

## **WRANGLING AND ANALYSING THE ARCHIVED TWITTER DATA OF WERATEDOGS.**

**By Olamide Olowoniyi**

The dataset contains data downloaded from the twitter archive of @weratedogs also known as @dog\_rates, a twitter account popularly known for rating people's dogs using funny comments. Like many real-life data, this downloaded data does not come clean. The goal of the project is to use python and its libraries in order to create interesting and trustworthy analyses and visualisations.

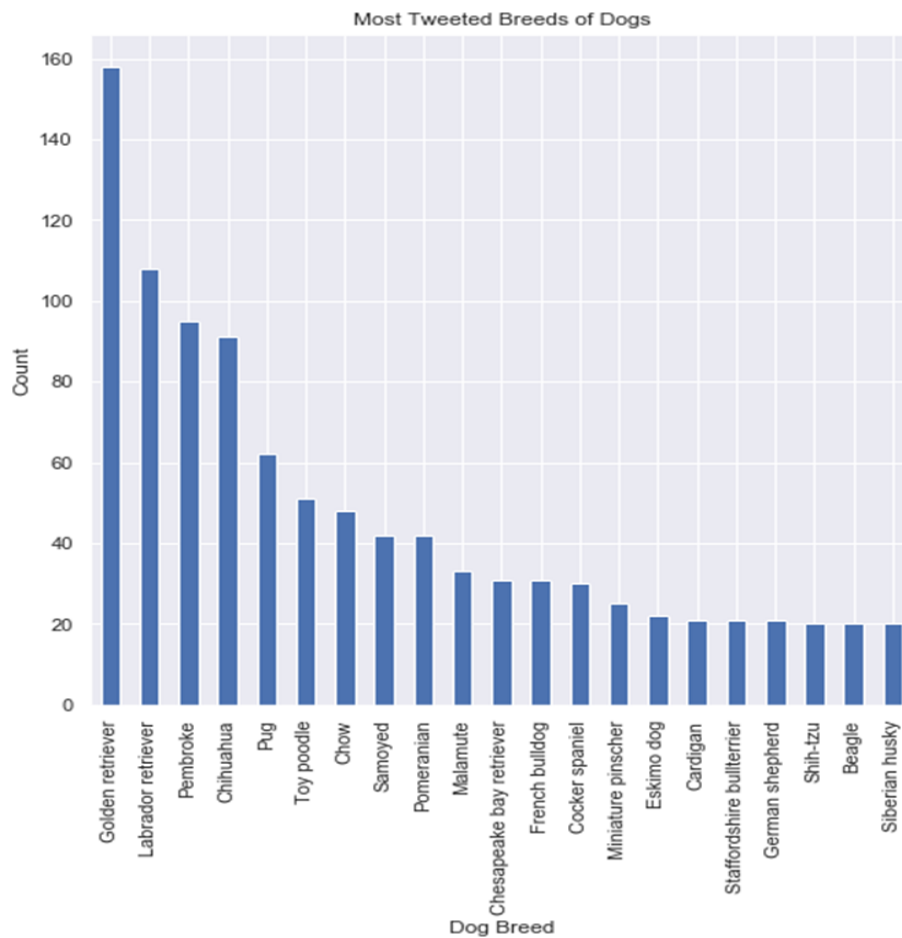
This report contains an analysis and a visualisation of the cleaned weratedogs data, including;

1. The most popular breed or most tweeted breed of dogs
2. The average ratings and average favourite counts for the most tweeted breed of dogs.
3. The average ratings and average favourite counts for the different dog stages.
4. Distribution of the different dog stages
5. The relationship between favourite counts and retweet counts

### **INSIGHT AND VISUALISATION:**

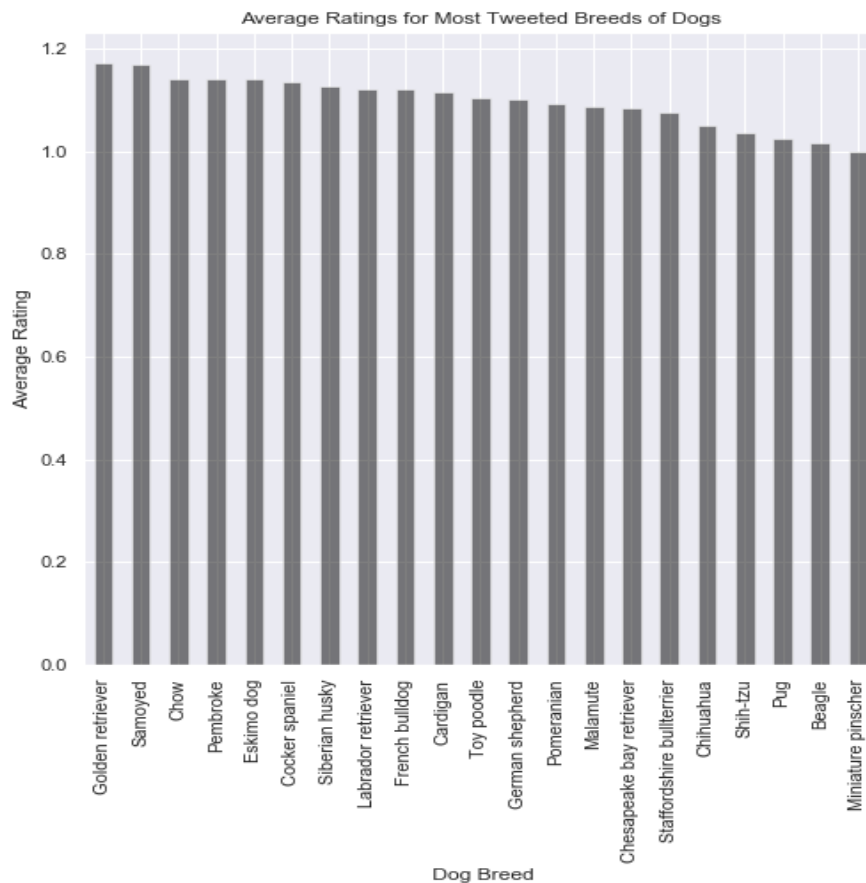
#### **The most tweeted breed of dogs**

I plotted a bar chart to check for the most tweeted about breed of dogs-this can also be said to be the most popular breed of dogs from the weratedogs account, taking into account only the breeds of dogs with at least twenty tweets. The golden retriever appeared to be the most tweeted with the Labarador retriever, Pembroke, Chihuahua and Pug coming in as the next top four.



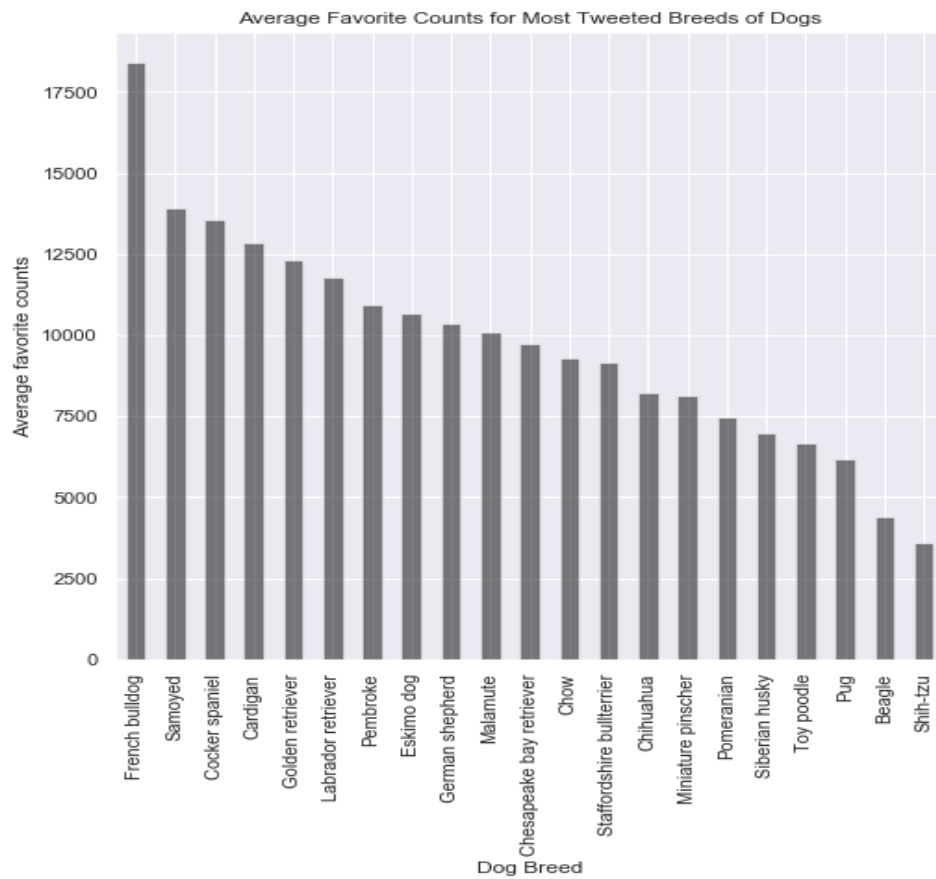
### The average ratings for most tweeted breed of dogs

From the plot, the top 5 most popular breeds of dogs with the highest ratings are in this order in order: Golden retriever, Samoyed, Chow, Pembroke and Eskimo Dog.



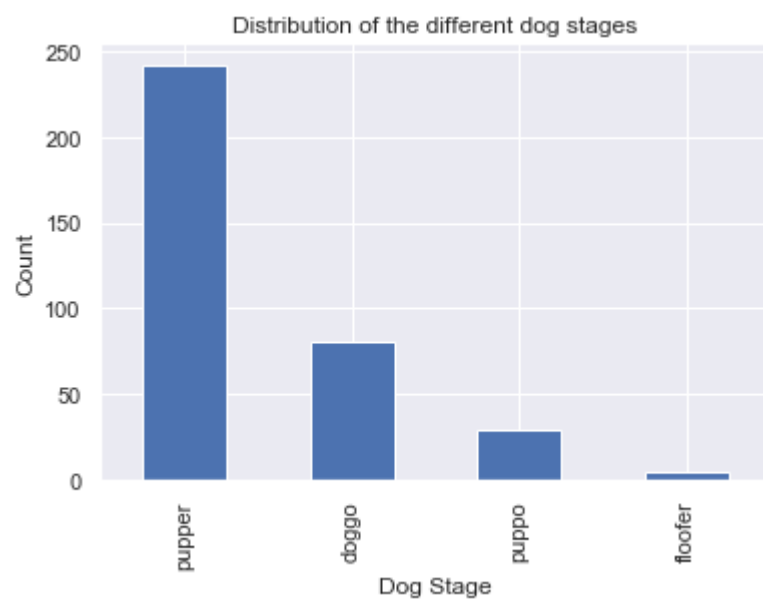
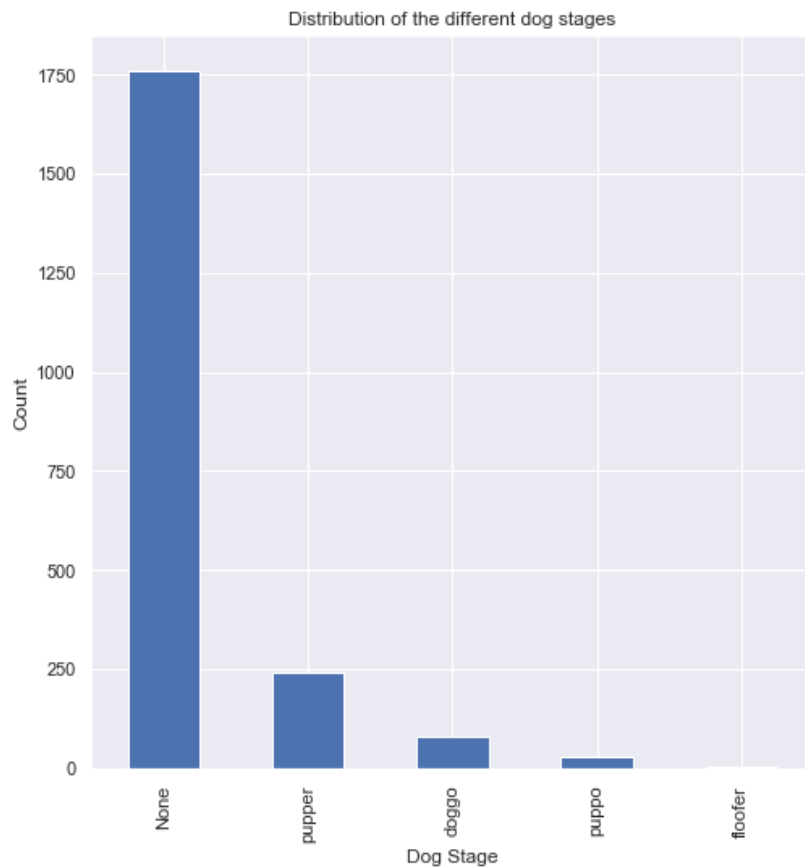
### The average favourite counts for most tweeted breed of dogs

The result showed that the top 5 most tweeted dog breeds with the highest number of likes are French Bulldog, Samoyed, Cocker Spaniel, Cardigan, and Golden Retriever. There are observed differences between most tweeted dog breeds with highest number of likes and those with highest ratings, while the dogs common in both groups are Samoyed and Golden Retriever.



### Distribution of the different dog stages

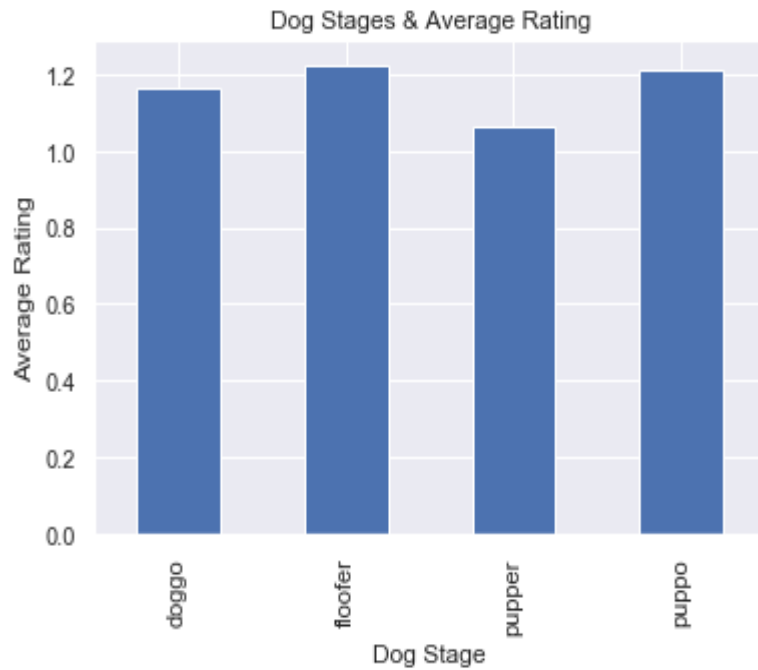
The result showed a really high number of dogs with unknown dog stage (this may be because there is a need for further cleaning in this section of the data). However, for the scope of this project, I just dropped the 'None' category.



From here, it was seen that most of the dogs are Puppies

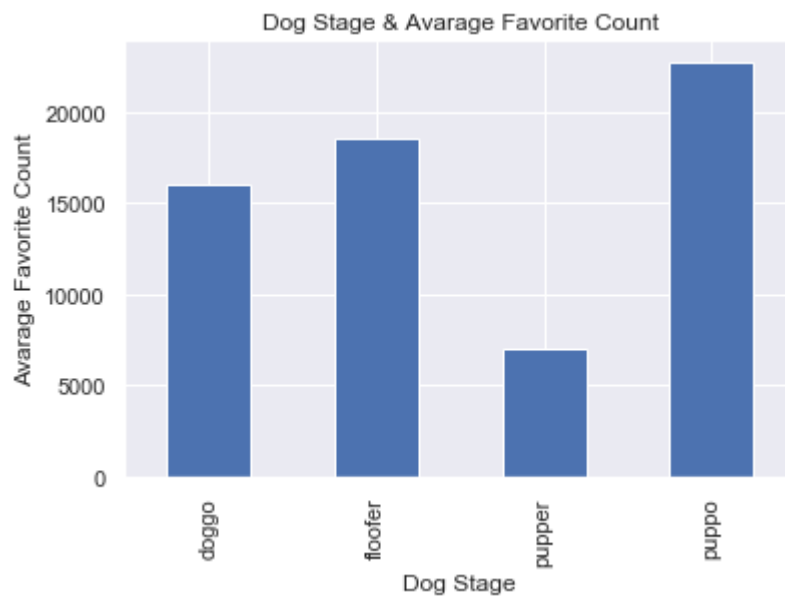
### The average rating for dog stages

The Floofer has the highest average rating, followed very closely by the Puppos. Over all, they all have very similar ratings.

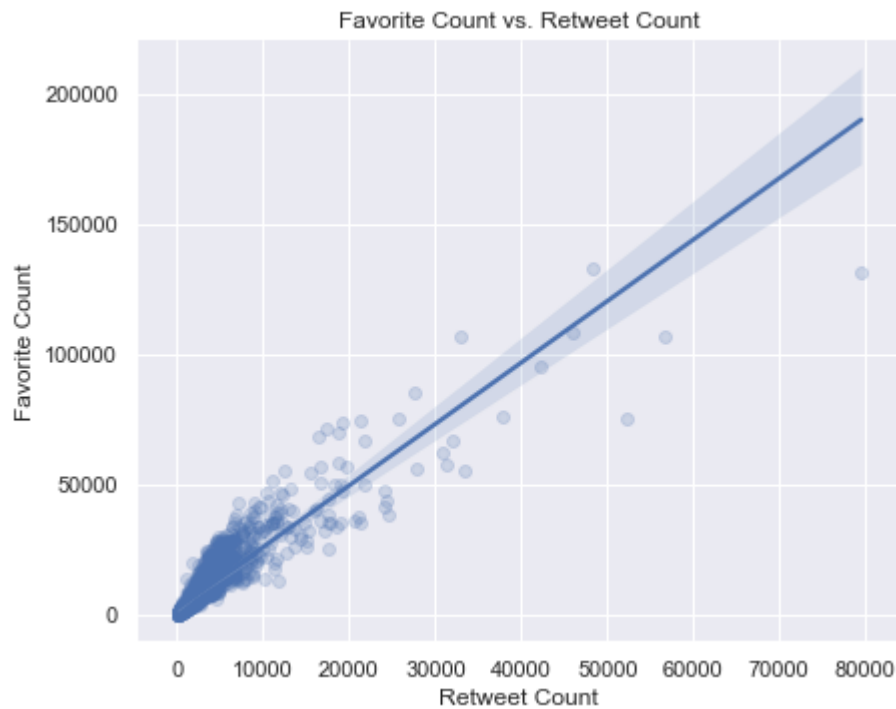


### The average favourite count for dog stages

Here, it can be clearly seen that the Pupperos have the highest number of likes



### The relationship between retweet count and favourite count



The scatterplot revealed that favourite counts and retweet counts are positively correlated, with the majority of the data falling below 50000 favourites and 10000 retweets.

In conclusion,

Since the weratedogs is generally a humorous account, it is understandable that we have the results we saw in the **average ratings for dog stages** where they mostly had similar and good ratings. Also, with the **average favourite count for dog stages**, result showed that Puppies have the highest favourite counts, even though the Puppies are more tweeted about, we can see they have the lowest rating and favourite counts. Overall, I believe I cleaned a significant amount of data but the dataset still needs more work e.g. the dog stages.