

By Nyabenge Sylvester.

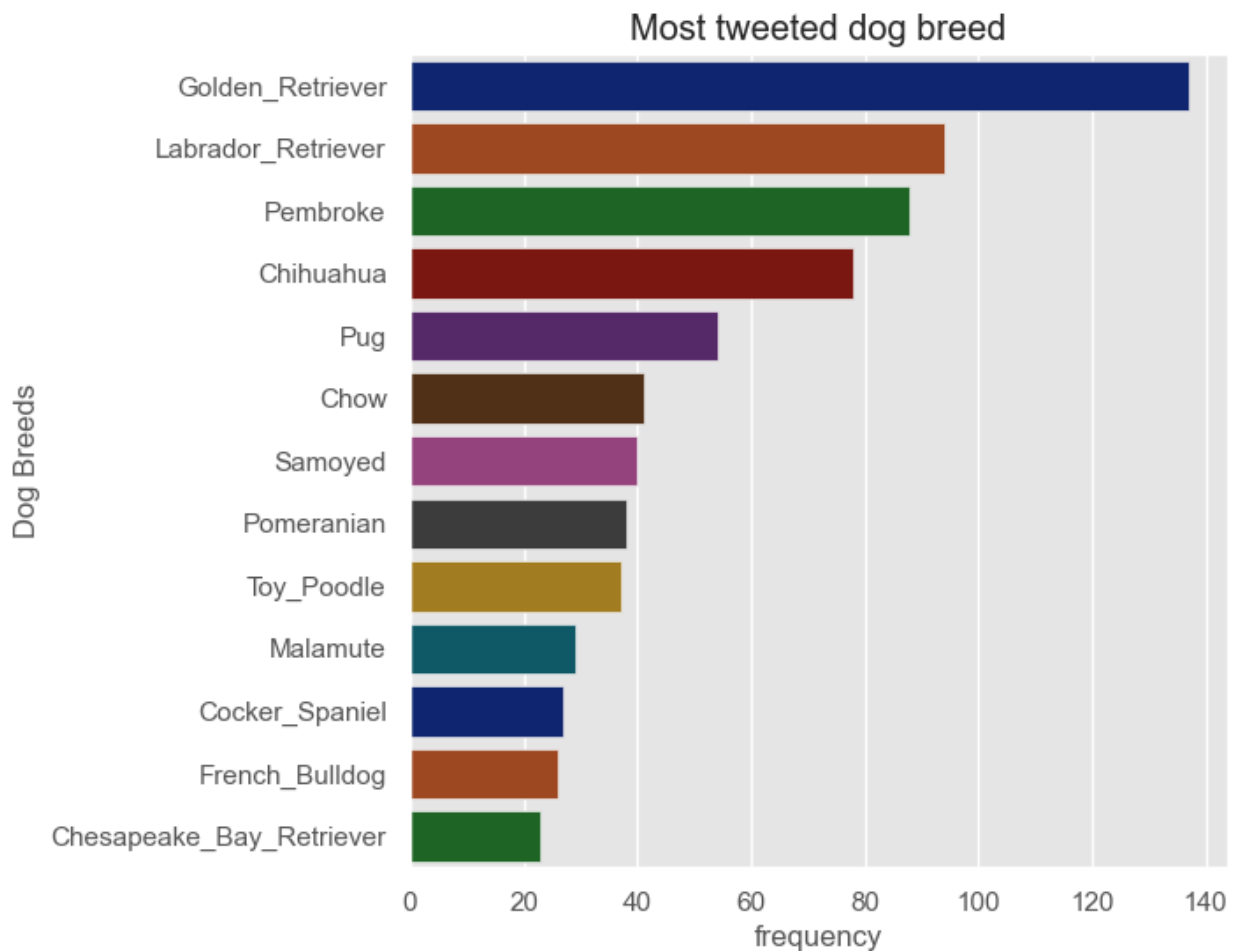
This is a report that communicates insights and visualization produced from data wrangling. The project was part of the requirement of the data wrangling section of the Udacity Data Analyst Nanodegree program.

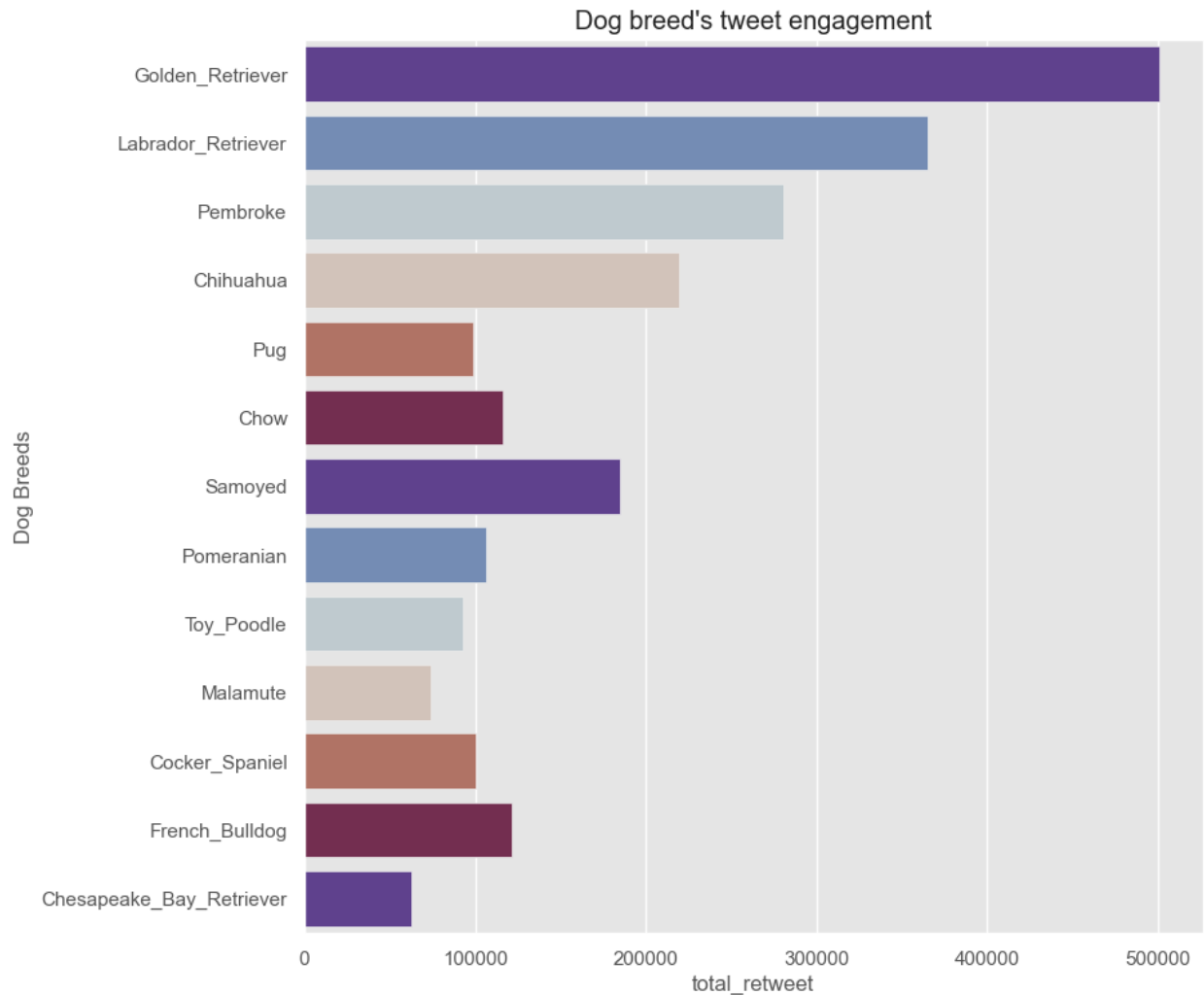
1. The most prominent Dog breed.

Using the `value_counts()` function, we calculate the frequency of various breeds.

The Golden Retriever was the most in number. A dataframe with 3 columns: breed, frequency and total retweets was used to store this data for visualization.

The following visualization was generated:

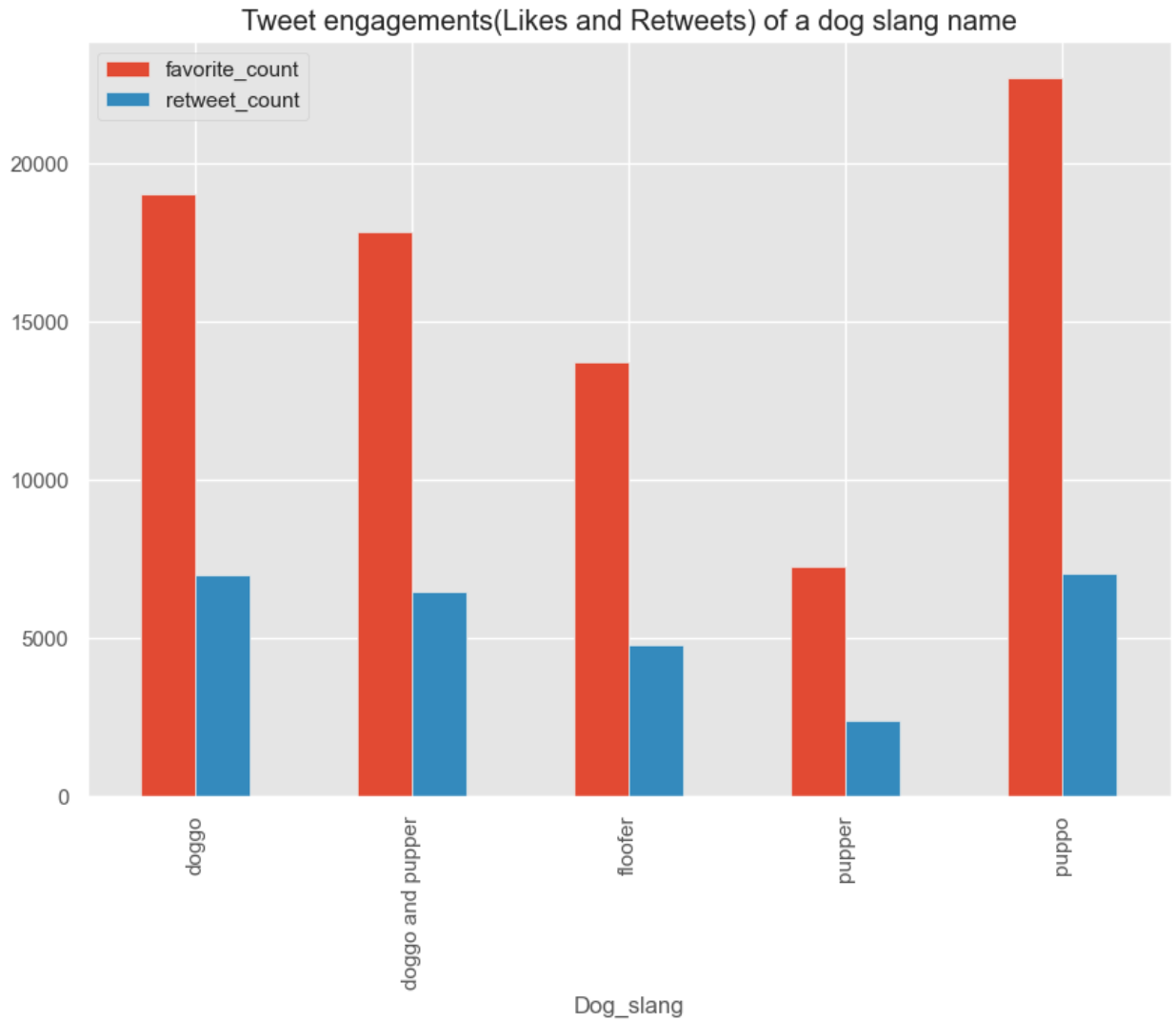




The Golden Retriever is the most prominent breed.

2. The most prominent dog slang/stages.

This was best achieved through visualization. The `groupby()` function was used to group dog slang by retweets and favorite counts. Through this we form the basis of our visualization. The following visualization was achieved:

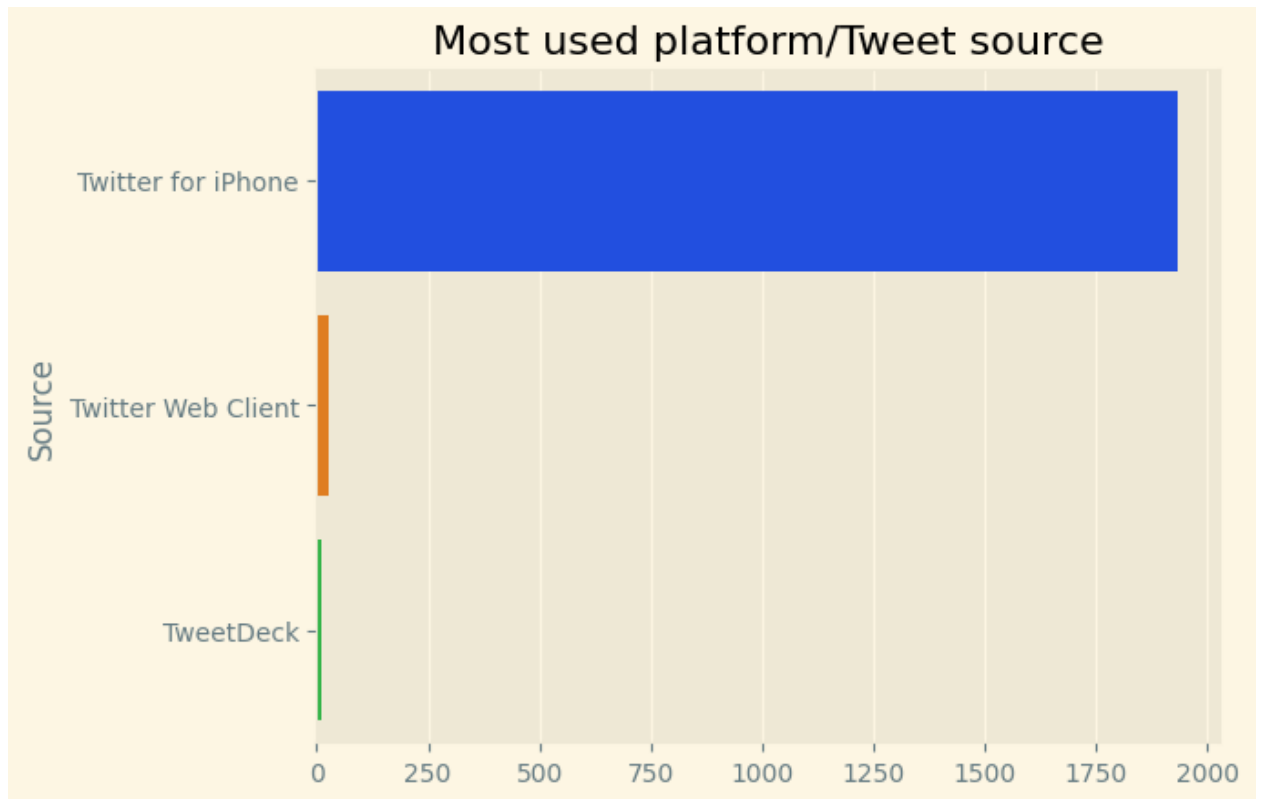


The conclusion made here is that any tweet with “puppo” generated a lot of engagements. On the other hand, “pupper” was least popular and brought less engagements.

3. The Most prominent source

Here we looked at the most popular platform used to tweet.

The `value_counts()` function was used to get the most popular source by numbers. This was also used to create the foundation for analysis and the following was generated:

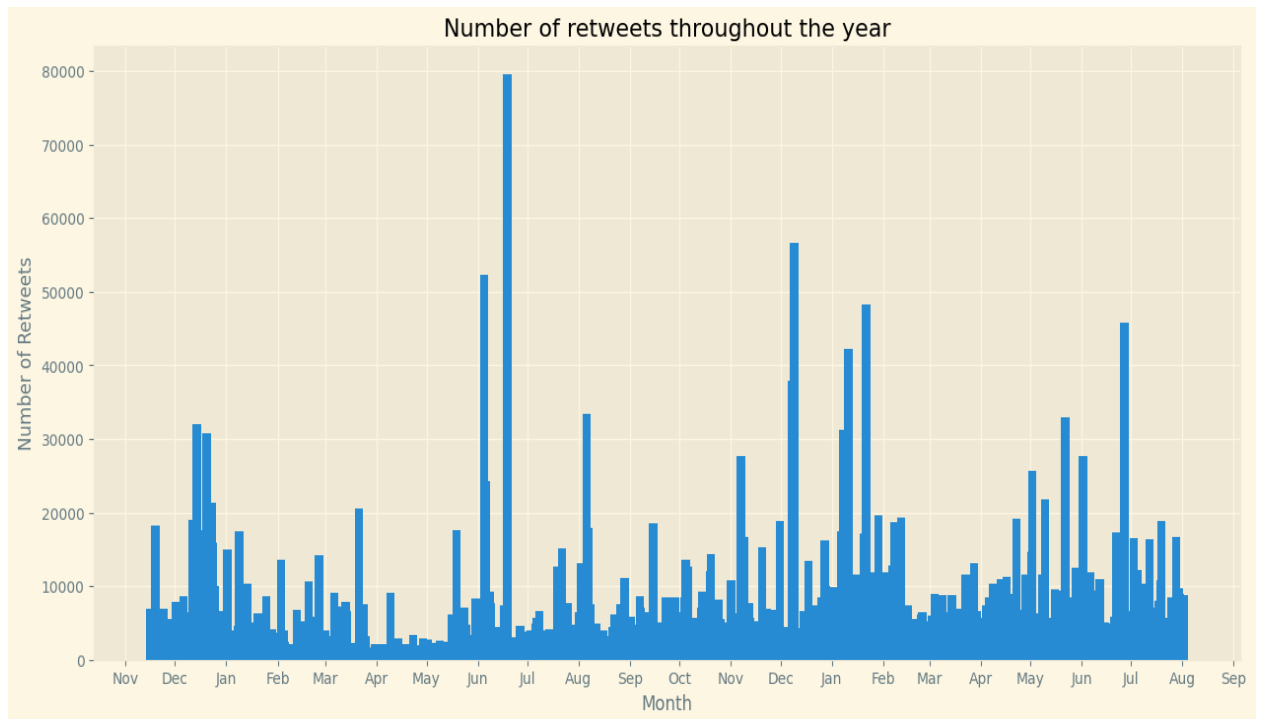


The "Twitter for iPhone was by far the most popular source. The iPhone was mostly used to tweet and retweet.

4. Month with the highest tweet interactions.

This was best displayed using visualization. The mdates library from Matplotlibs helped generate months in our visualization that was plotted against retweet counts with the help of timestamp.

The results was the visualization:



The month of June had the most retweets.