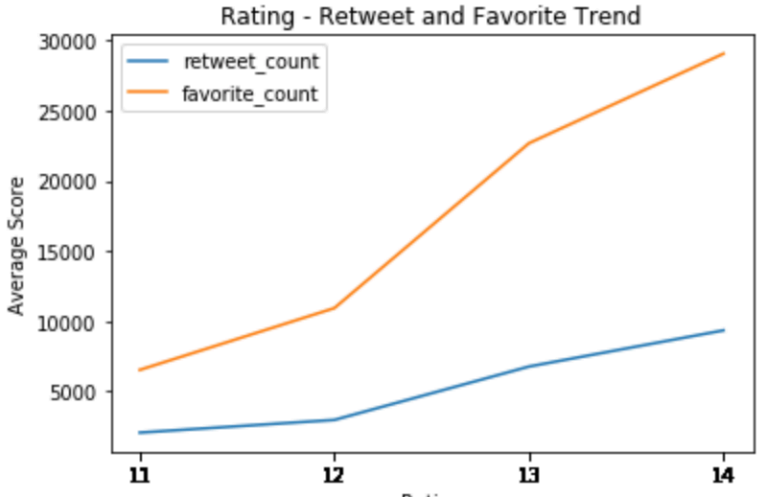
WeRateDog Twitter Data Analysis
Insights & Visualization Report
By **Brian Luk**

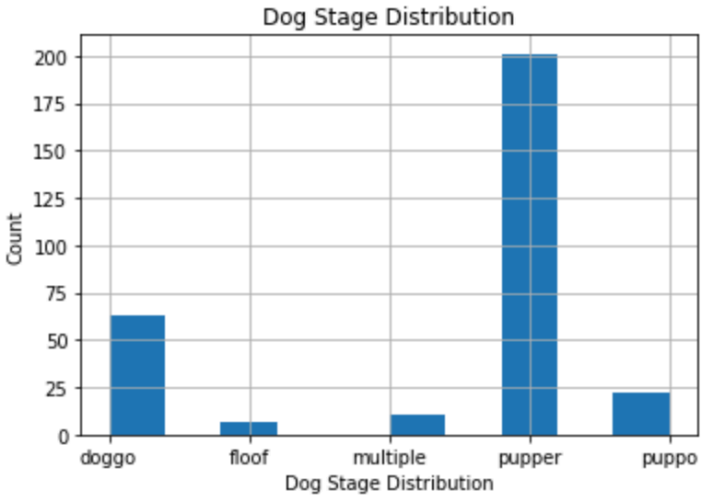
My analysis and visualization are mainly on the number of retweet and favorites based on dog stage and the WeRateDog rating.

The first question I wanted to investigate is the relationship between WeRateDog rating and number of retweet and favorites. After my data wrangling, I notice the rating data range is from 11 – 14, while any data outside of the range are all outliers. I plotted a line of the average retweet count and favorite count for each of the rating score.



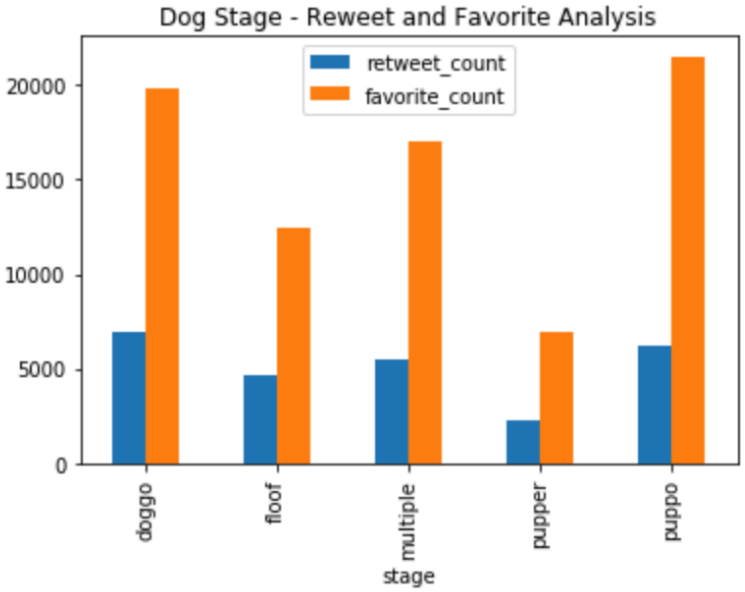
From the line graph, I notice that there is an obvious positive correlation between the WeRateDog rating score and the retweet and favorite count.

Then, I investigated the data surrounding dog stages. I was interested in discovering the distribution of dog stages. So I plotted a histogram.



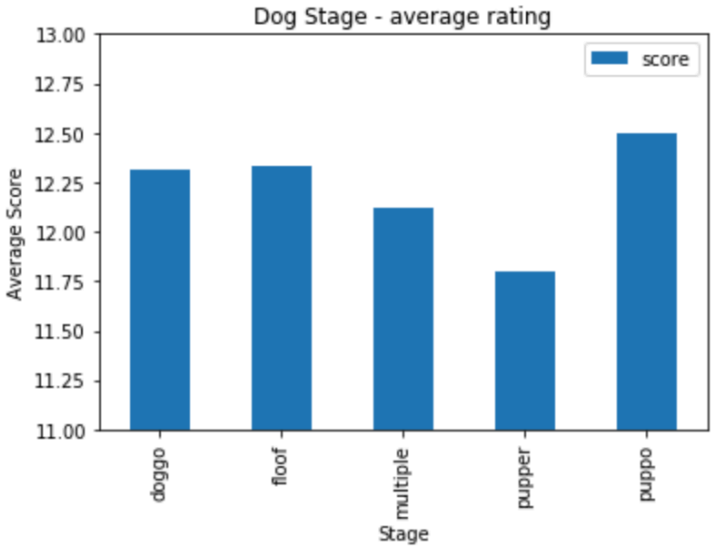
I notice that pupper has the highest frequency. However, based on the value_counts of the data stage column available in the data, only around 300 of the data has dog stage label in the 2000+ rows of data.

After that, I plotted a bar chart of the average retweet count and favorite count grouped by dog stage.



My insight is that most of the dog stage have similar retweet and favorite count except for pupper, which incidentally also has the highest frequency in tweet.

I then asked the question what could be the factor of that? Maybe WeRateDog simply rate pupper with a lower score than the other dog stages. Therefore, I plotted a bar chart of the average WeRateDog rating score grouped by dog stage.



Very interestingly, pupper does have the lowest average score of all the dog stages.

For further research analysis, I would like to know why is that? Is it misleading data interpretation simply because all the other dog stages don't have enough score simples? Or Twitter just doesn't like pupper as much as the other dog stages. It will be an interesting discovery.