

ACT REPORT ON THE TWITTER DATA  
ARCHIVE OF THE WERATEDOGS  
TWITTER ACCOUNT

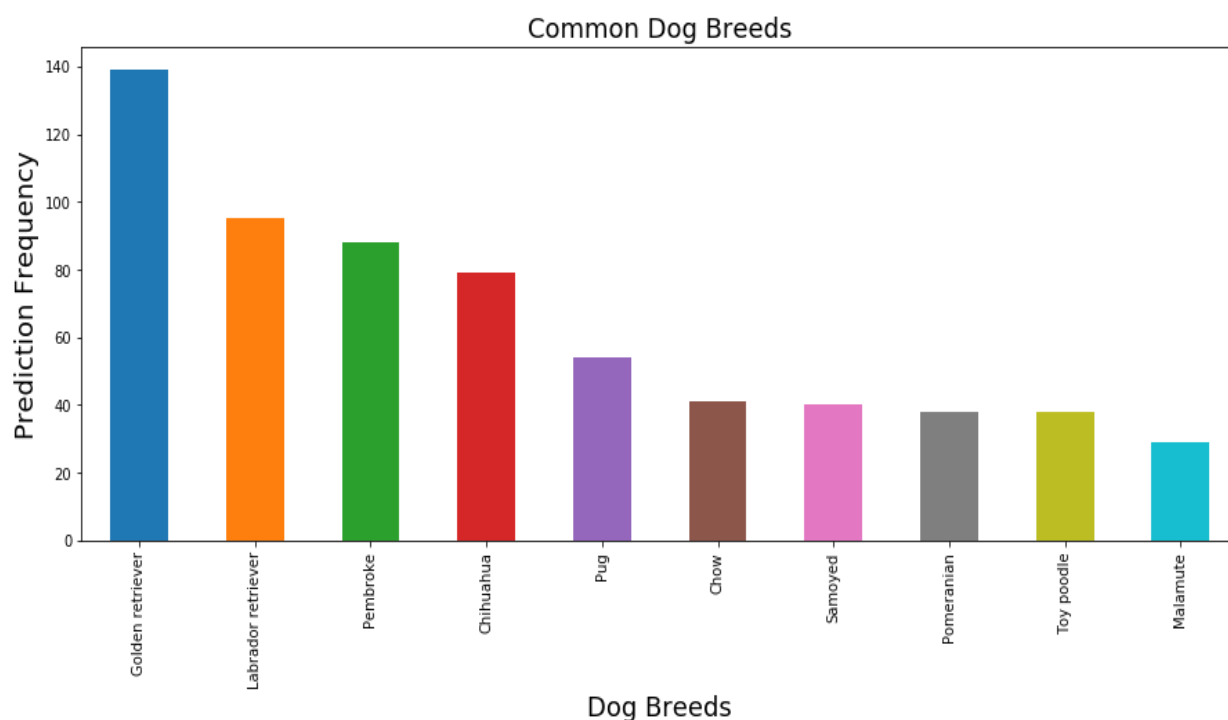
BY:  
OJODU UTHMAN

# Introduction

The aim of this report is to gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it using python and its libraries. The dataset I am attempting to wrangle is the twitter data archive of the user @dog\_rates also as WeRateDogs which was provided as one of the projects of the Udacity Nanodegree Program. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "[they're good dogs Brent.](#)" WeRateDogs has over 4 million followers and has received international media coverage.

After gathering, assessing and cleaning the entire dataset, I got the following results and made the following observations.

## Most Common Dog Breed

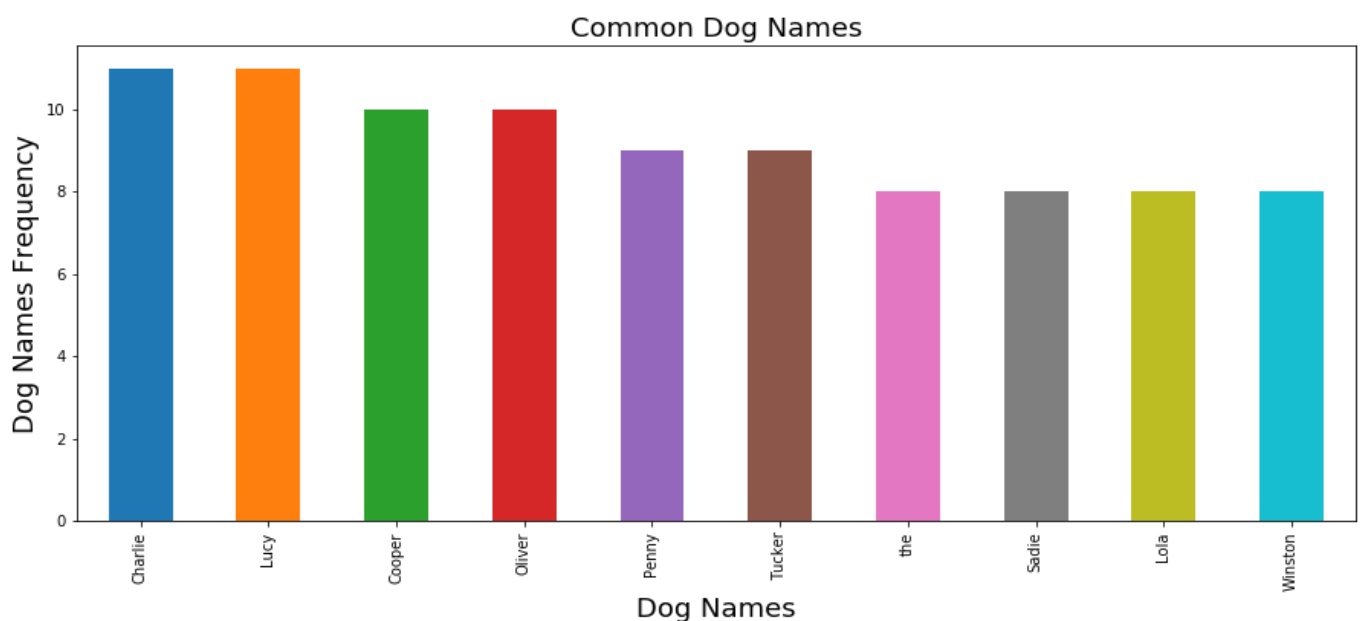


From the visualization shown above, the most common popular Dog Breed amongst the ones tweeted to the WeRateDogs accounts is the Golden Retriever. This was followed by the Labrador retriever and Pembroke breeds with the Malamute breed closing out the top 10.

## Least Common Dog Breeds.

I found that there are up to 7 really uncommon dog breeds (Standard schnauzer, Japanese spaniel, Groenendael, Scotch terrier, Silky Terrier, Entlebucher, and Clumber) that were being tweeted.

## Most Common Dog Name



The above visualization shows that the names Charlie and Lucy are the most common dog names on the WeRate Dogs twitter archive. These are closely followed by the names Cooper and Oliver with the name Winston closing out the top 10.

## Most Retweeted Dog Images

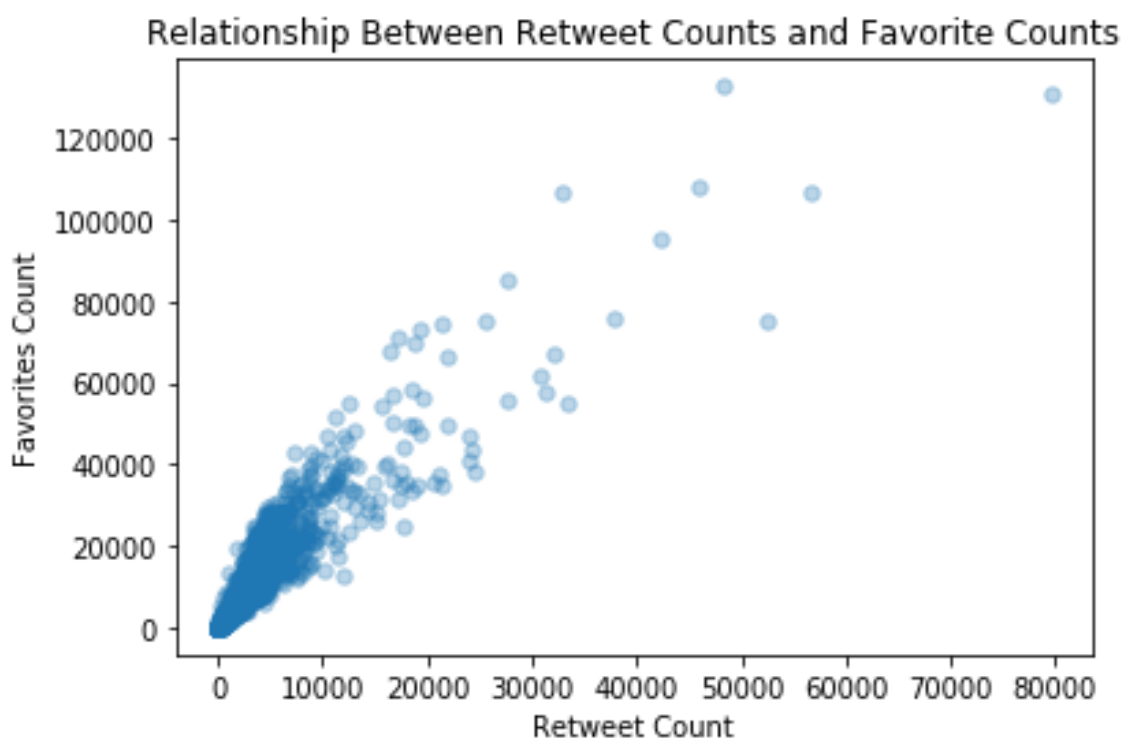




The above three images are the three top retweeted dog images on the WeRate Dogs Twitter account archive.

# Relationship between the Retweet and Favourites Count

This was both calculated and shown visually. The calculations gave us a correlation value of 0.911732 which shows a very strong relationship between the two columns. We confirmed this by drawing a scatter plot of the two variables which confirmed the initial strength of the correlation between the two variables.



## Conclusion

I found the most common dog breed to be the Golden retriever, the least common breeds to be seven different breeds. I also found the relationship between the retweet count and favourites count to be a strong one implying that a high retweet number is almost always followed by a high favourites number.

The limitations of this analysis includes the many null values in some columns which were not recoverable and some accuracy problems with some of the column (such as some of the dog names being a, an and so on) which makes this analysis a bit limited.