

# WeRateDogs

By Bukola Ajayi



This project is focusing on providing insights and visualisations from data gotten from WeRateDogs Twitter.

We would also be making use of "image\_predictions.tsv" hosted by Udacity to enhance the data gotten and use the predictions to make informed decisions.

We will be drawing some insight from the analysis done on this project.

## Processing our Data

The "WeRateDogs" Twitter archive enhanced csv: This file will be downloaded manually by clicking the following link provided.

The tweet image predictions tsv: This file image\_predictions.tsv is hosted on Udacity's servers and will be downloaded programmatically as well, following the link provided;

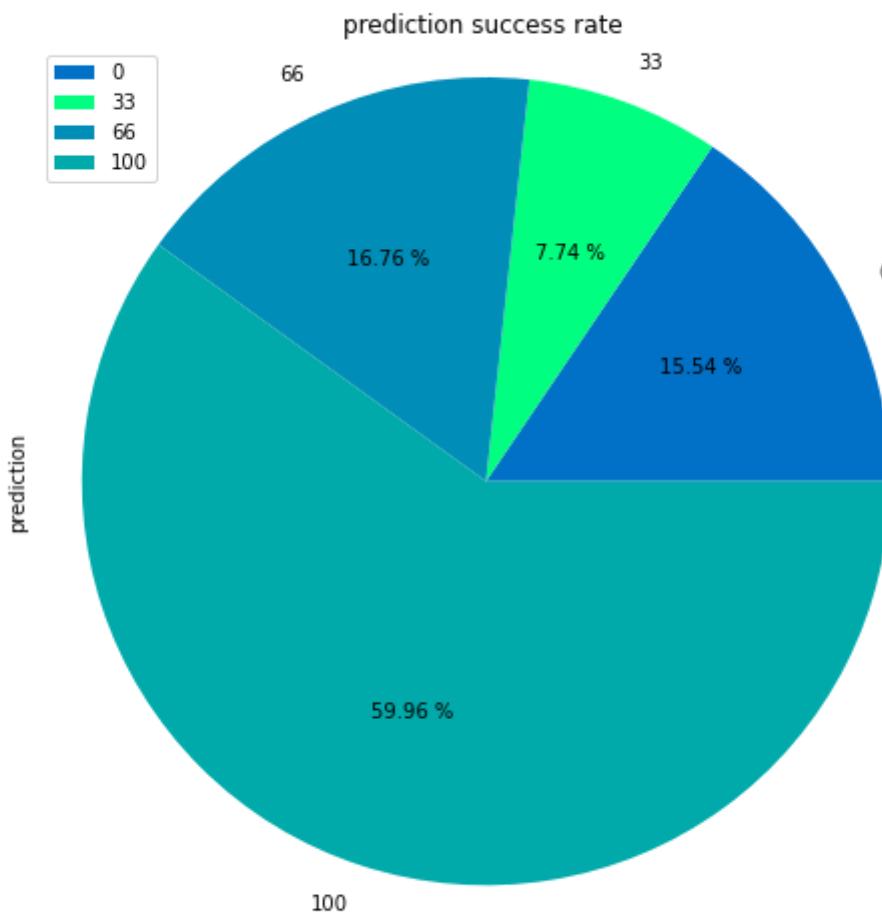
Using the tweet ID's gotten from the WeRateDogs Twitter archive csv, we will query the Twitter API for each tweet's JSON data using Python's Tweepy library and store each tweet's entire set of JSON data in a file called tweet\_json.txt stored as a txt file.

This gives a total of three files.

These three files were analyzed, cleaned up, and then brought together by merging them based on their unique identifiers to give a clean master file which was used for our analysis

## Percentage prediction of a dog to be true

According to the data gotten, we observed that not all predictions were dogs, so the success rate of dog prediction became apparent. We would see that 60% success rate for all prediction was true.



## The 20 most popular tweeted dog breeds

A total of 379 dogs were identified however, after drawing visualisation and gotten insight on the top 20 popular dogs mostly tweeted about, the image below indicates that the golden retriever breed tops the list.



golden\_retriever



Labrador\_retriever



Pembroke



Chihuahua



pug



chow



toy\_poodle



Pomeranian



Samoyed



cocker\_spaniel



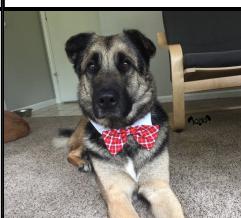
malamute



miniature\_pinscher



French\_bulldog



German\_shepherd



Chesapeake\_Bay\_retriever



Siberian\_husky



Shetland\_sheepdog



Eskimo\_dog



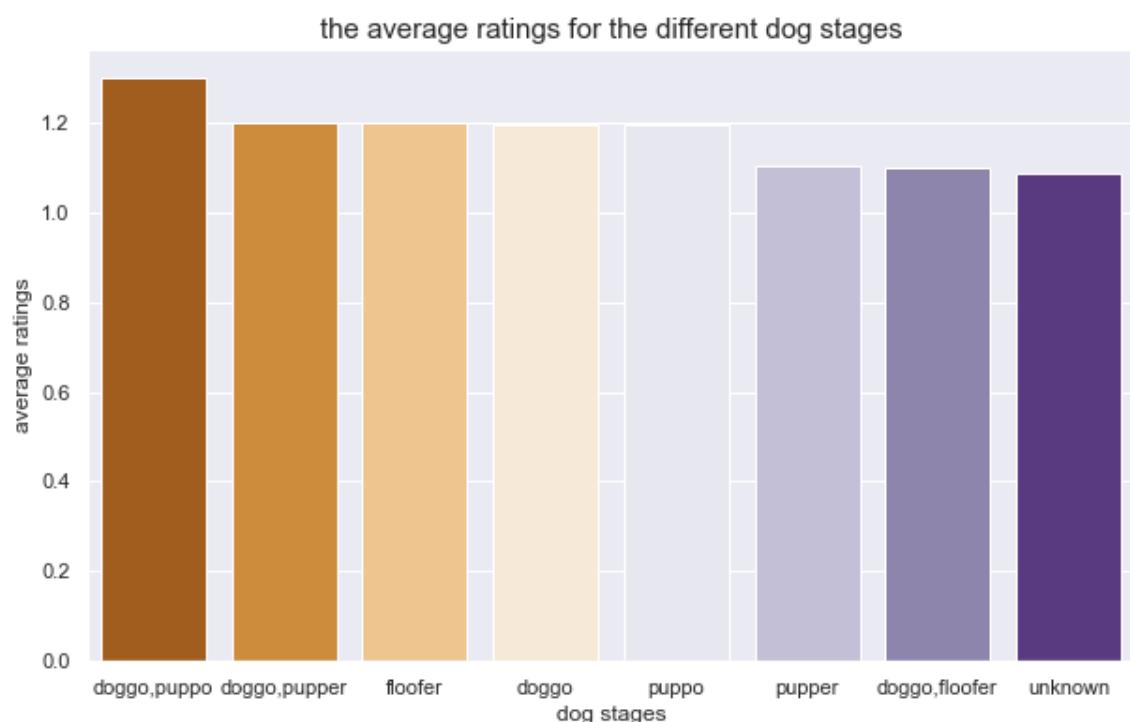
beagle



Cardigan

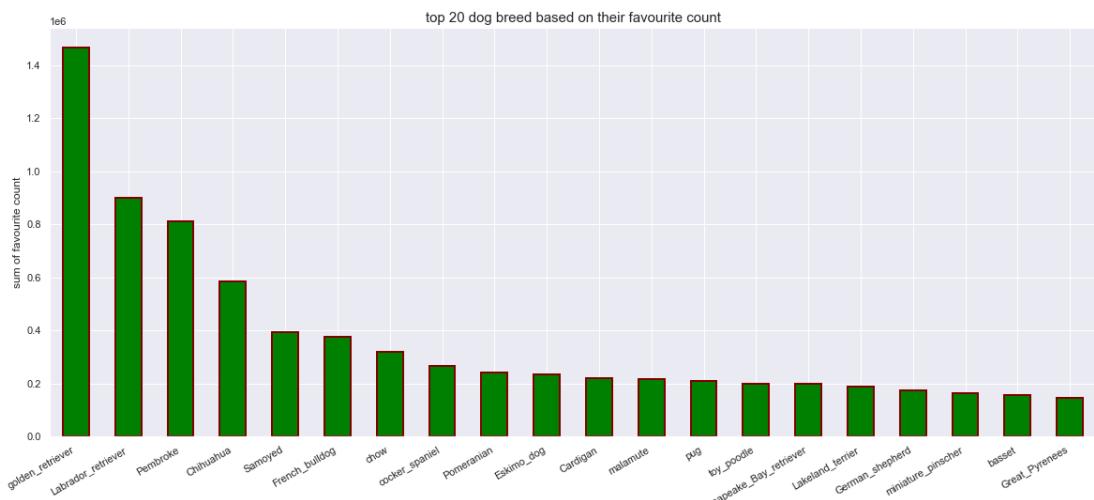
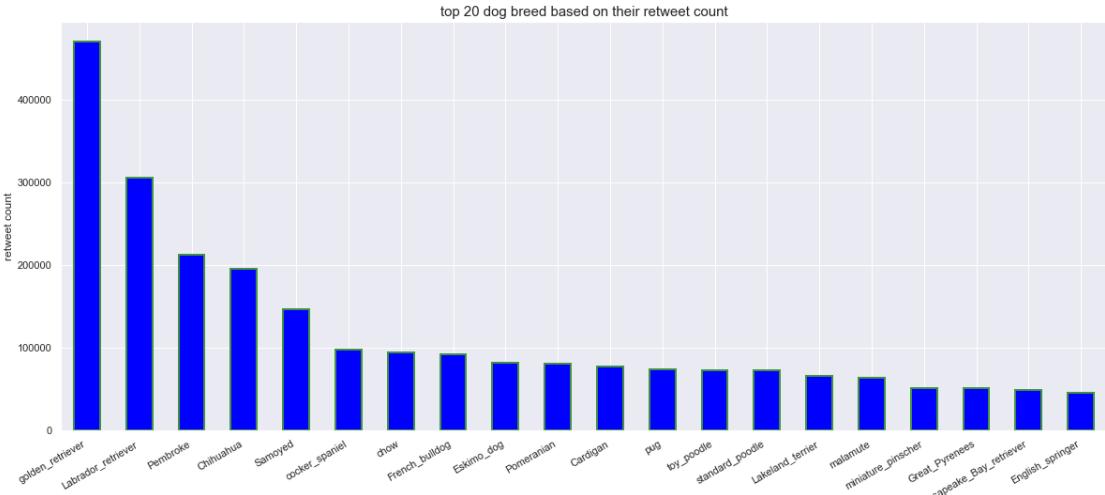
## Most tweeted 'Dog Stage'

Some of the tweets contains multiple dog stages as some of the tweet contains several dogs. The following graph shows the most tweeted dog stages. We can clearly see that the doggo,puppo stage has the highest average dog ratings.



## The top 20 dog breeds based on retweet and likes

As both of the visualizations of below shows, the top 20 most common dogs on WeRateDogs coincide with the most favorite and retweet dog breed. It can be seen that Golden retriever is the most popular dog and it along with Labrador Retriever, Pembroke, Chihuahua and Samoyed complete the top 5 common dog breeds in the data.

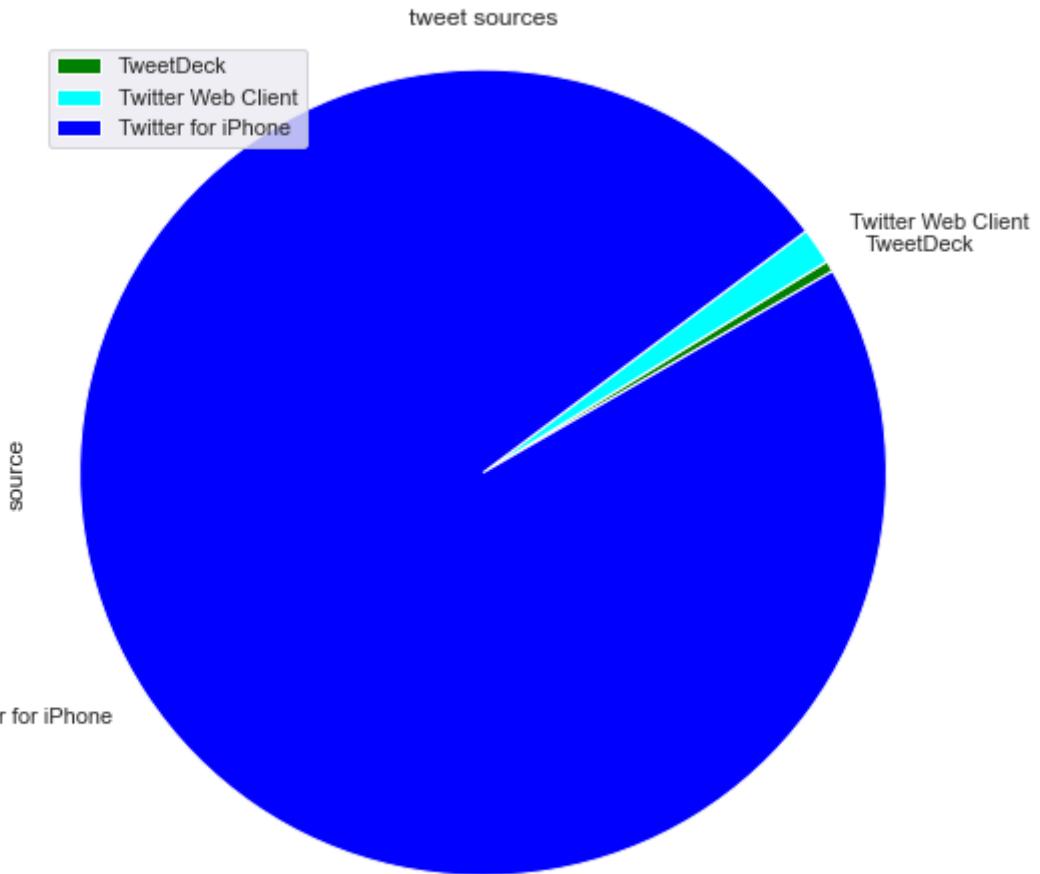




Sample image of Golden retriever breed gotten from the dataset

## The most used devices in this twitter project

The following bar plot shows that most users are iPhone users, the indicates that the largest portion of our dog lovers were iPhone users.



## The top 20 dog breeds based on retweet and likes

As both of the visualizations of below shows, the top 20 most common dogs on WeRateDogs coincide with the most favorite and retweet dog breed. It can be seen that Golden retriever is the most popular dog and it along with Labrador Retriever, Pembroke, Chihuahua and Pugs complete the top 5 common dog breeds in the data.

From the other insights derived from this analysis, I found that there is a strong positive relationship between the favourites count and the retweets count of a given tweet.

## Conclusion

This project was mostly a data wrangling project. During the course of the data cleaning, all identified issues have been fixed, and I have also merged separated data frames into one and added some missing values. The final data frame was stored as `twitter_archive_master.csv`. From the overall exploratory analysis, certain conclusions have been drawn, which include the following:

Based on the data gathered, I have assessed the most evident issues (17 issues in total) and documented them to create a record of modifications. Later, in the data cleaning process, I fixed all

identified issues to complete, and I also merged separated data frames into one and added some missing values. The final data frame was stored as `twitter_archive_master.csv`.

The Golden Retriever breed, which seems to be a popular choice for pet dogs, is the most liked dog based on its interestingly high retweet and favourite count.

There is a strong linear relationship between the number of favourites and the number of retweets of a tweet, with a positive correlation coefficient of 0.86. This relationship might already be established as people will favourite what they retweet or the other way around.

The Doggo stage has the highest average ratings.

The dog breed with the highest average rating is the clumber. This indicates that the average rating of a dog is independent of the retweets and the likes.