

Act_Report

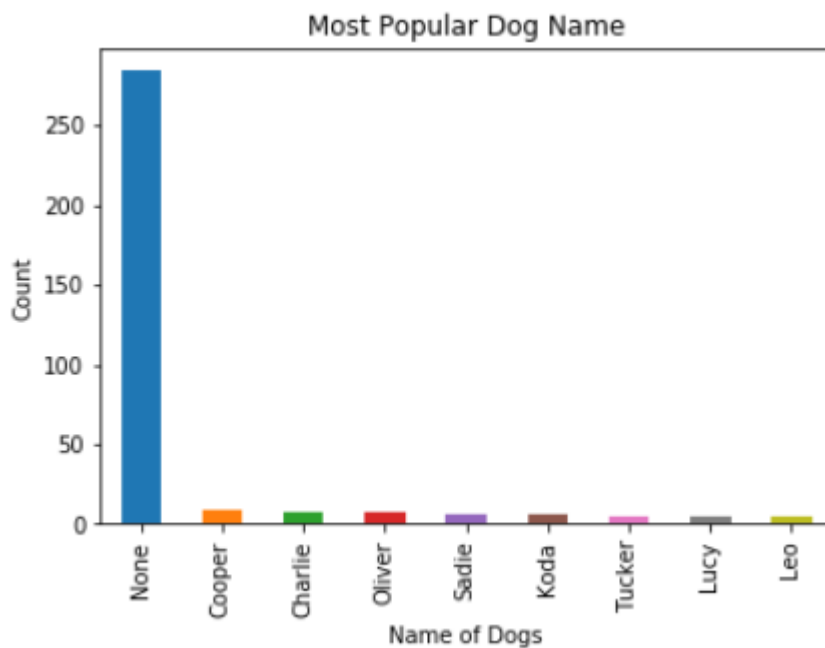
Introduction

This report briefly describes the various insights gotten from WeRateDogs data wrangling. 4 insights and 2 visualizations were reported. After the whole data wrangling process, I went ahead to get some insights and carry out some visualizations on the final dataset. There were a few observations ranging from the most popular dog name to the most used twitter source, and so on. The final dataset contained dog names, source, tweet id, favorite counts, retweet counts, different dog predictions, etc.

The insights and visualizations from the dataset are as follows:

- 1. Most dogs do not have names (represented as "None"), but they have ratings and other information. So "None" is the most popular dog name.**

The first insight was to check for the most popular dog name. It was seen that most dogs do not have names, making “None” the highest value in the name column.



- 2. There are many dogs that do not belong to any dog category. These dogs are represented as "None".**

The second insight was to find out how many dogs are in different classes. A plot was created for the different dog classes. Visualization showed that over 900 dogs did not belong to any class. This was represented as “None”. It is either these dogs were not classified, or the data is missing. The next most popular category was the “pupper” category.

- 3. The most used twitter source is "Twitter for Iphone"**

There were 4 categories of sources namely; twitter for iPhone, Vine-Make a scene, twitter web client, and tweetdeck. It was seen that “Twitter for iPhone” was the most used twitter

source with over 1000 values. It looked like many users made use of an iPhone as at that time and less people made use of TweetDeck.

4. There is a positive correlation of 0.92 between favorite counts and retweet counts.

A scatter plot was used to carry out the visualization and it was observed that there was a positive correlation of 0.92 between both variables.

