

ACT REPORT

After wrangling and obtaining the master dataset, `twitter_archive_master.csv`, the dataset was finally ready for analysis. Some insights we wished to obtain from the datasets were:

Insights:

1. The popular dog breeds
2. Dog distribution among the four stages
3. The common dog names
4. Relationship between likes and retweets
5. what is the most popular source of tweets for our dataset?

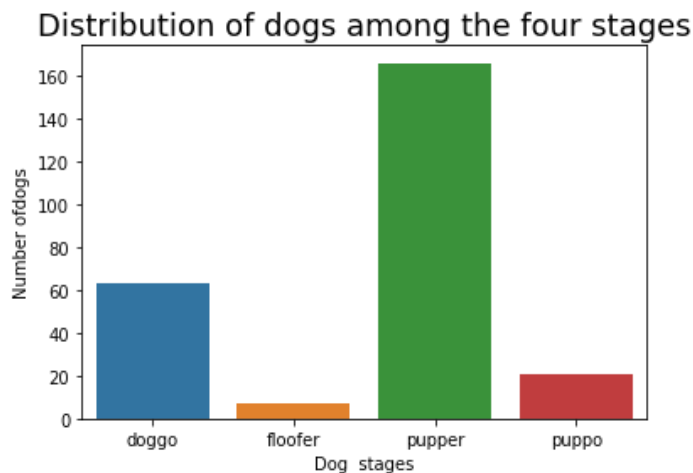
- The common dog names

From our analysis, Charlie, Lucy and Cooper appear to be popular names for the dogs in the dataset with 10 dogs bearing each name respectively, followed closely by Oliver and Tucker with 9 each.

- What is the most popular source of tweets for our dataset?

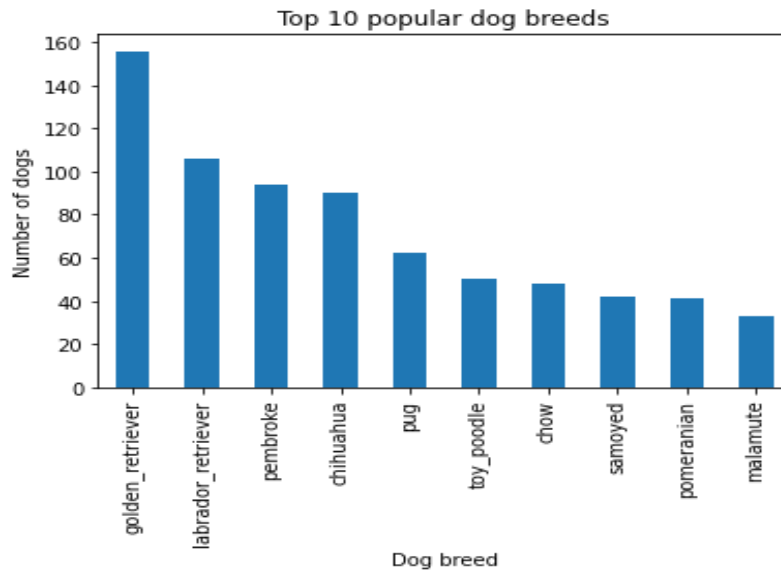
Most of the tweets came from an iphone as it makes up the majority with 1635 tweets out of 1666 coming from it. Other tweet source includes: twitter web client and tweet deck.

- Dog distribution among the four stages



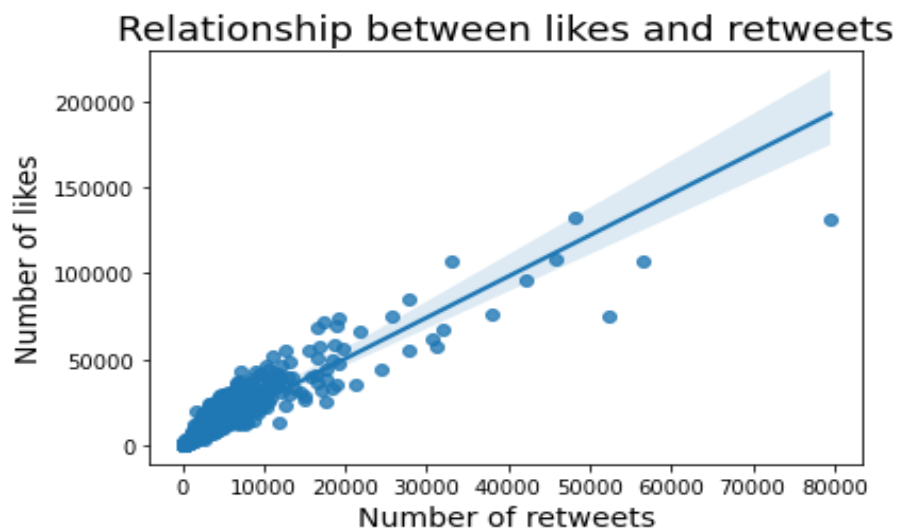
From the bar chart above, it can be observed that majority of the dogs are in the pupper stage followed by the doggo stage. The least dog stage is the floofer stage which has less than 20 dogs in it.

- The popular dog breeds



The bar chart shows the top 10 popular dog breeds with the golden retriever being the most popular dog breed followed by the labrador retriever.

- Relationship between likes and retweets



From the reg plot above, there is a positive relationship between likes and retweets, the more the retweets, the higher the number of likes a tweet gets.