

WeRateDogs Data Archive - Act Report

The WeRateDogs Twitter archive contains data extracted from their twitter account, these data is about dog ratings that were taken from the text of the tweet or pictures uploaded to the tweet, using an algorithm to predict the breed of the dog from the image, and the stage of the dog from user's texts.

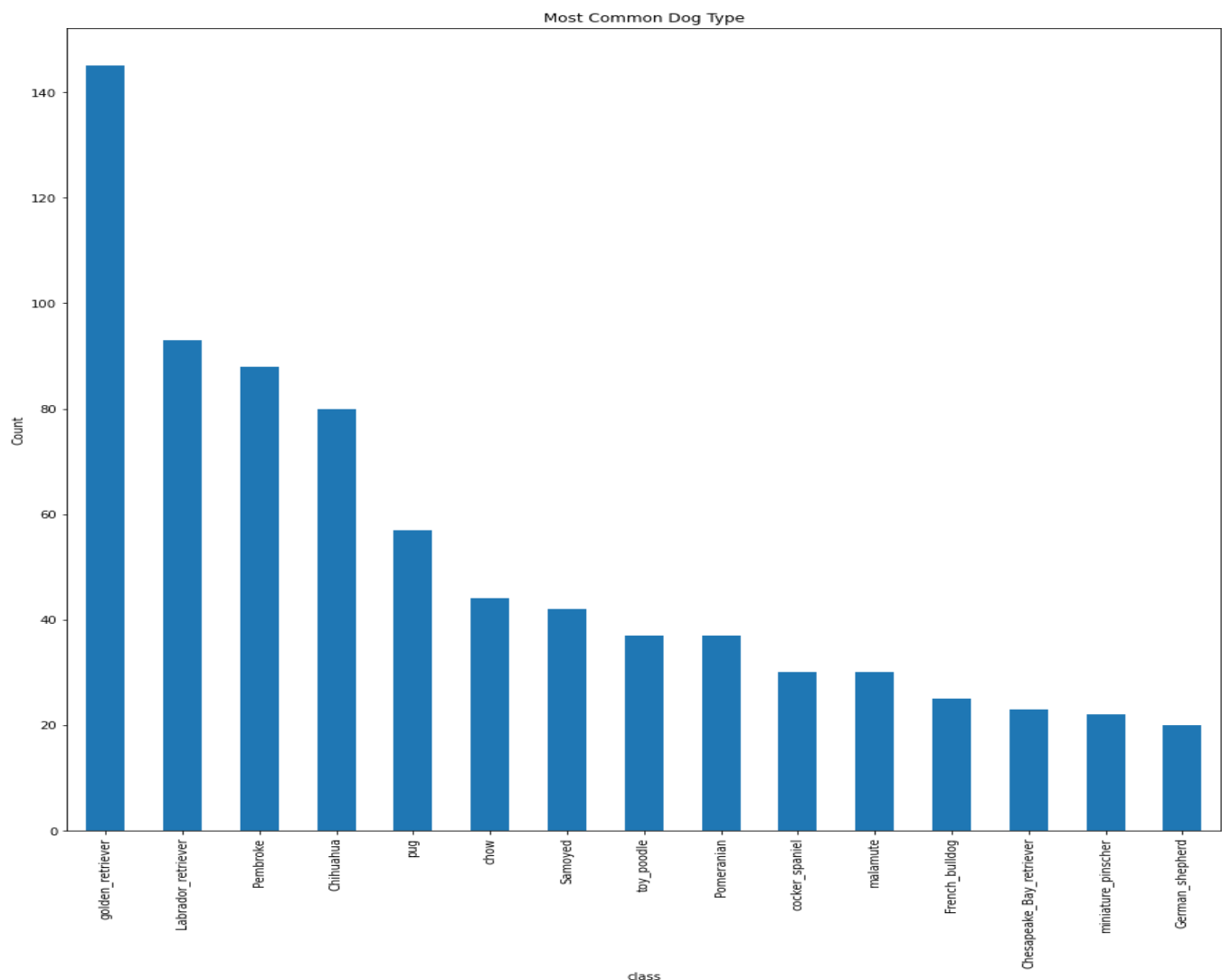
The data sets that have been used here was for a twitter downloaded archive, programmatic download for an image prediction file and later augmented by downloading more data using twitter API.

Before the analysis, the data had to be wrangled to generate the following insights. Data was assessed on both verticals visually and programmatically for quality and tidiness issues, After cleaning many of the issues found during the assessment, there were about 1489 record with good quality data, and here is what they tell us.

- **First Insight:**

“Golden retriever, and Labrador retriever are the most common dog according to WeRateDogs Data, and AI, While West Highland white terrier is the less common dog according to WeRateDogs Data, and AI”

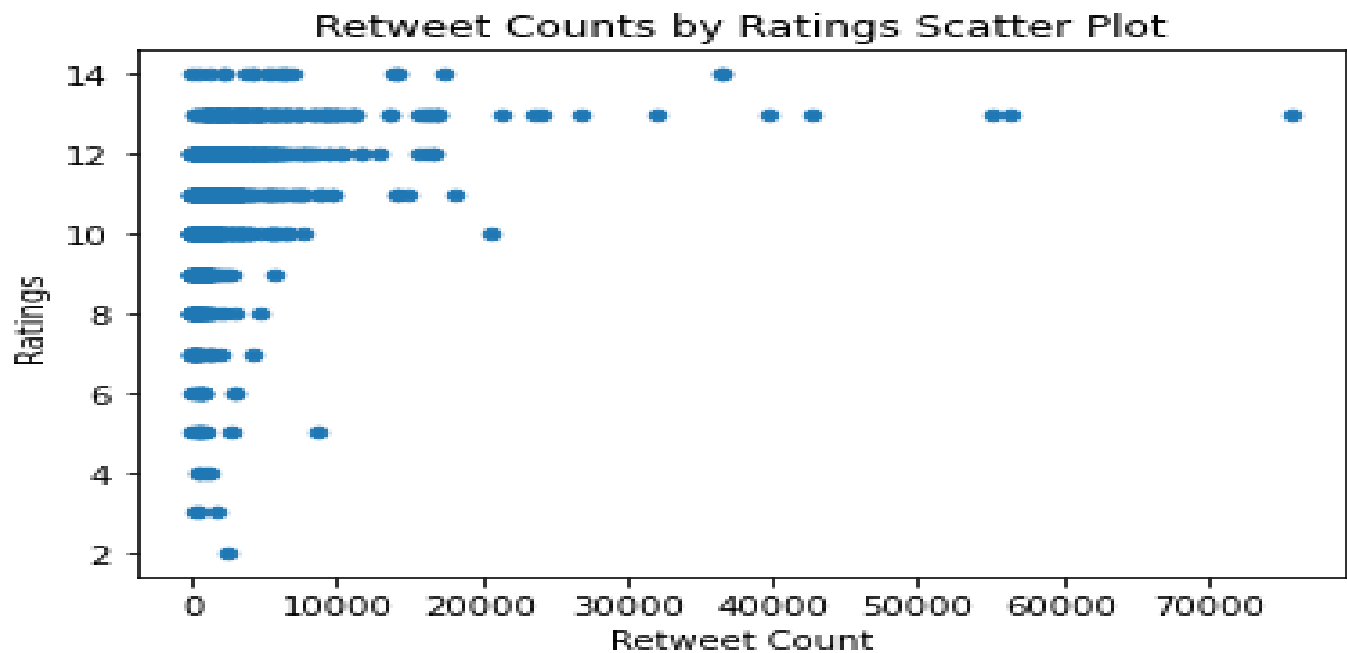
When the images from the tweets were fed through a neural network to predict the dog breed, we got predictions for 145 of Golden retriever, and 93 of Labrador retriever, and another few different breeds, Grouping all the tweets by the dog breed showed that the top 2 breeds of dogs with the most number of tweets are **Golden retriever, and Labrador retriever (see the graph below)**. Also there are 3 additional breeds in common: Pembroke, Chihuahua, and pug.



- **Second Insight:**

“Top Rated Dogs have less retweet count, and vice versa”

When you compare the relation between the top rated dogs based on the rating numerator after removing the outliers, you will find that the top rated dogs (or dog types) don't have the most retweets or favorites, (see the plot below)



- **Third Insight:**

After that if you check the screenshot below you will find that there is no correlation between the dog type popularity and the acquired ratings on @WeRateDogs twitter account, (see screenshot below), based on the following we find that

1. The top rated dog class is not the most common class.
2. The most rated dog breeds are Saluki, Tibetan mastiff, briard , Border terrier, standard schnauzer, and silky terrier

```
Out[244]: class_prediction
Japanese_spaniel      5.000000
soft-coated_wheaten_terrier  8.250000
Scotch_terrier        9.000000
Norwich_terrier       9.000000
Ibizan_hound          9.000000
...
standard_schnauzer    12.000000
Border_terrier        12.142857
briard                12.333333
Tibetan_mastiff       12.400000
Saluki                12.500000
Name: rating_numerator, Length: 110, dtype: float64
```

- **Fourth Insight:**

By looking at the two pictures below we will find out that **(Dog ratings has no correlation with the most common dog)**

Fig 1: (the average on the top rated dog breeds)

```
Out[244]: class_prediction
Japanese_spaniel          5.000000
soft-coated_wheaten_terrier 8.250000
Scotch_terrier            9.000000
Norwich_terrier           9.000000
Ibizan_hound              9.000000
...
standard_schnauzer        12.000000
Border_terrier            12.142857
briard                    12.333333
Tibetan_mastiff           12.400000
Saluki                    12.500000
Name: rating_numerator, Length: 110, dtype: float64
```

Fig 2: (The top rated dog types)

