Date: July 24, 2022

Analysis and visualization report

Overview: The data in the WeRateDogs tweets dataset consists of names, ratings of dogs, timestamp etc. as well as favorite counts and retweet counts of each of the tweets. As with any data set given, we want to gain meaningful insights by asking questions based on the data. For instance, we could ask questions like "*What are the popular names for dogs?*" or "*What is the average dog rating?*" etc. Analysis of the data set produced the following insights:

- The average dog rating was 12.08
- The most popular dog names found in the dataset were Null, None and Bo.
- The higher the rating, the higher the favorite counts.
- Most popular source of posting is the phone

Table 1: Descriptive statistics of dog ratings, favorite counts and retweet counts

	Rating	retweet_count	favorite_count
count	**180.000000**	**2354.000000**	**2354.000000**
mean	**12.077778**	**3164.797366**	**8080.968564**
std	**4.984283**	**5284.770364**	**11814.771334**
min	**1.000000**	**0.000000**	**0.000000**
25%	**11.000000**	**624.500000**	**1415.000000**
50%	**12.000000**	**1473.500000**	**3603.500000**
75%	**13.000000**	**3652.000000**	**10122.250000**
max	**75.000000**	**79515.000000**	**132810.000000**

a. The denominator is omitted since it's a constant value (10) and the numerator was renamed as the rating.

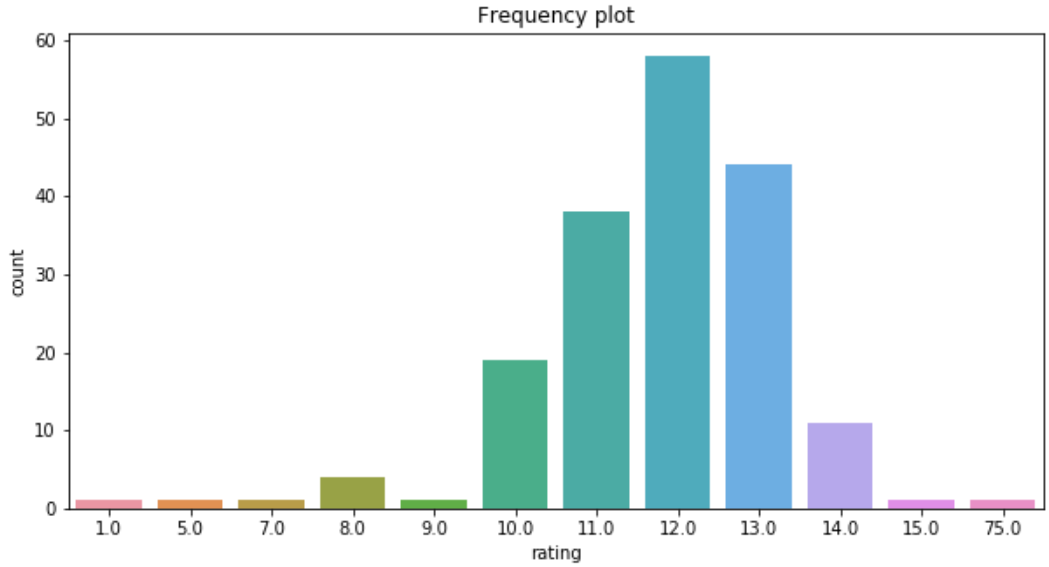

Figure 1: The frequency plot of ratings.

The relationship between dog rating and favorite counts

We investigate the relationship between the retweet count and favorite counts (figure 2). It appeared that favorite count increased with the rating although the trend seemed like a curved one. This could mean that the more favorite counts a tweet gets the higher its chances of getting retweeted and vice versa.

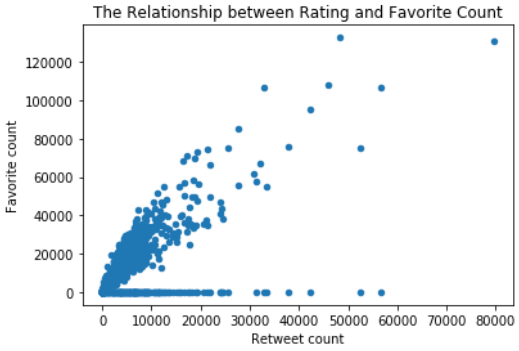

Figure 2: The relationship between the retweet counts and favorite counts

Frequency of various sources

It seems thar when it comes to posting on twitter the iPhone is the most popular choice as shown in the graph below.

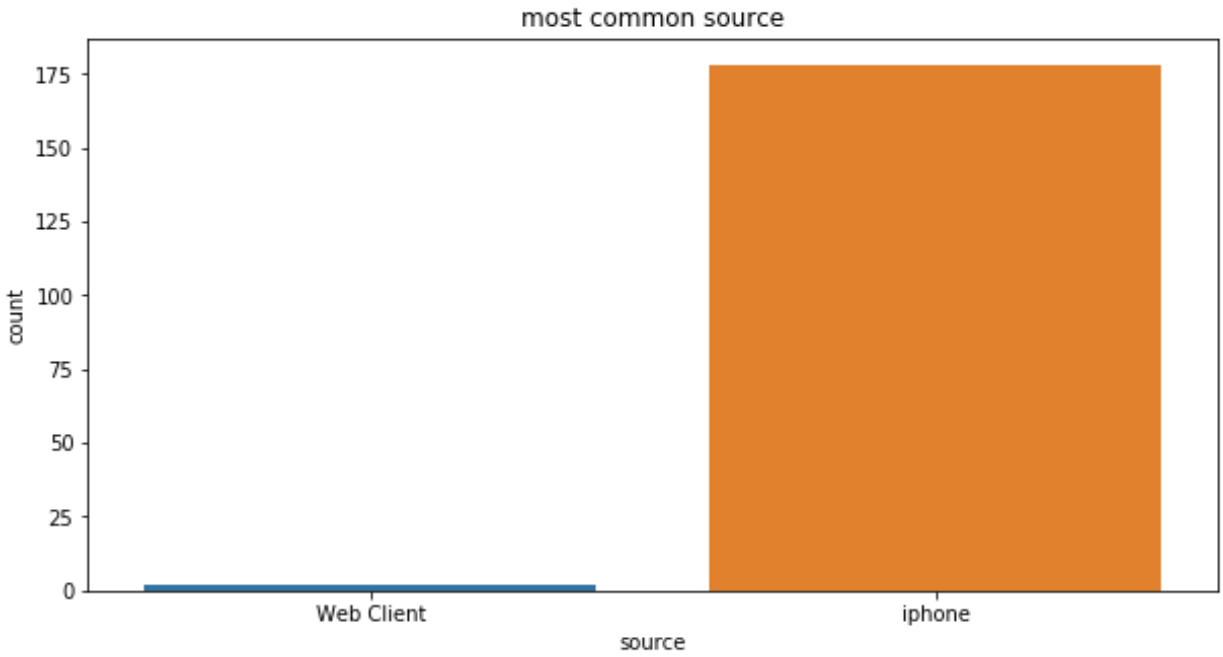

Figure 3: The frequency of the various sources.