

Wrangle and Analyze Data

Adel Abu Hashim

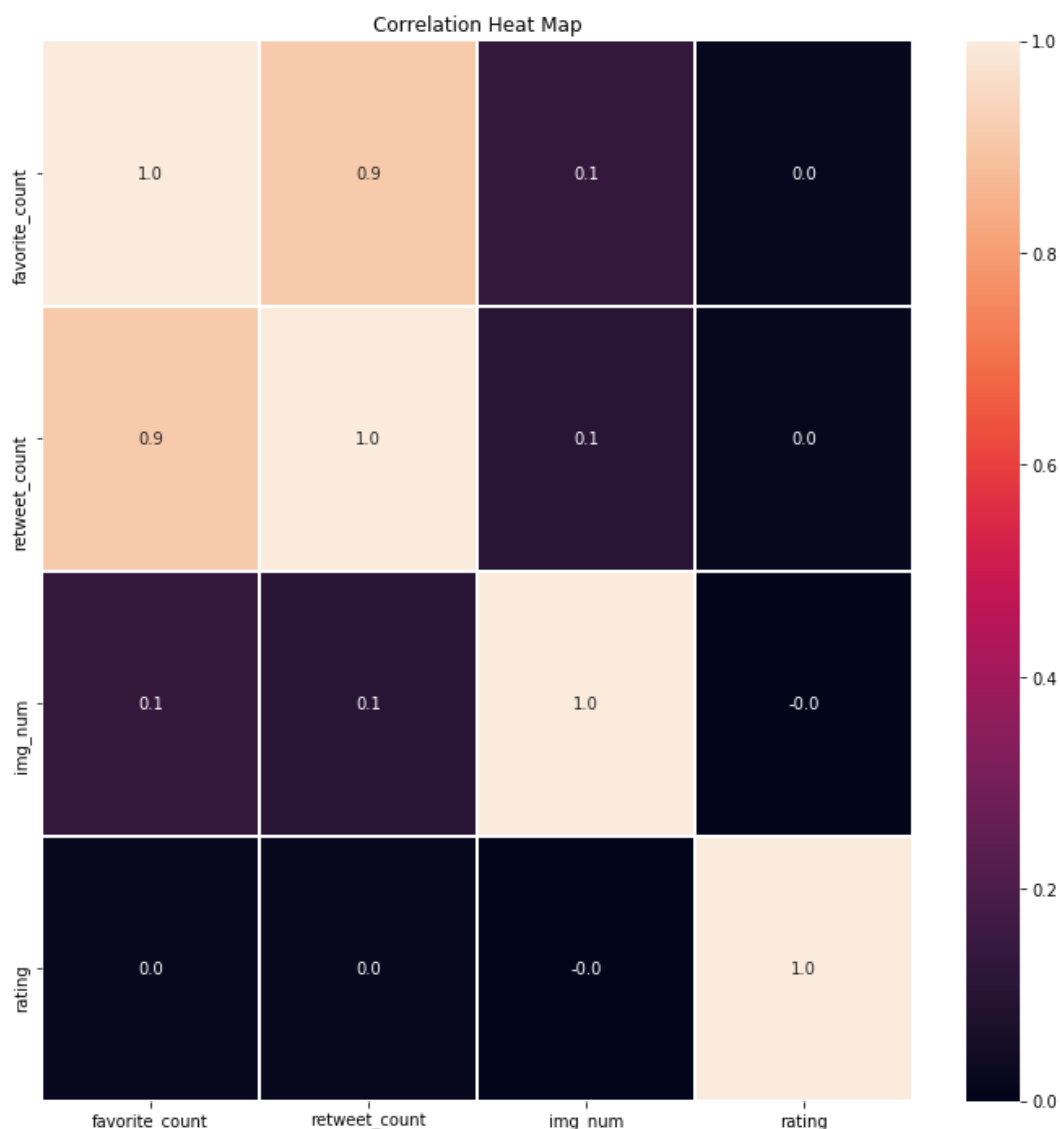
Data Analysis Process:

Here I will visualize and deduce some information after wrangling.

I did some analysis on dogs' facts; like dog names(choosing the highest 10 popular names) also about the favourite dog type by comparing and visualizing counts of the four types.

I also did analysis in rating like searching for the time of the year in which tweets are popular and find the average rate or the most rate on dogs tweets.

You can see Correlation Heat map which states correlation between some columns:

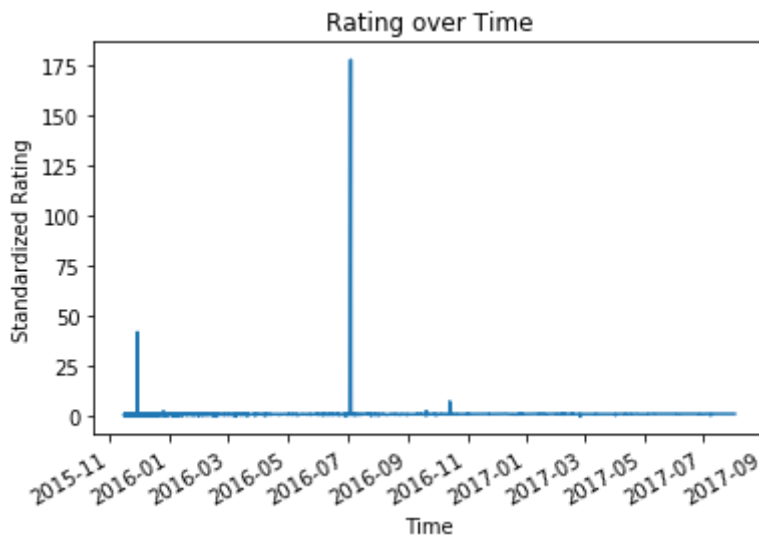


After this Conclusion we can see:

The strong correlation between favorite and retweet; it seems that it was an expected conclusion but even normal things must be analyzed.

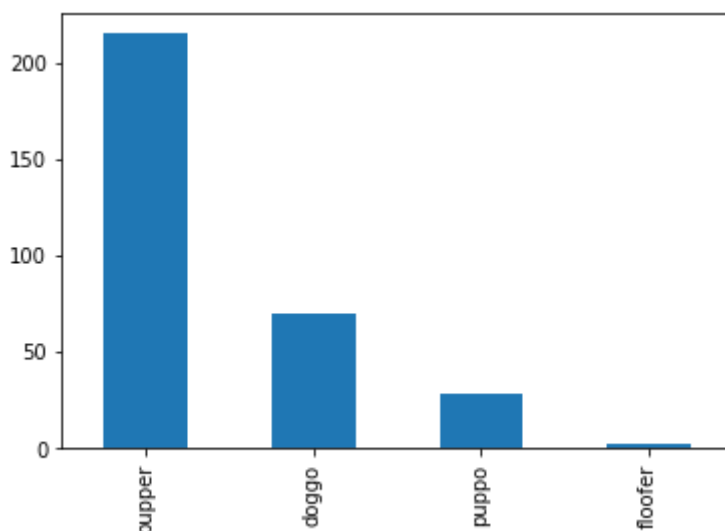
My analysis is about answering some questions:

Q1: Which tweets are favorite over time?



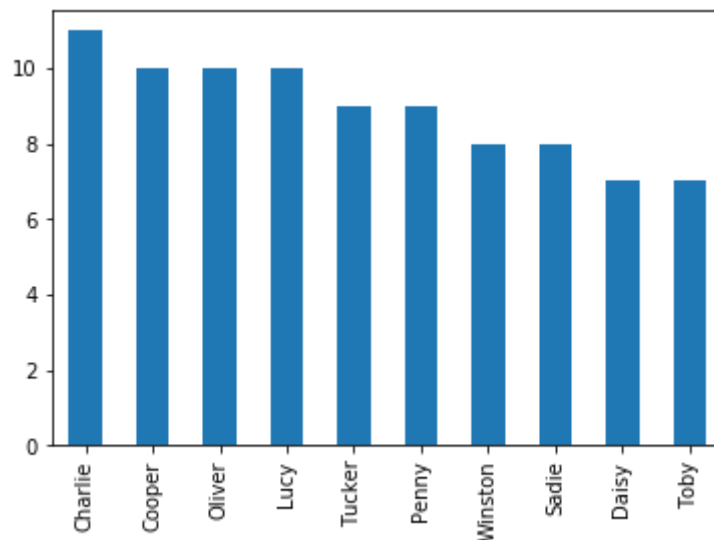
As we can see in the visualization Tweets in july 2016 contains high rates

Q2: Which stage of dogs are used most on tweets?



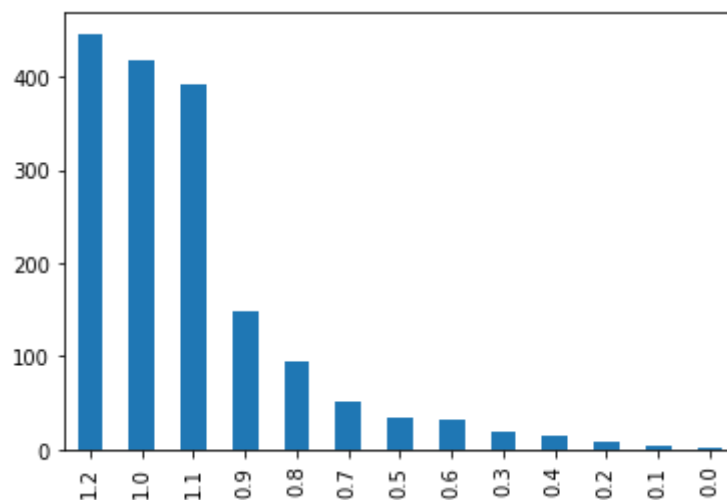
Pupper dog type is the favourite stage, we can see that it is more favorite than the sum of the rest three stages!

Q3: What is the most popular dog name?



Charlie. After deleting extraneous names like; 'a', 'the', 'nan', 'not', 'one', 'O', 'all', 'his', 'by', 'life' we deduced that charlie is the most popular dog name.

Q4: What is the most common rating?



the most common rating is **1.2**; we go this after filtering our data by removing outliers

Conclusion

I can say that the twitter handle of WeRateDogs gives a lot of interesting insights about how the sentiments of the people work when it comes to different breeds of dogs within their different life stages. It also tells us that having more followers on twitter is not directly proportional to more retweets of a users tweet.