Data Analyst Nanodegree Program

Wrangle and Analyze Data Project

(WeRateDogs Twitter account)

Name: Abdulrahman Alghaligah

September 2021

Introduction:

After wrangling the data, it is time to do the analysis, I started the analysis by creating master dataframe that merge all the three dataframes, it has 1986 rows and 22 columns. This dataframe is ready to help us and start asking questions and answer them using visualization.

Analysis and visualization:

In this section, I have asked three question and I explored them with visualization using matplotlib library.

## Question1: Which of the dog breeds gets more retweets?
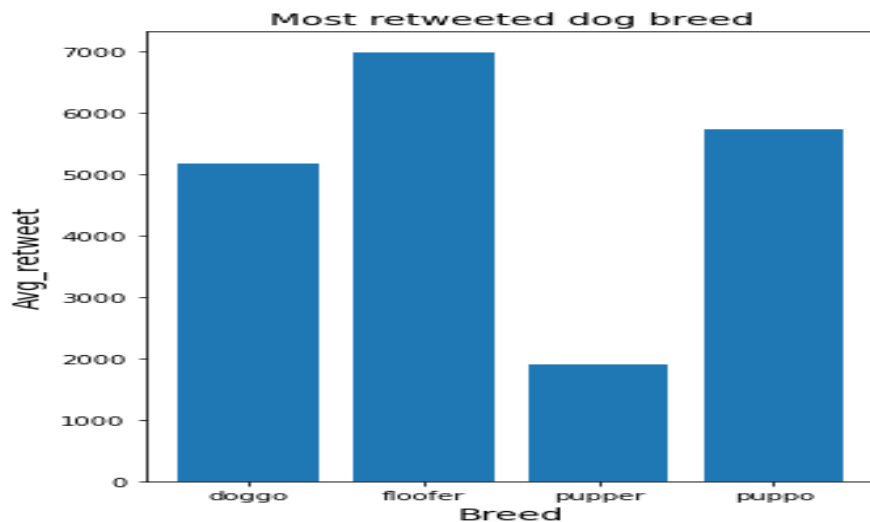
The above graph shows that floofer breed is the most breeds that get a retweet, it gets around 7000 retweets per tweet, after that puppo and doggo breeds respectively and lastly the pupper breed.

Question2: Which source is the most used in our dataset ?
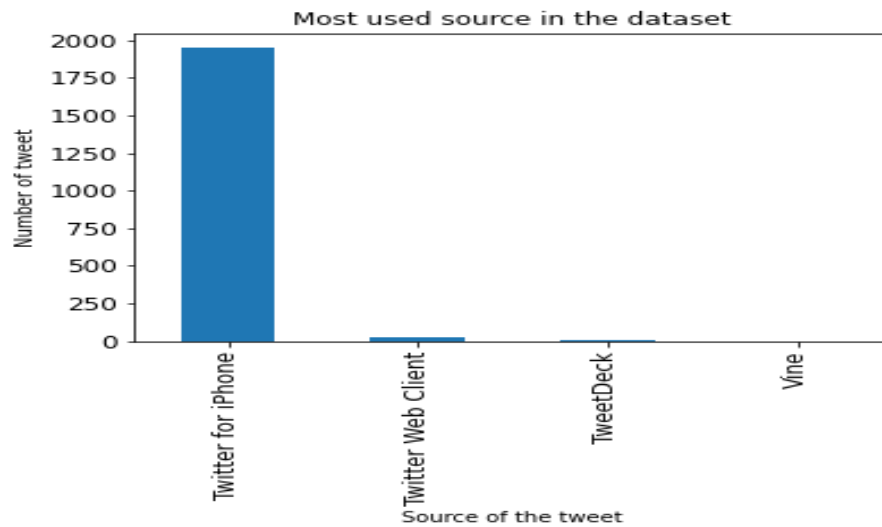
It seems that Twitter for iPhone is the most used source in our dataset tweets with a huge gap between it and other sources as shown in the graph. And it is remarkable to note that Vine source has never been used.

Question3: What is the most common dog name ?
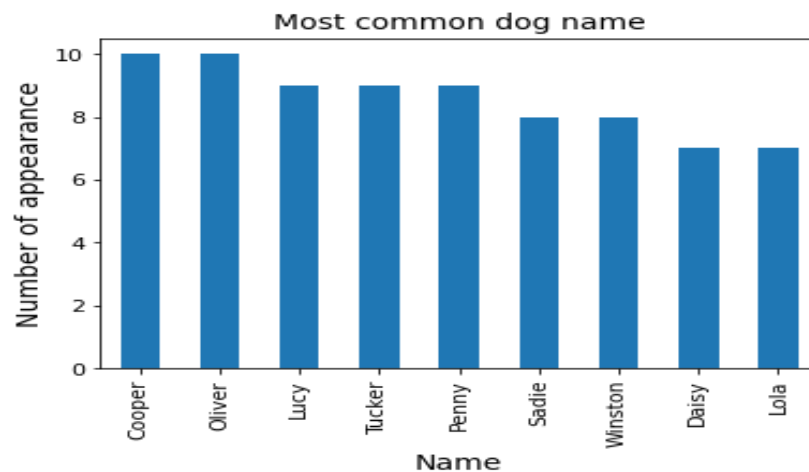
Cooper and Oliver are the most common names in our dataset, after that, we can see names like Lucy, Penny and Tucker. There is no big difference in the appearance of each name, and they are all appeared more than 5 times in the data.

Conclusion:

After wrangling the data, it is clear that the data is ready and easy to be explored and help to ask questions and answer them about WeRateDogs Twitter Account, also it is important to note that I have cleaned the data in a way that help me answer these questions, it is totally ok to reassess and clear the data again in future if needed.