# PROJECT REPORT ON WRANGLING AND ANALYZING DATA

**By Onyedikachi B. Ogbonna**

## Introduction

In the world of data science, data analysis (or any field that has to do with data), getting the data is one part, but Real-world data rarely comes clean (maybe due to mising values, duplicate rows, etc), and hence the hard part is mostly wrangling, cleaning or preparing the data before analyzing or feeding it to a model for training. In this project, I will be wrangling the data (Tweets) gotten from twitter (WeRateDogs) and then do some analysis.

I used the following python packed for my project

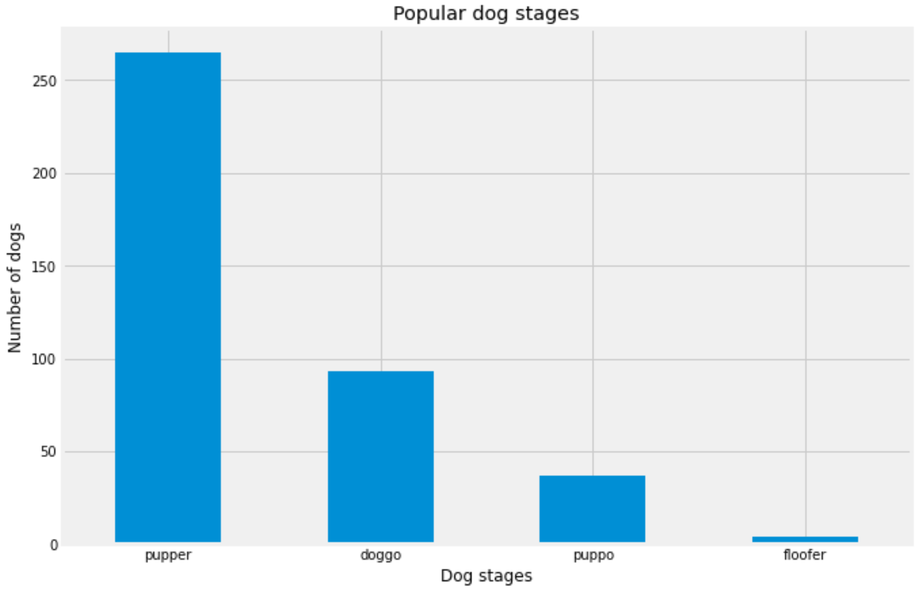
- pandas
- NumPy
- matplotlib
- json

So, after the datasets were cleaned and merged, I did some analysis.

## What is the most popular dog stage?

This question can be answered by plotting a bar chat showing all dog stages and the number of dogs that fall under that stage.

| | |
|---|---|
| • pupper | 265 |
| • doggo | 93 |
| • puppo | 37 |
| • floofer | 4 |

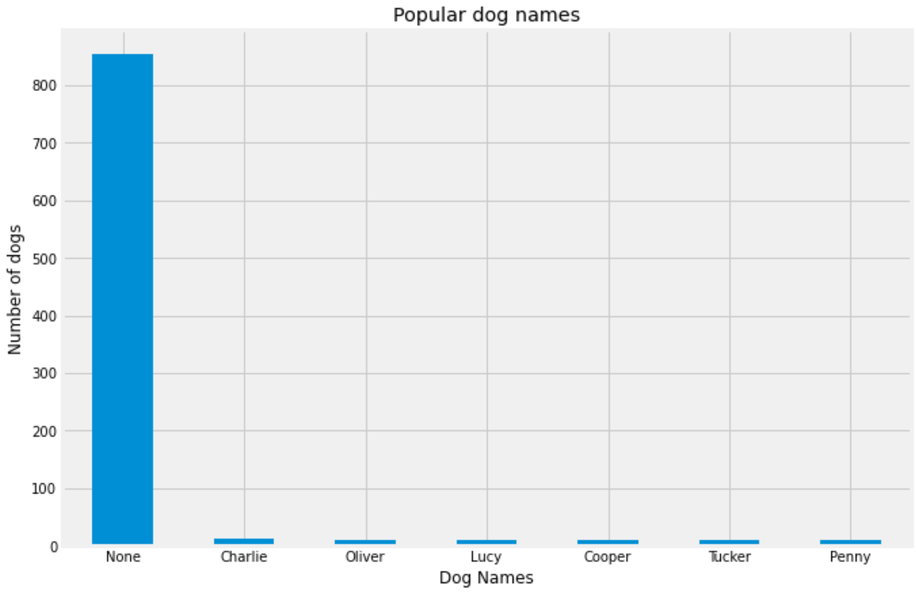


We can see that pupper has the highest amount of dogs

### What is the most given dog name?

Although this column had so many missing values and some dog names were not filled and hence has the None value. But we can still get a good number of the most given dog names.

| | |
|---|---|
| None | 854 |
| Charlie | 12 |
| Oliver | 11 |
| Lucy | 11 |
| Cooper | 11 |
| ... | |
| Rudy | 1 |
| Georgie | 1 |
| Raphael | 1 |
| Gerbald | 1 |
| Augie | 1 |



For some reasons, there seems to be too many names missing here, hence we got none to be the highest, followed by Charlie, etc.

### What is the most popular source people use to post on twitter (from the dataset)?

| | |
|---|---|
| Twitter for iPhone | 2221 |
| Vine - Make a Scene | 91 |
| Twitter Web Client | 33 |
| TweetDeck | 11 |

From our result, most users post from an iphone.

## Conclusion

Visualizations are very important for analysis as it gives one an idea of what the dataset is all about, or even see relationship between features.

Many questions can be asked about this dataset to get more insights about it. We may want to also check for the most popular dog breed, the number of tweets over time, etc.