# Report

on

We Rate Dogs Data

Carried out as part of

UDACITY PROJECT

This documentation focus on the analysis and insight of the final data. There are several analysis which I have carried out in order to answer three different questions.. They are as follows:-

1. WHAT IS THE MOST COMMON DOG NAMES?

The data column containing the names of dogs has a lot of invalid names which includes single alphabets and in some cases none. These issues were sorted out by setting those invalid names to none. It is observed that more than 60 percent of the name are categorized as none. From the analysis carried out so far with the data at hand, the first eight most common names to dogs: Charlie, Cooper, Lucy, Oliver, Tucker, Penny, Sadie. Excluding the none part of the names.

Below is a chart that shows this
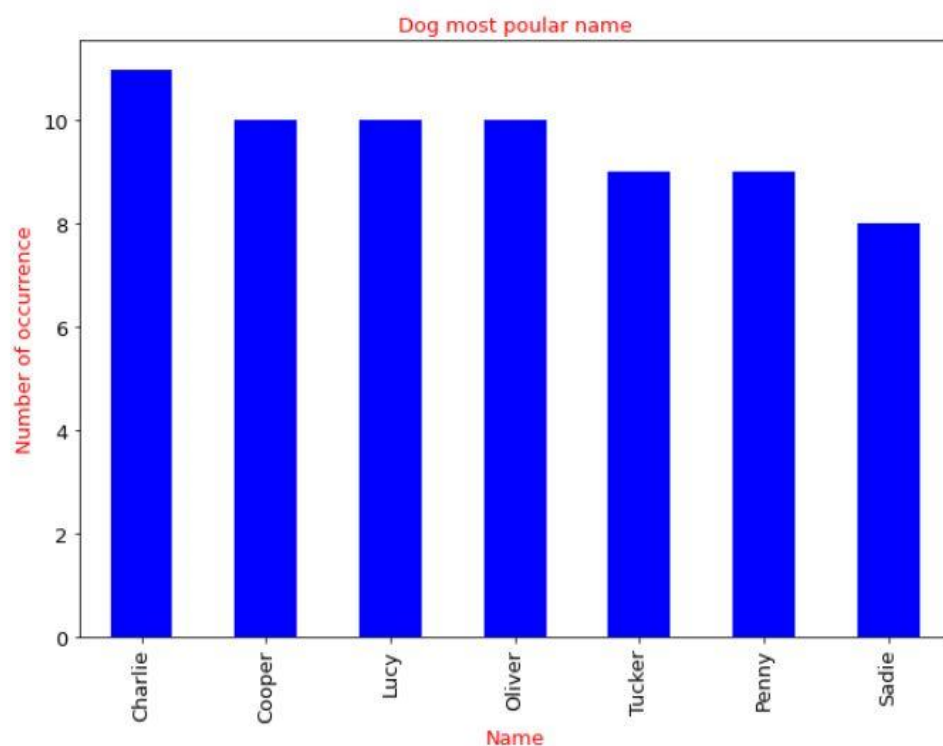
2. WHAT IS THE CORRELATION BETWEEN FAVOURITE COUNT AND RETWEET COUNT?

A correlation help find the statistical measure of the relationship between two variables. Using a scatterplot, we were able to assess the relationship between the two variable been considered. The favourite count and the retweet count.

The image below shows the correlation between the favourite counts and how many times a post was retweeted. From this scattered chat, we see a positive correlation between this two variables. The variables tend to move in the same direction that is when one variable increases, the other variable also increase.
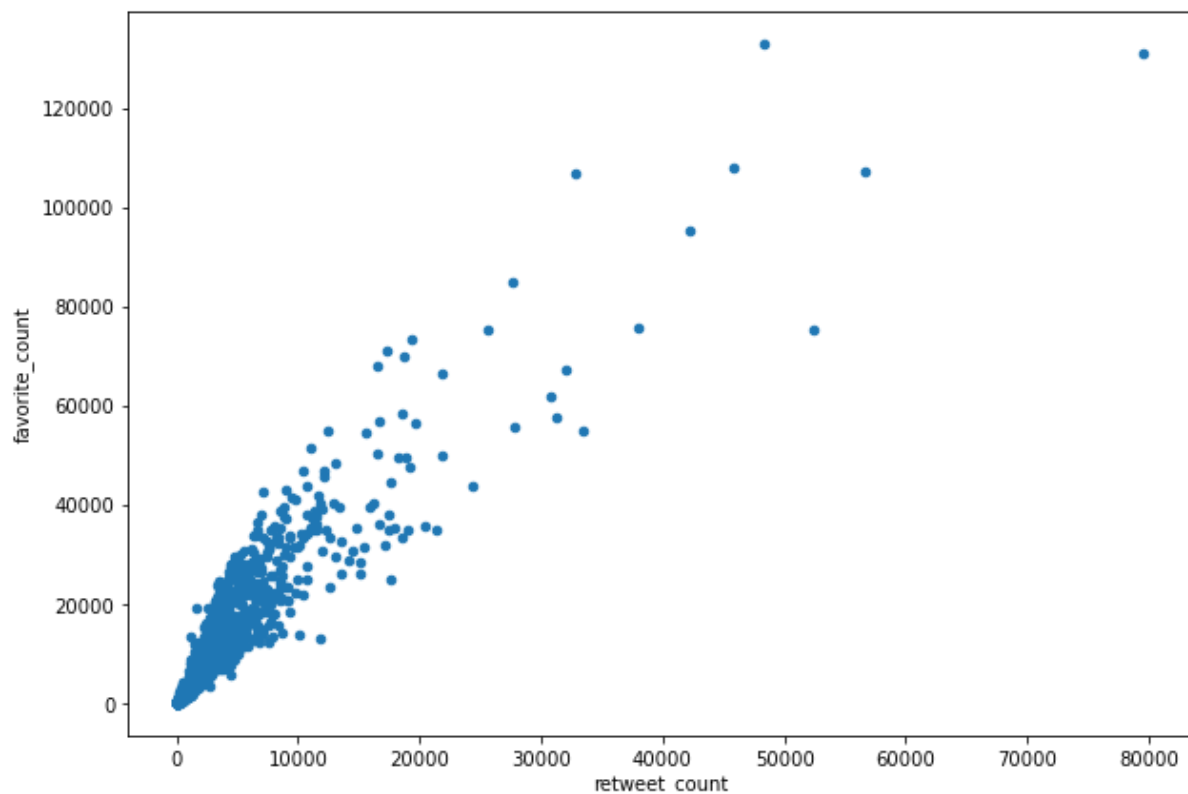
3.WHAT IS THE MOST USE SOURCE FOR TWEETS

Another analysis done was to determine the most used source for tweeting. This was possible after carrying out a cleaning process on the source data by removing the excess text and link attached together. From the analysis carried out, it was discovered the over 98 percent of the tweets have their source from an Iphone while the other 2 percent was shared among the twitter web clients and the TweetDeck users.

Most Used Twitter Source

Twitter for iPhone — 98.0%

TweetDeck — 0.6%
Twitter Web Client — 1.4%