

# Data Analysis and Visualization

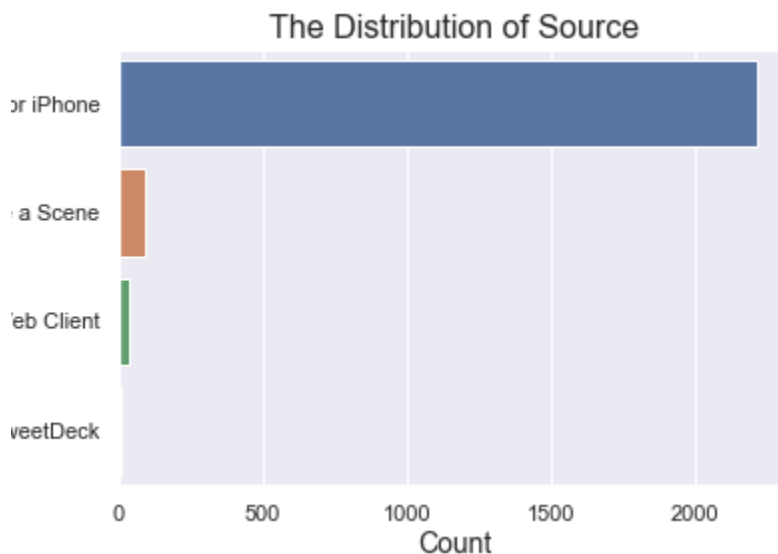
23 May 2019

## Introduction

This act report includes the basic data analysis of WeRateDogs twitter account data. I analyzed the information in the clean, combined data frames and then created customized visualizations of my findings using Matplotlib in Python.

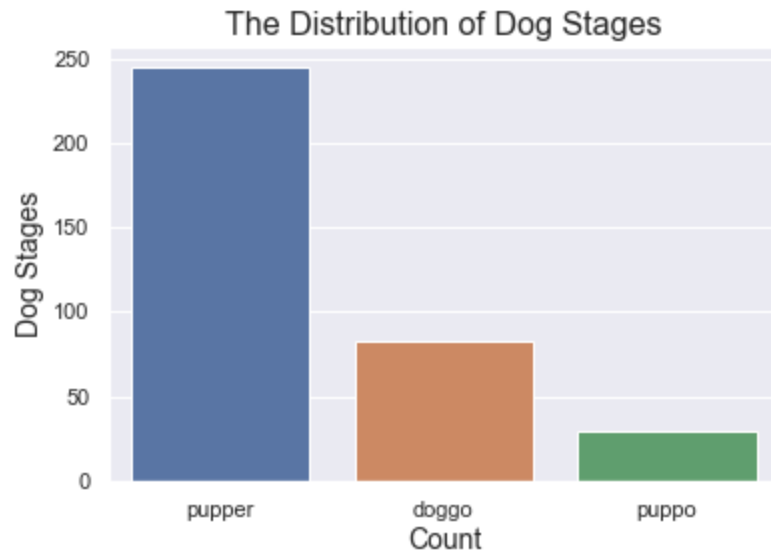
### 1. The Distribution of Source:

The dominate source of tweets is from iPhone twitter app, which is 94% in the total. That means the twitter app is the main channel for people using to tweet, retweet, post, and others, while the TweetDesk is pretty rare (less than 1%).



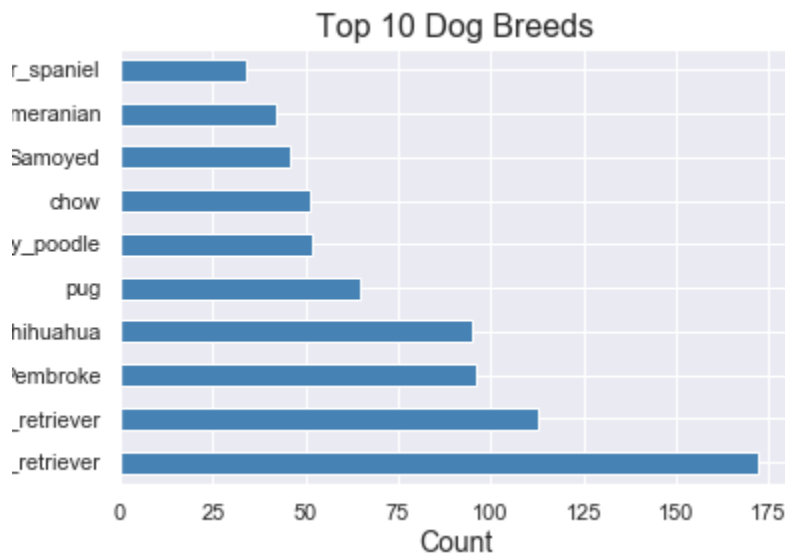
### 2. The Distribution of Dog Stages:

It shows that 'pupper' (a small doggo, usually younger) is the most popular dog stage, followed by 'doggo' and 'puppo' comes in the last stage.



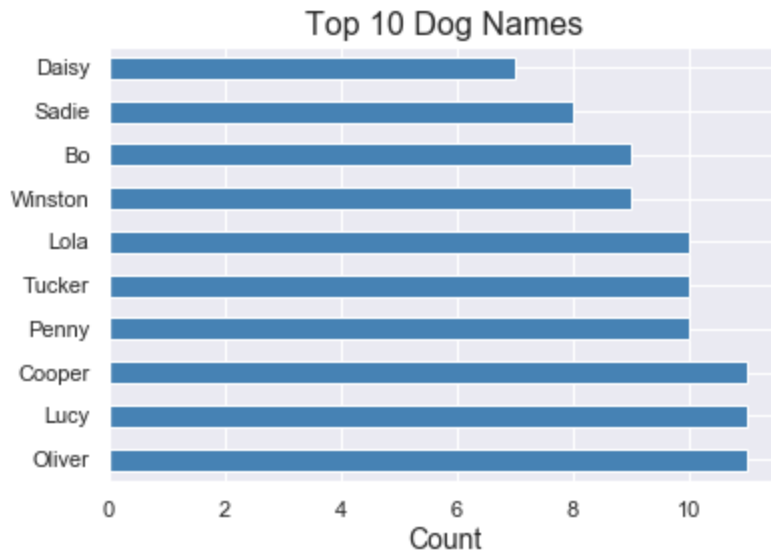
### 3. The Top 10 Dog Breed:

There are more golden Retrievers than any other dog in the dataset. Labrador Retrievers are the second most common.



### 4. The Top 10 Dog Names:

Oliver, Lucy and Cooper are the most popular dog names.



## 5. The relation between retweets and favorite count:

There is a strong positive correlation between number of retweets and favorite count. That is reasonable, the more a post is retweeted, the more eyes view the post, the more favorites the post receives

