Data Analysis and Visualizations

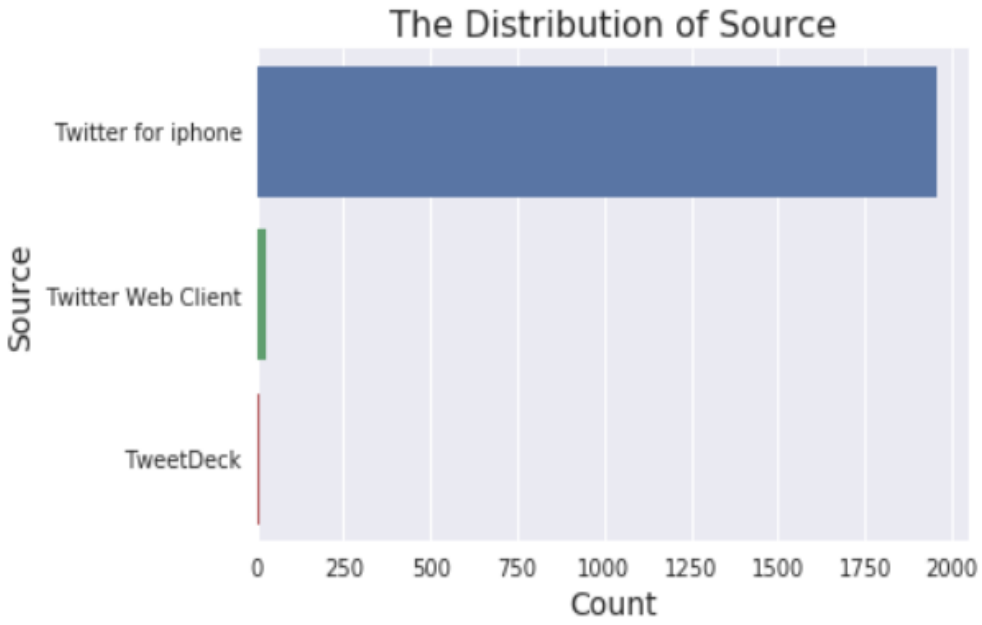
Introduction

My act report includes the analysis that I have done on two different datasets in order to analyze the twitter account. I have provided two insights from the analysis and visualization results.

Data Analysis and Visualizations

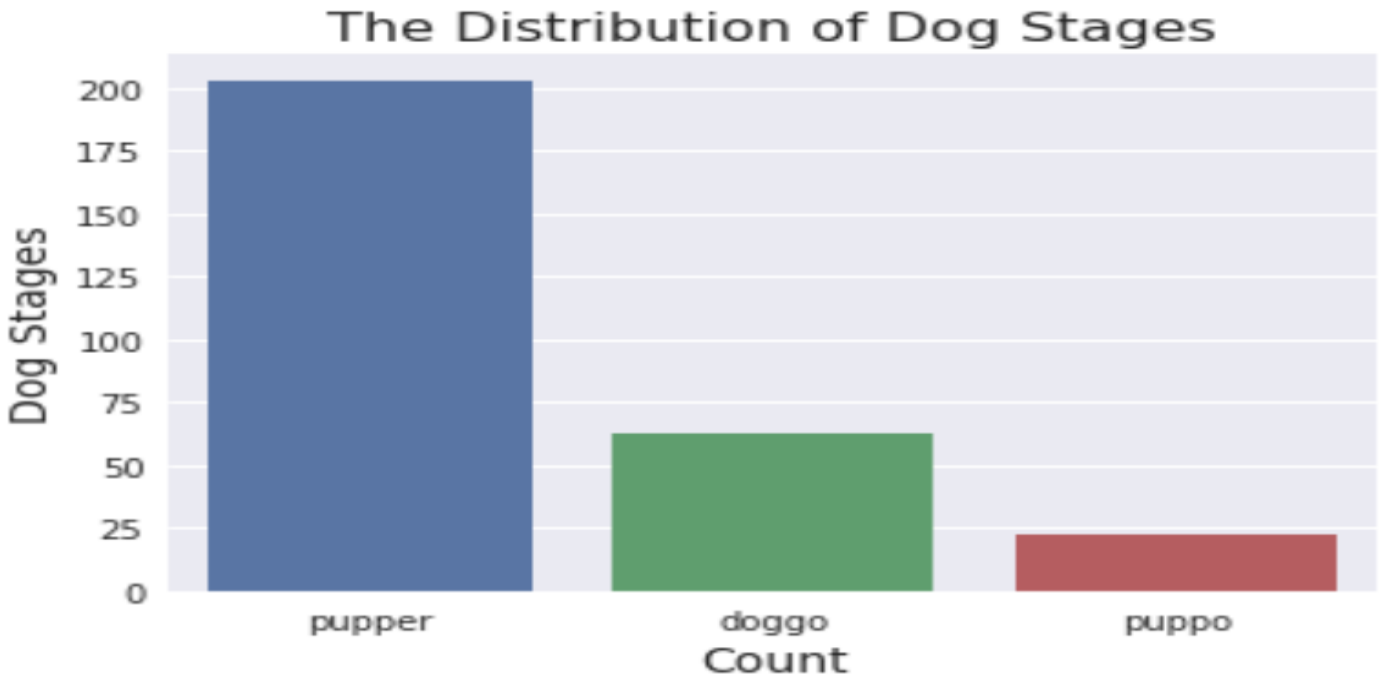
The Distribution of Source

The below plot represents the distribution of source. We can see that the most powerful source of tweets is iPhone twitter app, which dominates 94% in the total, which means twitter app for iPhone is the main channel for people to tweet, retweet, and post. In the other hand, the TweetDesk is rarely used with (less than 1%).



The Distribution of Dog Stages

Similarly, I created one stage that includes all the different kinds of dogs. Then, I check the distribution that stages. It shows that ‘pupper’ (younger) is the most popular followed by ‘doggo’ and ‘puppo’. It seems to me that is due to the cuteness of the young. Since there is a huge amount missing data in dog stages, the distribution may not demonstrate the reality.



The Classification of Dogs Results Analysis

The df_predictions table stores the result of a classification of dog breeds through a neural network. I analyzed the results to see how that works. The breeds below are the top 10 dog breeds have been predicted by this model. You can see that Golden retriever and Labrador retriever are top two predictions. I can conclude by saying those two are most common breeds in U.S based on the results.

```
df_predections['first_prediction'].value_counts().head(10)
```

```
Golden_retriever      150
Labrador_retriever    100
Pembroke               89
Chihuahua              83
Pug                    57
Chow                   44
Samoyed                43
Toy_poodle             39
Pomeranian             38
Malamute               30
Name: first_prediction, dtype: int64
```

Refrences:

https://twitter.com/dog_rates?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor