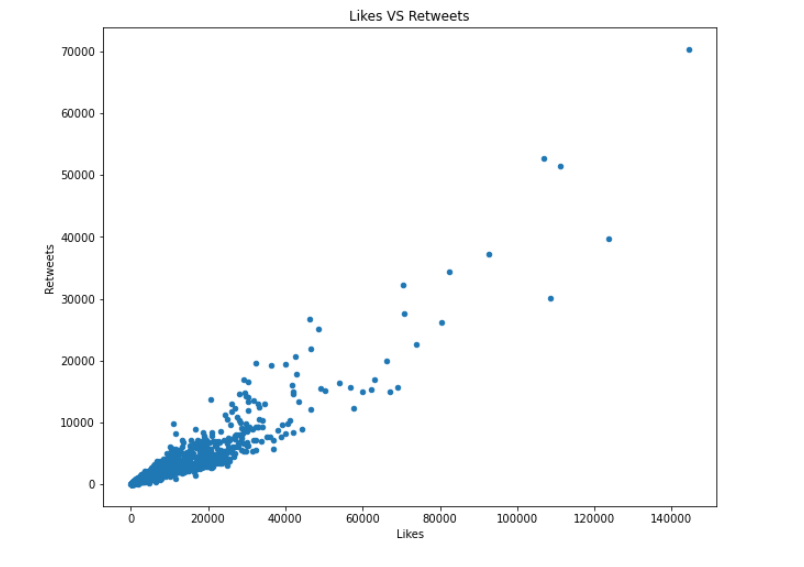
After assessing the data set one after the other, I wrangled and cleaned them programmatically making it ready for analysis.

I started my analysis by first checking the basic info and statistics of the master dataset, from which I moved on to using the pandas describe() function to get a basic statistics of the dataset which revealed some interesting details about the data below:

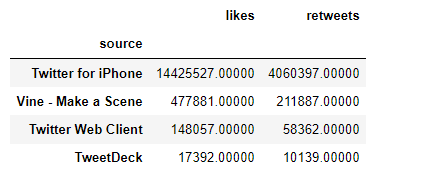
* The tweet with the lowest number of likes was just **45 likes** & the highest **144,343 likes**
* The tweet with the lowest number of retweets has just **1 retweet** & the highest **70,383 retweets**
* The best/most accurate dog prediction using the first prediction **p1\_conf** is **99%** accurate

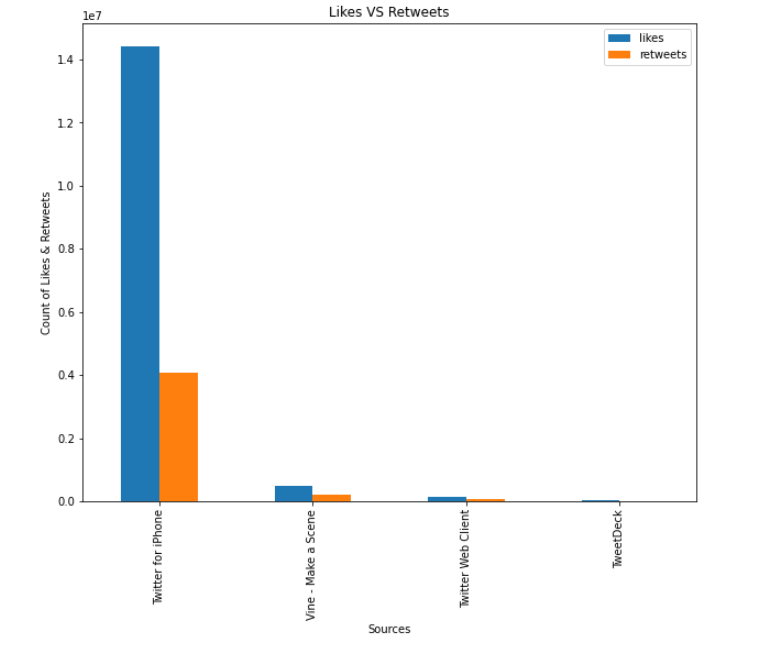
Then I moved on to check if there’s a correlation between the number of likes and retweets in the dataset, and as expected there is a strong positive correlation of 0.92 which means as the likes increase the number of retweets increase also.

But we must be careful not to imply that likes causes retweets to grow as this is not enough proof to go on. A chart of a scatterplot with the correlation below:

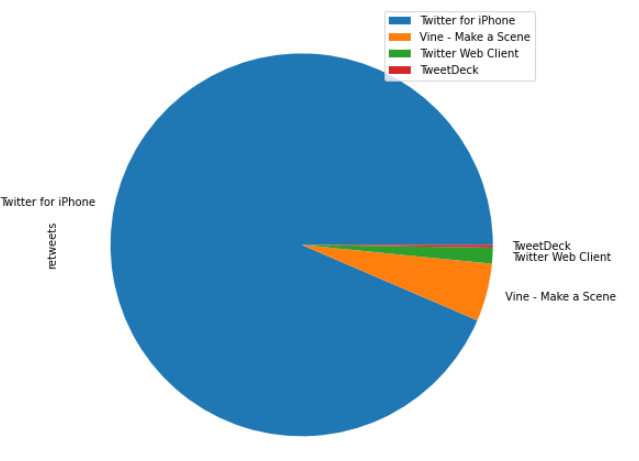
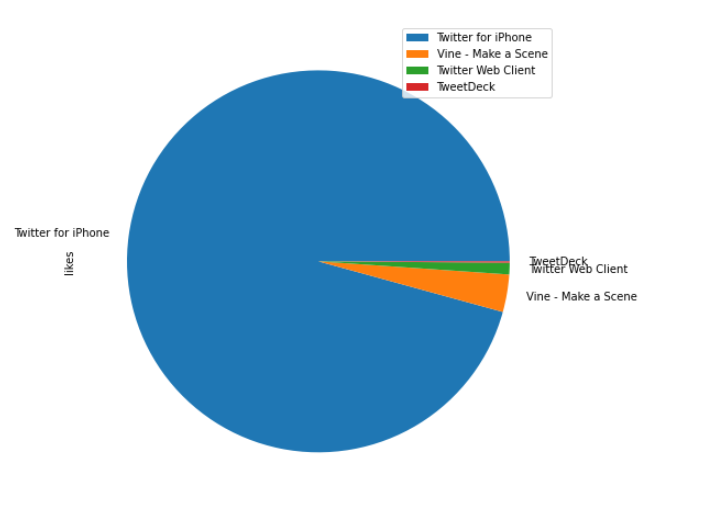


My next analysis was to check the distribution of likes among the various sources of tweets which showed that tweets from iPhone got a greater number of likes and retweets though this is also because most of the tweets were from iPhones. See table and chart below:

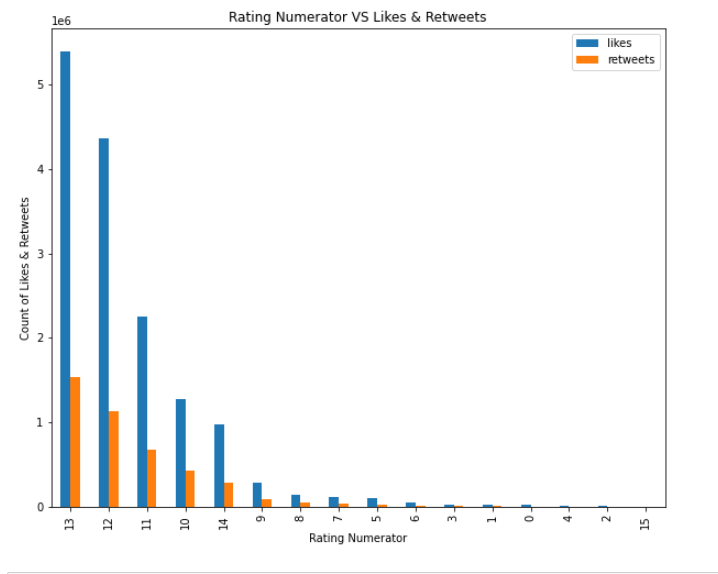
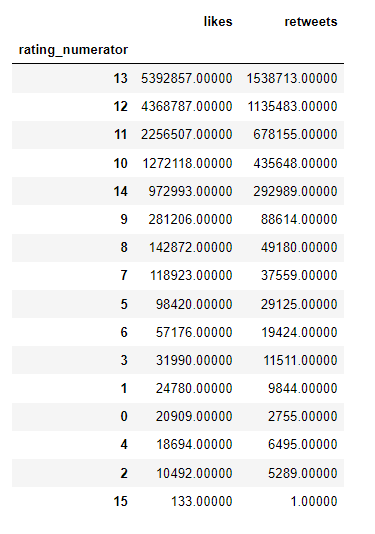




I also used Pie charts:



Then finally my last analysis was checking if the rating\_numerator had any effect on the number of likes and retweets, which the analysis showed that the tweets with dogs with higher ratings had more likes and retweets. The chart and table is below:



* The above results goes to show that the higher the rating the more likes and retweets the tweet gets except for the case of rating(15). We should also note that this does not neccesarily implies causation as further statistical analysis would need to be done.

This goes to complete my analytical process and insights gotten from the dataset.