

WeRateDogs Twitter Archive – Act Report

The twitter project contains three data. The *archive_df* table has 2356 rows and 17 columns, which provide the information dog name, stage, rating score and relevant information.

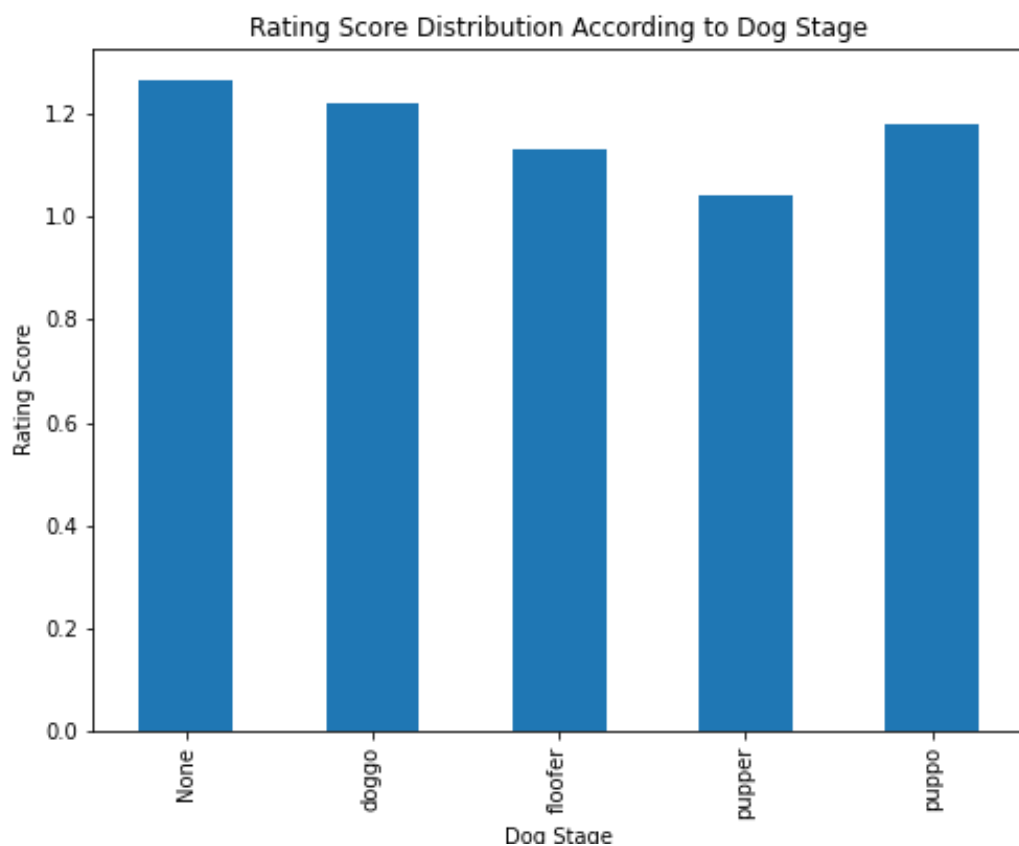
The dog breed information is provided by the *prediction* table, which has 2075 rows and 12 columns. It is the result to predict the dog breed via three algorithms in the neural network. To clear the prediction dog breed, I also extract the predicted dog breed which show the true value in *df_cof* and get the highest confidence value among the three algorithms.

The tweet data is json data gather from twitter. It has 2354 rows and 3 columns and provide the essential two columns retweet count and favorite count.

After accessing and cleaning the data, the three datasets were merged into a single dataset named 'twitter_archive_master.csv'. This consolidated dataset contains information from all three datasets and serves as the final output of the data cleaning process.

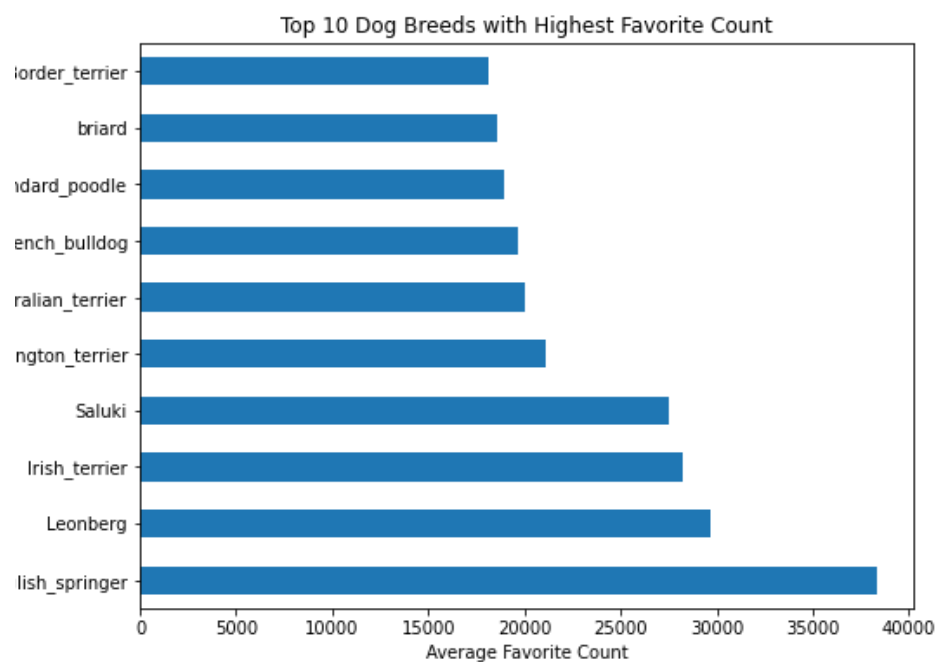
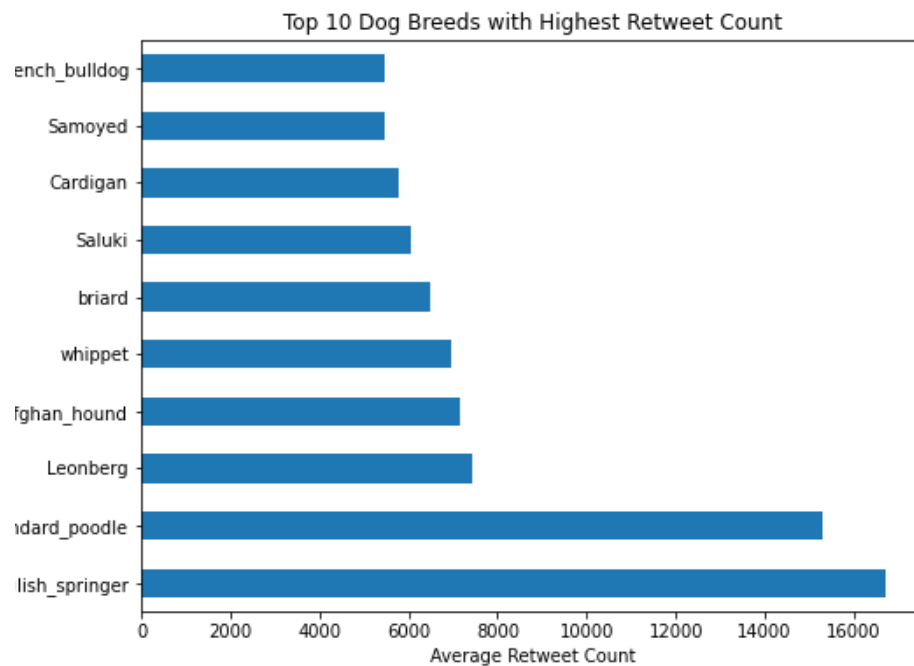
Insight

1. Which stage of dog received the highest rating score?



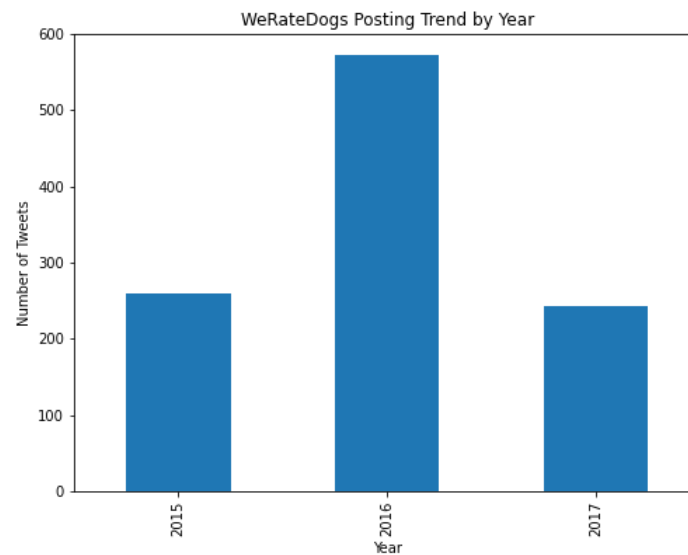
Based on the provided plot, it can be concluded that the dog stage labeled as "None" received the highest rating score. However, among the known dog stages, the "doggo" stage appeared to have the highest rating score.

2. Which dog breed type garnered the highest favorite count?
 There is two count present favorite count: favorite count and retweet count.

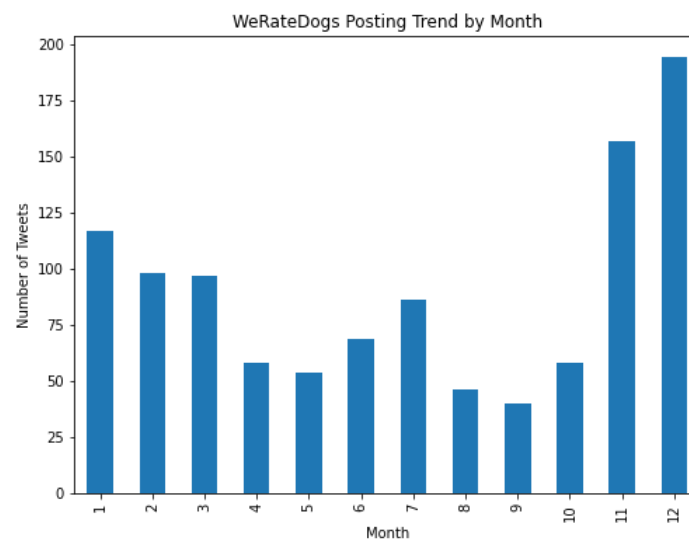


Since there is two count : favorite_count and retweet_count, representing the favorite count. Considering these two plots, we conclude that English Springer is the most favorite count dog breed in WERATEDOG tweet. The Standard_poodle, Leonberg, briard, Saluki, French_bulldog are listed in top 10 favorite dog breed.

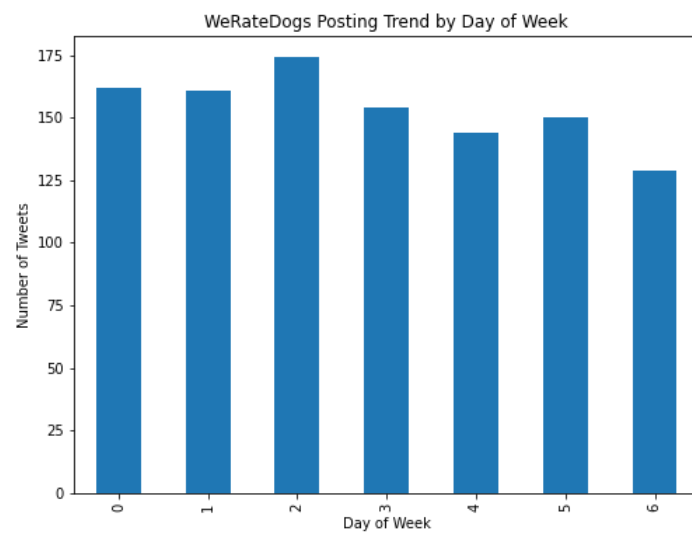
3. How does the posting trend of WeRateDogs vary by time?



From above analysis, we observe that the WeRateDogs account has posted a significantly varying number of tweets each year. In 2016, the maximum number of tweets was more than double the number in both 2015 and 2017. Additionally, the number of tweets in 2015 slightly exceeded the number in 2017.



The posting trend by month shows significant variations in the number of tweets. The months of October and December have the highest number of tweets, whereas the summer months of August and September experience the lowest points in terms of tweet activity.



The number of tweets posted per day exhibits slightly variation, only showing a small posting peak on Tuesday.