

# WRANGLE AND ANALYZE DATA PROJECT

## ANALYSIS AND VISUALIZATION REPORT

Adedamola Babayato | Udacity Data Analysis Nanodegree | June 22, 2022



### Table of Contents

- Introduction
- Report Aim
- Analysis and Visualization
- Conclusion

### Introduction

The dataset WeRateDogs, is the tweet archive of Twitter user @dog\_rates. WeRateDogs is a Twitter account that rates dogs with humorous comment about the dog. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) This Project worked through the entire data wrangling process, gathering, assessing and cleaning data. However, there are some insights that I derived from this data.

### Report Aim

The aim of this report is to communicate the findings I derived from the cleaned master dataset saved to CSV file twitter\_archive\_master.csv and display visualizations produced from the wrangled data.

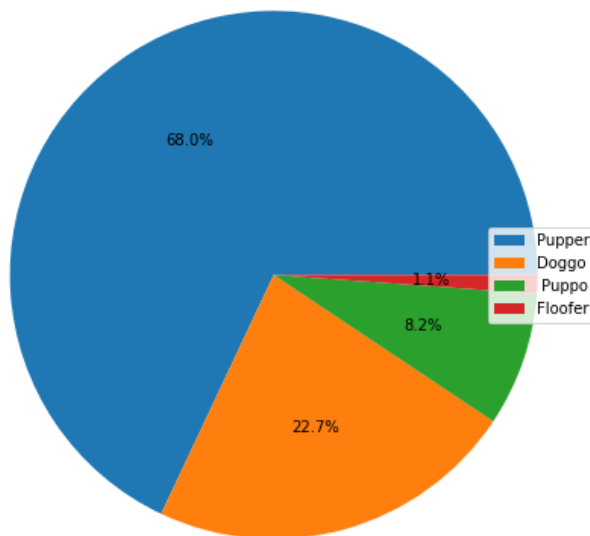
## Analysis and Visualization

I performed exploratory data analysis on the cleaned master dataset and I came about these three insights:

1. Popular Dog Stage by percentage
2. Most popular Dog breeds
3. Most dog tweet favorited and retweeted
4. Relationship between favorite count and retweet count

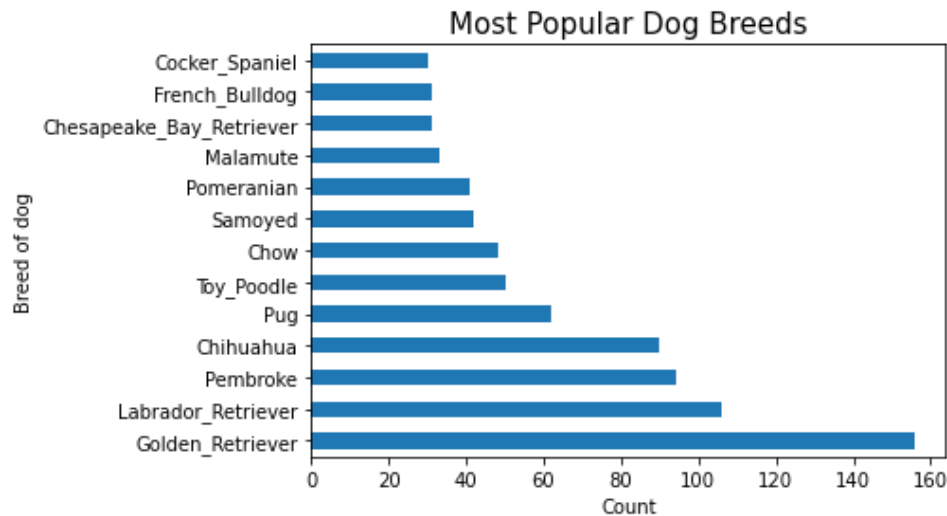
### Popular Dog Stage by Percentage

Percentage of different Dog Stages



Pupper is the Most Popular dog stage with 68% while Floofer is the least popular stage with 1.1%.

## Most Popular Dog Breeds



Golden Retriever is the most popular dog breed while Cocker Spaniel is the least popular dog breed.

## Most dog tweet favorited and retweeted

Performed the analysis on the favorite count and retweet count using Pandas' max() method to derive the most dog tweet favorited and retweeted. I discovered that the most favored dog tweet was from Lakeland Terrier breed, it was favored by 132810 users while the most retweeted dog tweet was from Labrador Retriever breed it was retweeted 79515 times.



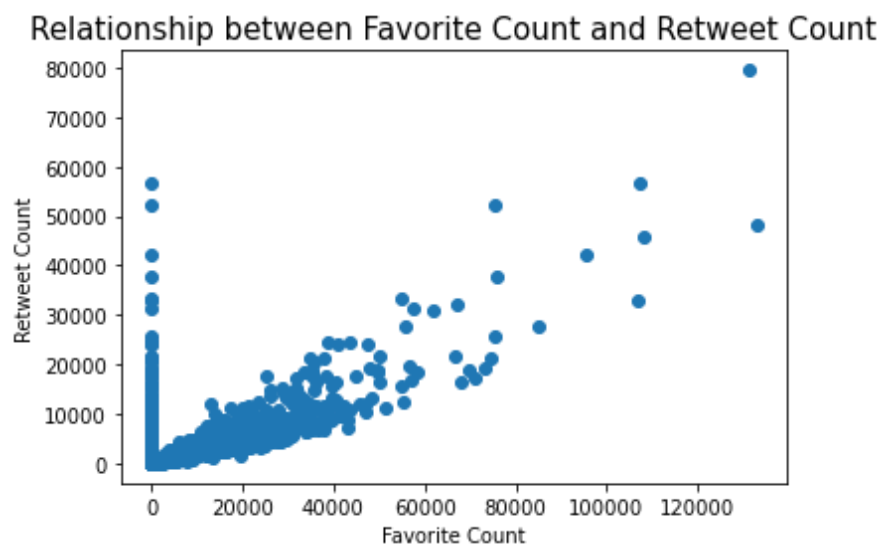
Lakeland Terrier

---

[ "Here's a doggo realizing you can stand in a pool. 13/10 enlightened af (vid by Tina Conrad) <https://t.co/7wE9LTEXC4>" ]

Favorite Dog Tweet

### Relationship between favorite count and retweet count



A positive correlation between favorite count and retweet count

### Conclusion

Overall, I was able to gather data from different sources, I assessed the data for quality and tidiness issues and cleaned the data to aid future analysis. I was able to derive three insights, however further research can still be done to uncover more insights as this is an interesting dataset.