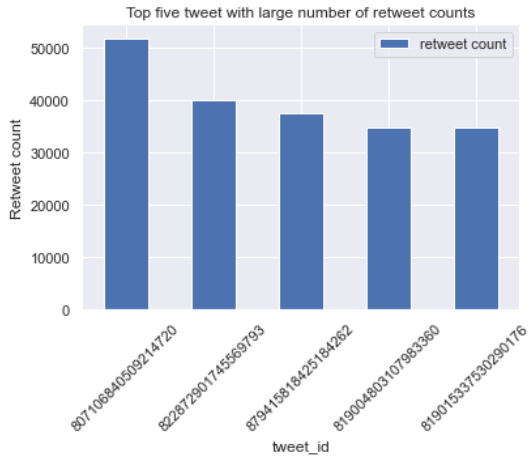
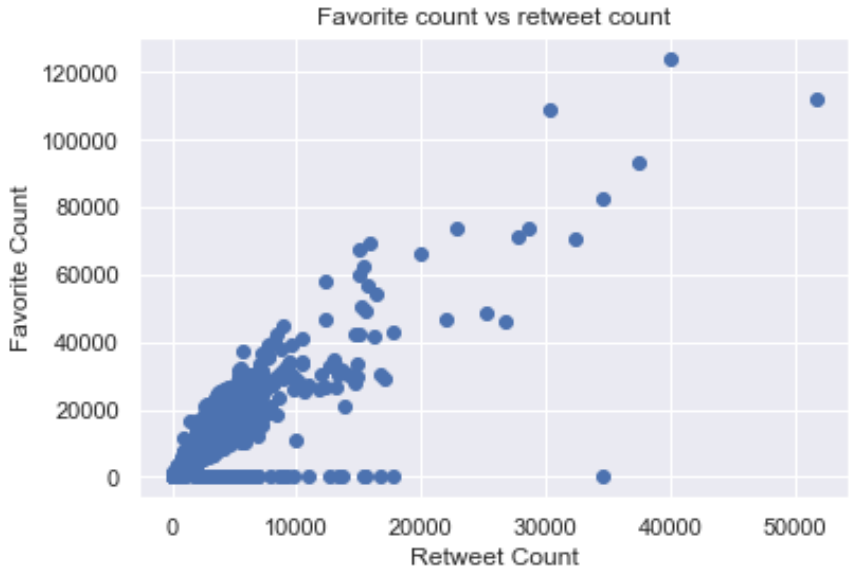
After properly wrangling the datasets, I jumped into the analysis part. In this part, libraries such as matplotlib and seaborn were used to provide some visualizations.

To begin, have the tweet about dogs, one could be curious to know the most retweeted tweets. Using seaborn, a graph showing the most retweeted tweets was provided.

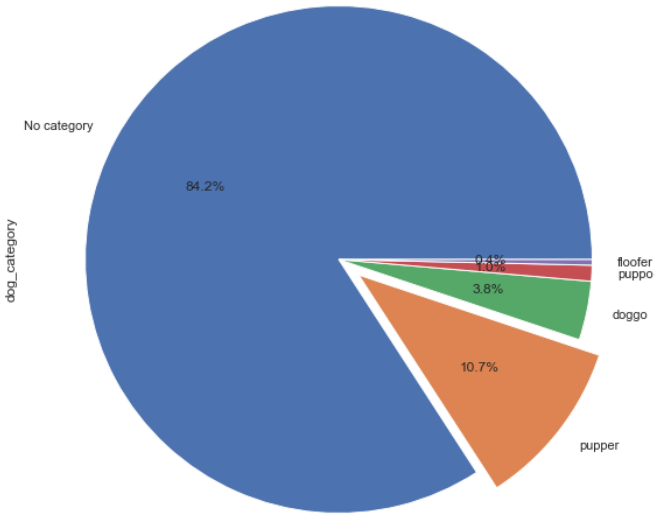


Using their URLs, the image with the most retweet was extracted.

With the retweet and favorite counts, we can see if the most popular tweets are also the most retweeted. For analyzing the relationship between the retweet and favorite counts, a scatterplot was provided.

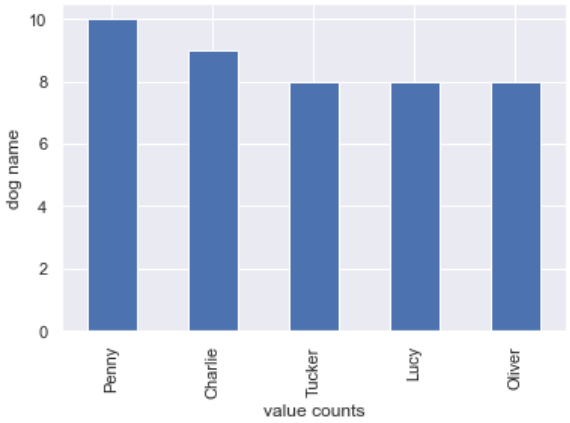


It can be seen in the figure above that there is a positive trend. This means that in most of the tweets as the retweet counts increase, the favorite counts also increase. Therefore, we can say that favorite tweets are those being retweeted.



The figure above depicts the category of the dog. As shown by the figure, we can see that above 80% of the data, we don't have a category for the dog. Among four categories, pupper is the most with 10.7% and floofer is the last with 0.4%.

Having the dog name column, we can see the most popular names of dog.



The figure shows that the penny is the most commonly used name of the dog in our dataset.

Limitations and further work

Poor meaning of some columns mostly for rating columns. With your feedback and deep research, I will try to work on the analysis of ratings and make my codes more modularized.