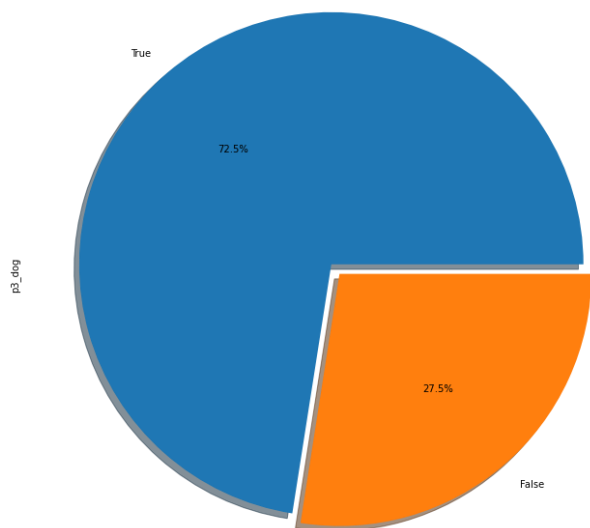
P1 Dog Predictions Accuracy



P2 Dog Predictions Accuracy

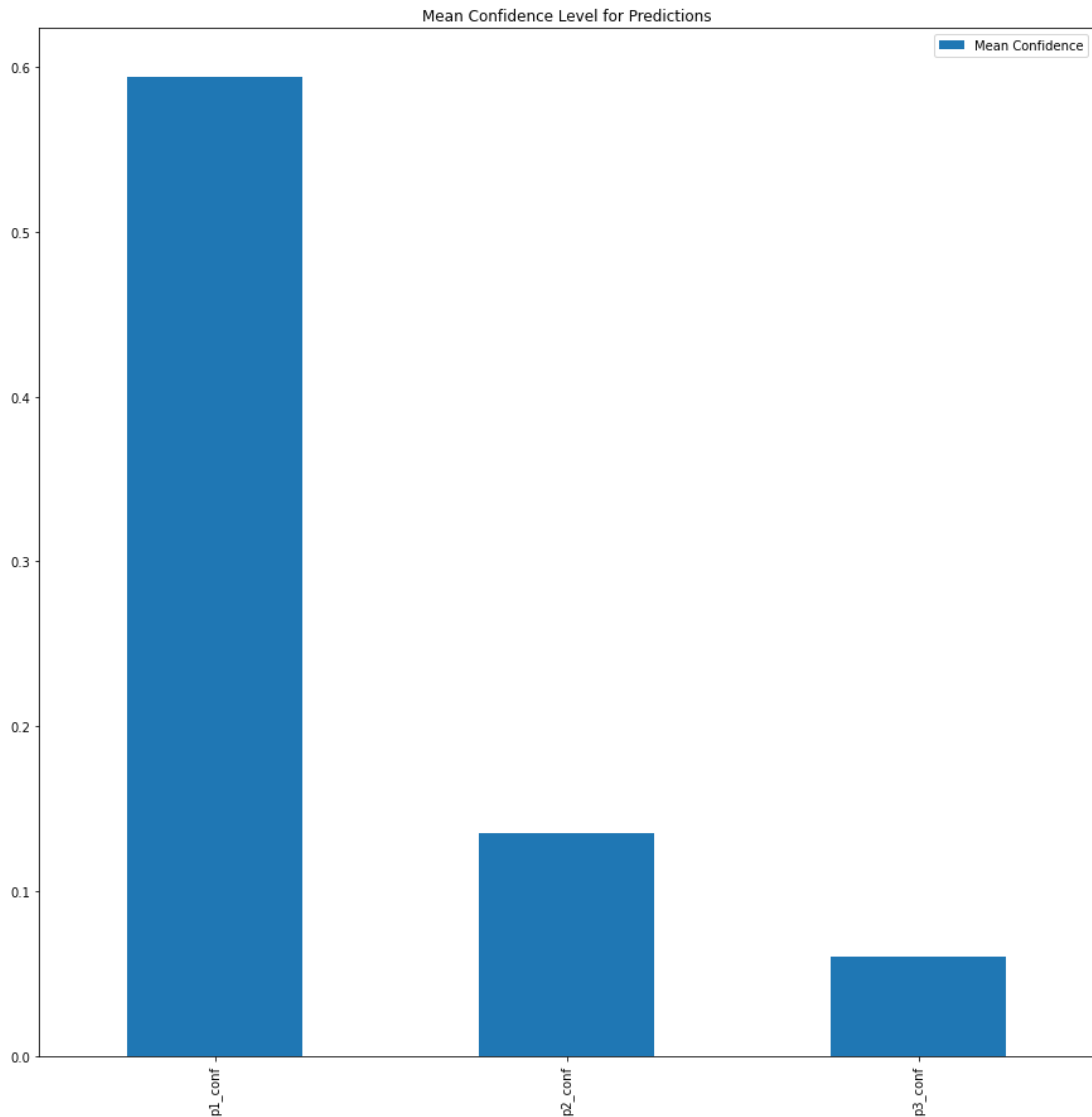


P3 Dog Predictions Accuracy



On average in the three predictions, about 1 in 4 predictions will be false. Of course, some of these are actually issues with computer vision, when the CNN could not recognize the dogs were separate objects from pillows, etc.

On average, how confident was the Convolved Neural Network used in the predictions when faced with the same image 3 times?



The network was around 59% confident in its prediction in the first instance. In the second instance it was only about 13% confident. In the last instance the network was only around 6% confident.

On average, the best prediction to go with for dog breeds would be the first one.

POSSIBLE CONCLUSIONS

The following are possible conclusions:

- I. It is possible to conclude that the most popular images are those of puppies or smaller dogs. So if you want your dog highly rated, join the trend and have your adorable puppy rated.
- II. The false image predictions were things like pillows and other items. This could mean that to get the best breed rating, a plain background for the dog would be better.
- III. Predictions made with the first prediction using the same machine learning model as was used on the dog images should be considered especially on the first prediction.
- IV. Higher ratings could cause more likes and retweets for any dog rated.
- V. Likes and retweets are mostly proportional.

