ACT REPORT

I analyzed the gathered data from WeRateDogs twitter feed. This is a twitter account that rates the picture of dogs that he is tagged in. Upon gathering and cleaning the data. I did some basic analysis of the cleaned data. They are as follows.

1 - I extracted the top 10 popular breed of dogs according to predication algorithm. This was done by selecting all the breed of dogs that was predicted by the algorithm except the case where the algorithm could not predict the type of breed. In such situation, I set the data to 'No prediction'. As can be seen from figure 1, golden retriever was the most popular type of dog breed as predicted by the algorithm with about 175 count, followed by Labrador retriever on the number 2.

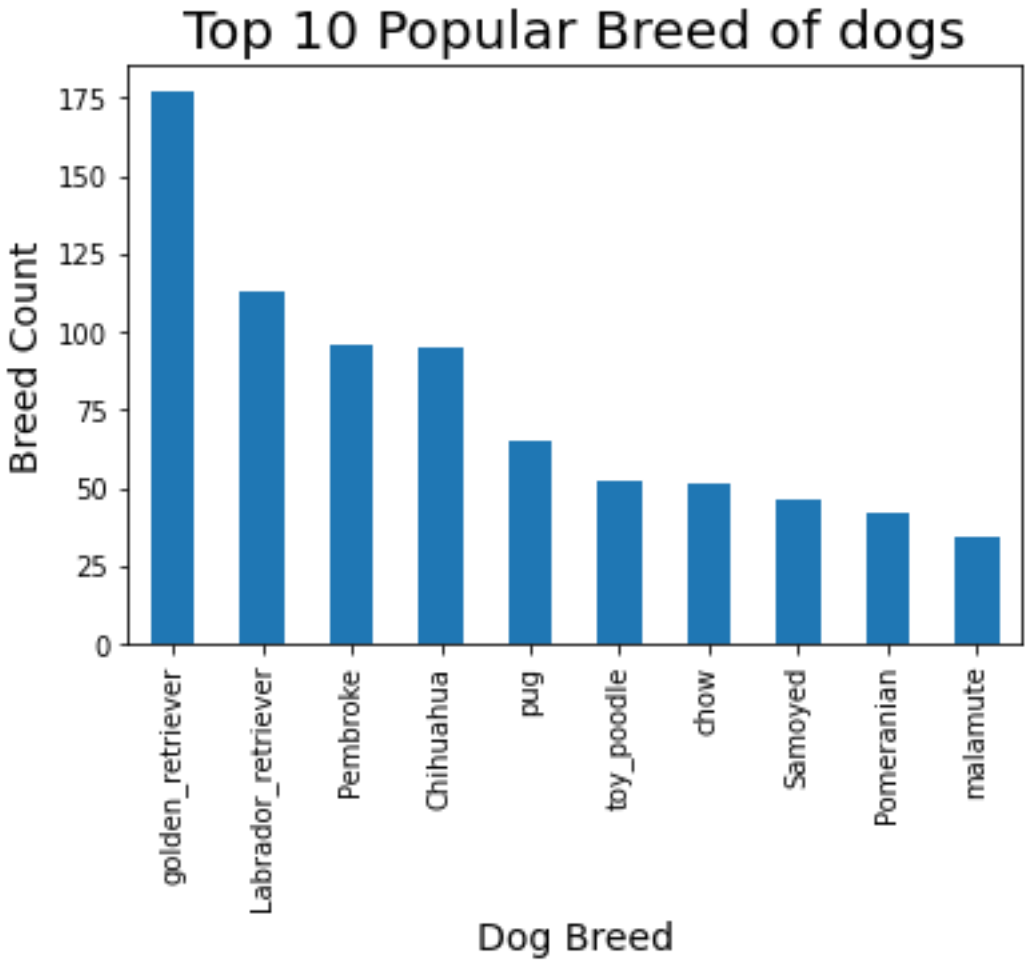

Figure 1 : Top 10 Breed of dogs by the algorithm

2 – According to the data, there are four stages of the dog stages. I analysed the gathered data to get the most popular dog stage. As displayed in figure 2, pupper is the most popular stage of dogs and it has more than tnwo times count the count of doggo that is second most popular.

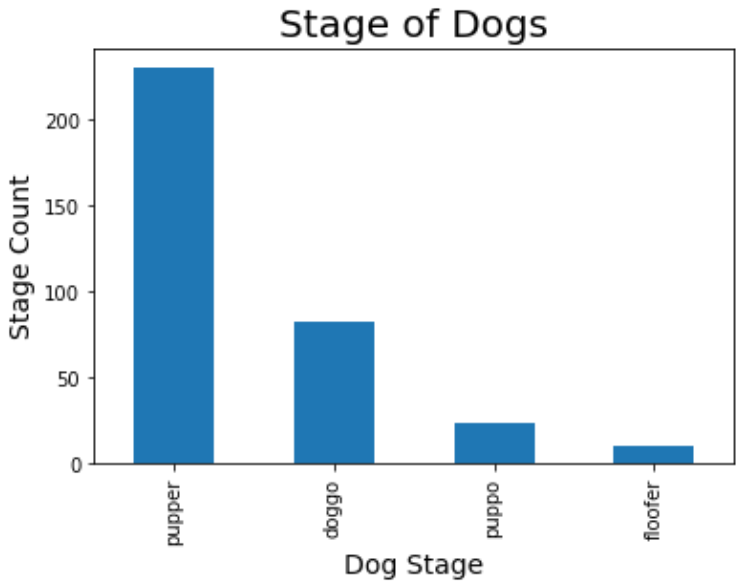

Figure 2 : Dog stage by popularity

3 – Finally, I did a quick check on the correlation between the retweet counts, favorite counts and the confidence rating of the predicted dog name. I wanted to see if there is any relationship between the three variables. AS expected there is a strong correlation of 0.71 between retweet counts and favorite counts but the rest of the correlation value suggests low to none correlation between the retweet counts and the confidence ratings.

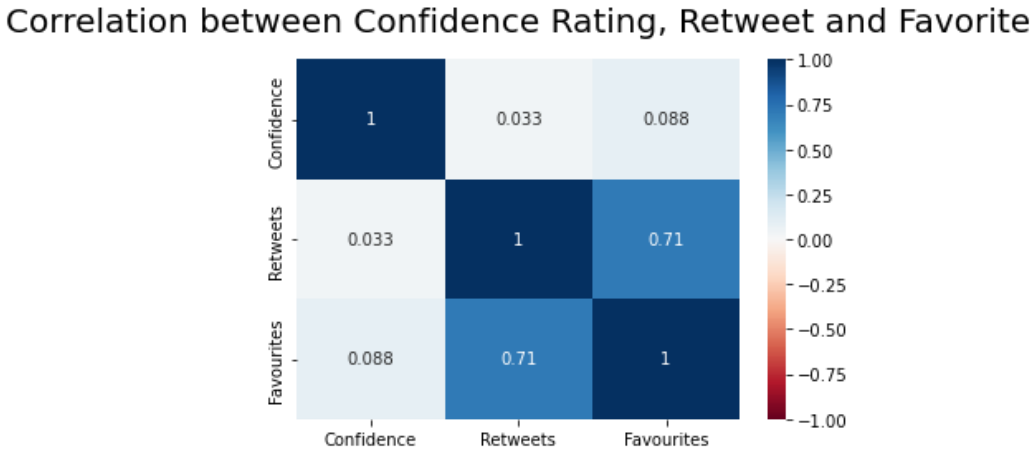

Figure 3: Correlation between confidence rating, retweet counts and favourite counts