WeRateDogs Data Analysis ACT Report

The aim of this report is to communicate the insights and display the visualisations produced from the wrangled date of the WeRateDogs data analysis project on Udacity.

After the wrangling of the data, the cleaned data was plotted on graphs using different methods to get different insights into the data.

The following methods were used for plotting the data:

- Histograms: These were used to show the numerical distribution of the different columns in the dataframe.
- Scatter plots: This was used to show the correlation between two (2) different attributes or properties of the dataframe.
- Pie charts: This was used to show the pre-dominance of the sources of the images.

HISTOGRAM OF THE DATAFRAME:

Using the hist() method of matplotlib in Python, a histogram of all the numeric columns of the dataframe was plotted. Then a few histograms were selected to answer a few questions. For instance, one question of interest was to know which rating occurred more frequently for the dogs.
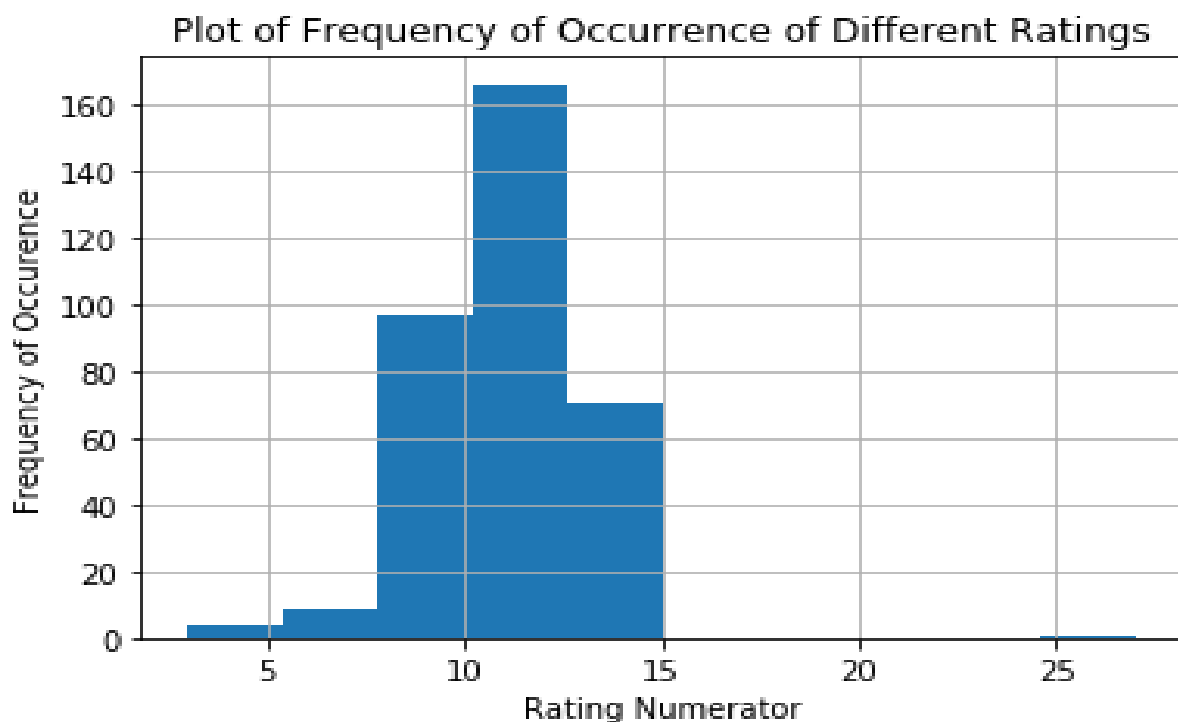
From the plot above, we can see that most of the ratings (which were mostly over 10) fell between 11 and 13. There a few outliers that were around the 20 range. However, considering its relatively low rate of frequency, it can be ignored. Below is that of the distribution of the denominator.

Plot of Frequency of Occurrence of Different Rating Denominators
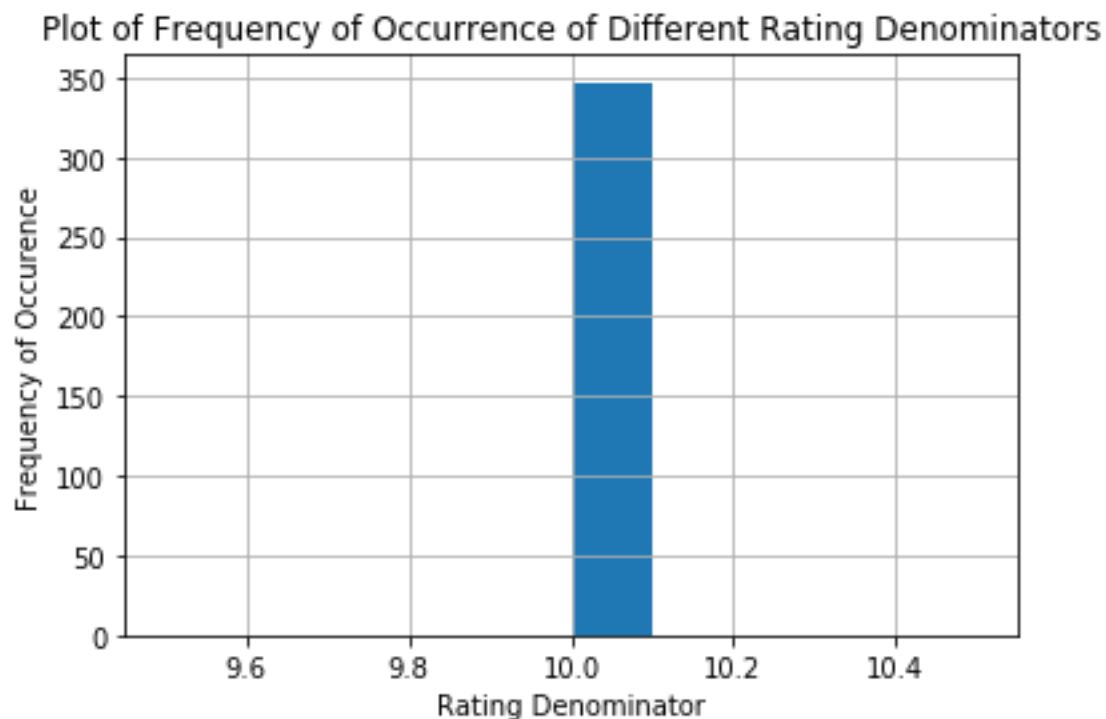
This plot clearly shows that most of the denominator rating is 10, as actually expected from the visual assessment of the data at the beginning of the data wrangling exercise.

**SCATTER PLOT OF RETWEET COUNT VERSUS FAVORITE COUNT**

A scatter plot to show the correlation between the favorite counts of the image tweets and the number of retweets of the images was done. The aim was to show if a particular attribute influenced the other. The scatterplot is shown below

Plot of Favorite Count VS Retweet Count
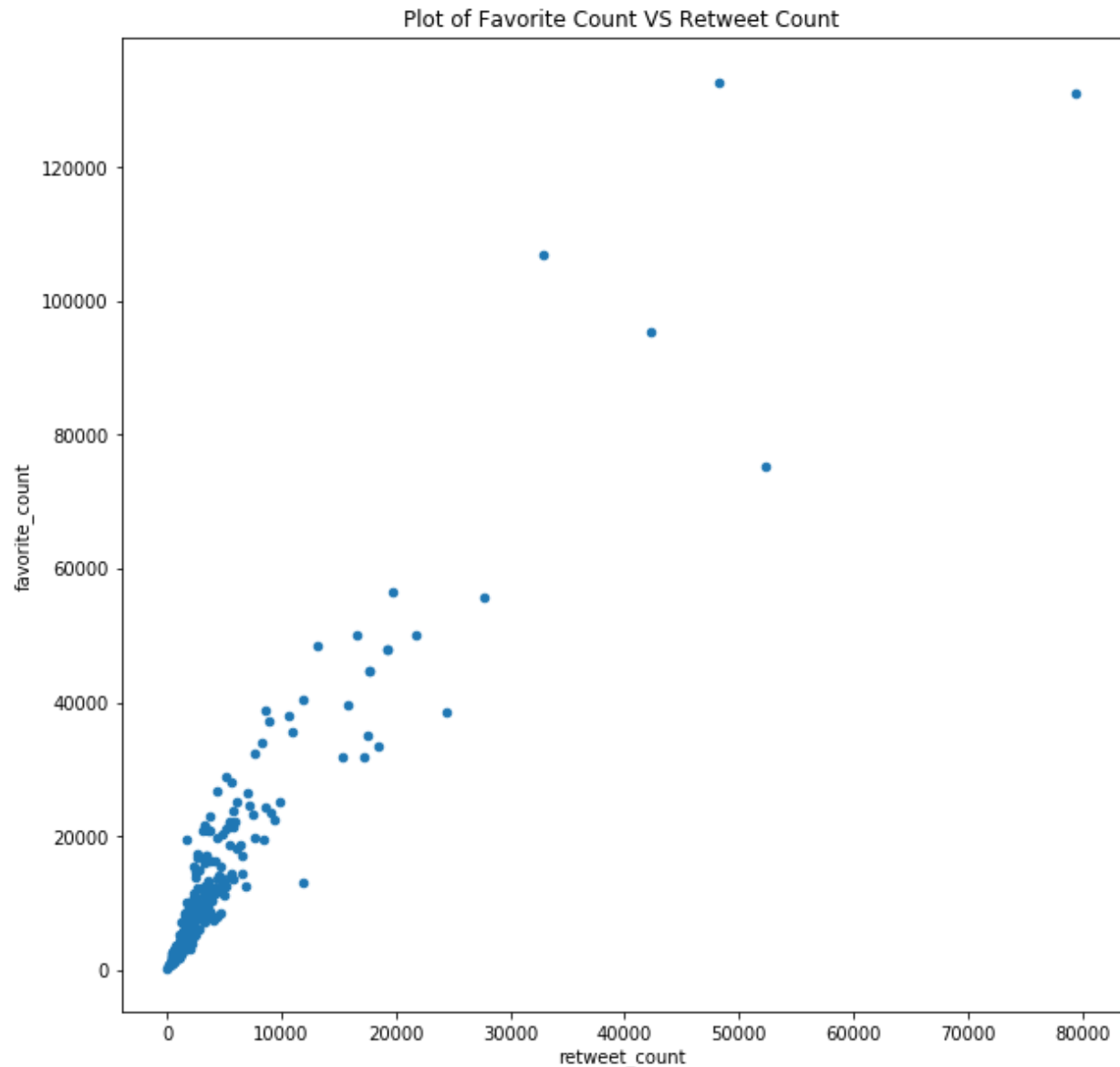
From this plot, it can be shown that there is a positive correlation between the retweet count and the favorite count. We can surmise that the more an image is retweeted, the more it is added to the favorites of the Twitter user.

PIE CHART OF SOURCES OF THE IMAGES:

A pie chart to show the how the images were sourced was also plotted using the plot function. The result is shown below.
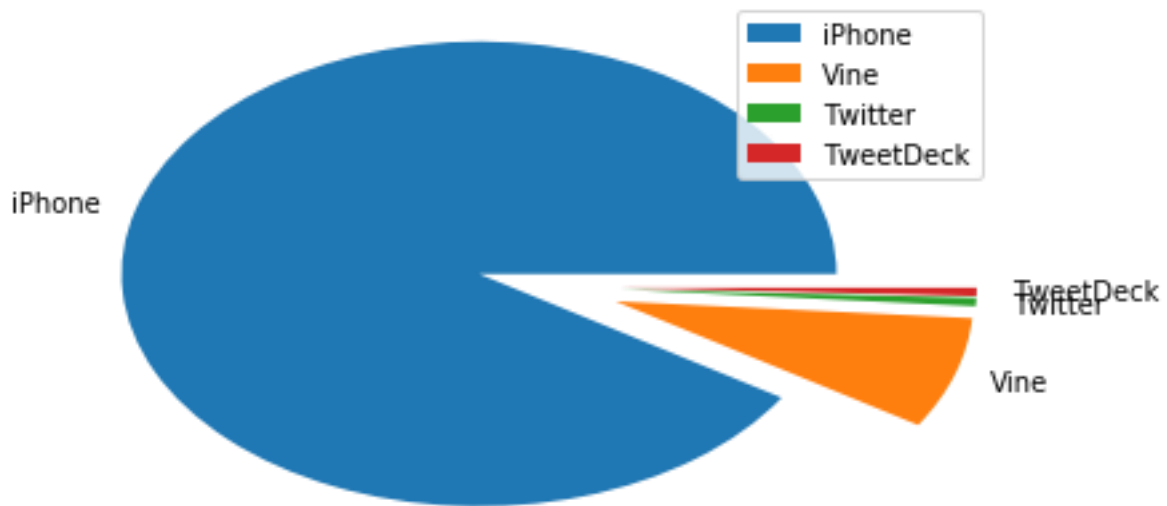
The plot clearly shows that most of the images were sourced from the Twitter app of an Apple iPhone followed by use of the Vine application. If anything, it shows that most of the followers of the WeRateDogs Twitter account are actually iOS users.