

Analyze and visualize data

After Gathering, Assessment and cleaning data, we want to analyze this data so that we can derive insights from this data and draw some visualization

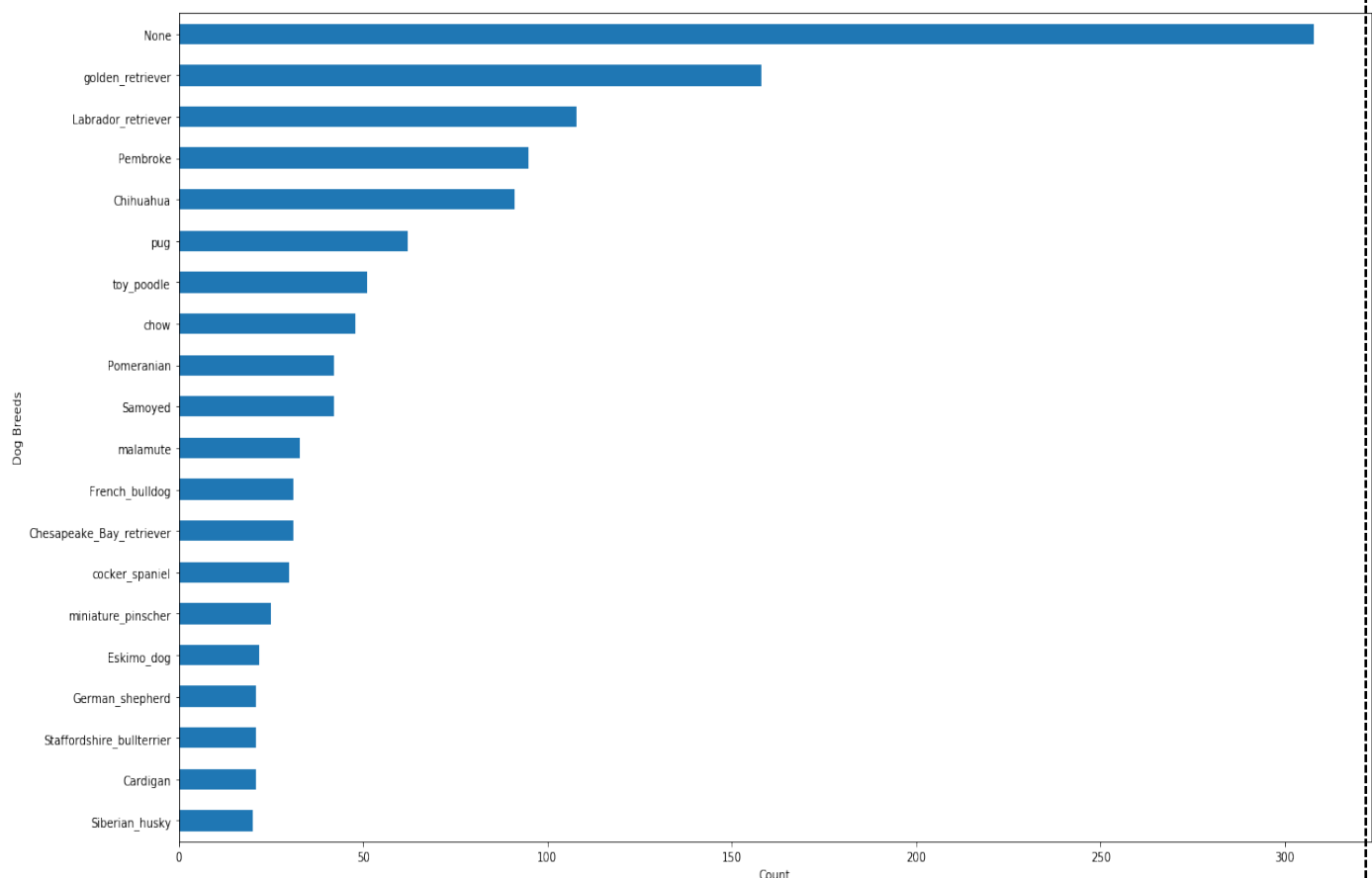
First of all we have stored our master cleaned data in a csv file using 'to_csv' function in pandas.

Some question we want to answer through analyzing data :

1. What are the most popular dog breed images in the tweets ?
2. The relation between number of retweets and number of favorite count (likes)
3. Percentage of dog stages presence in the dataframe
4. Average rating with each time interval
5. The most popular dog names among people

Question 1

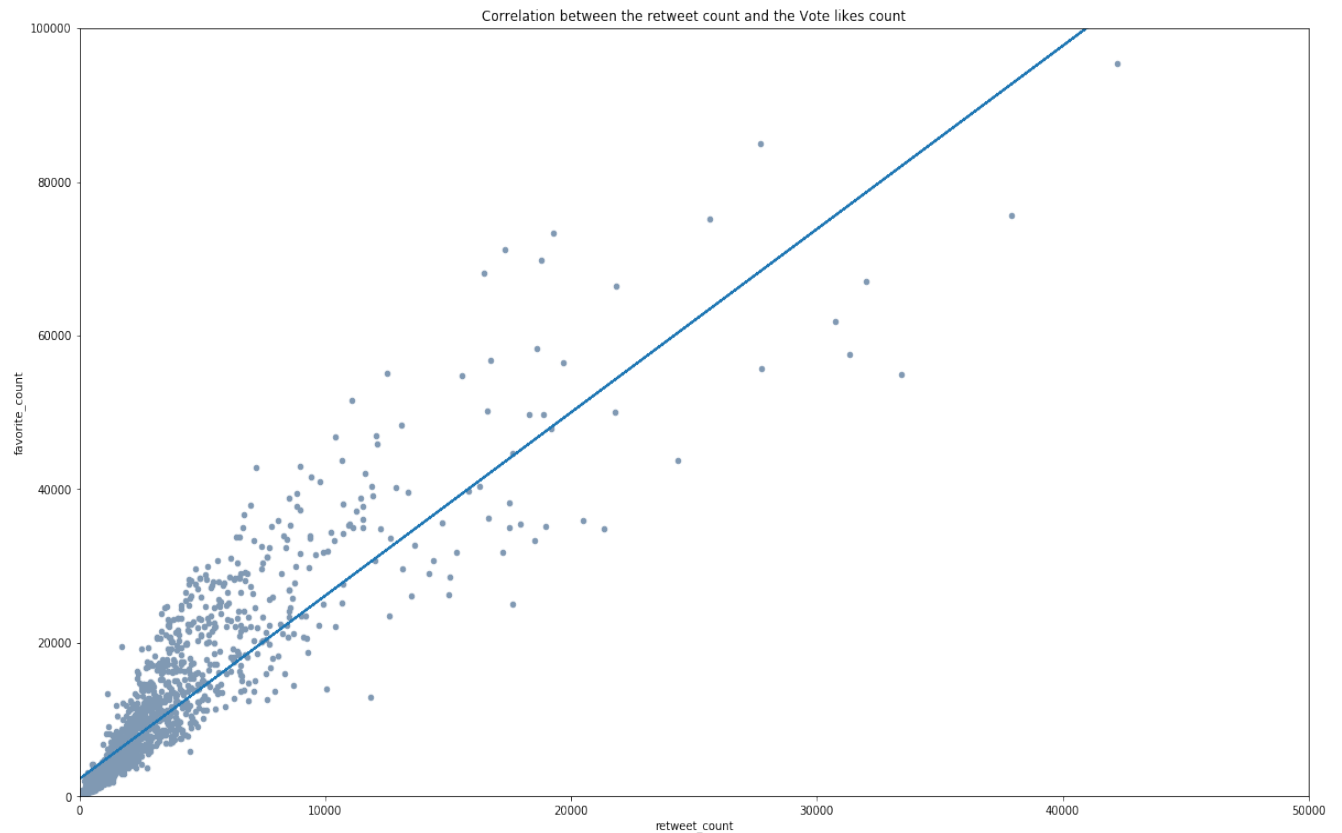
The below graph shows the count of the top dog breeds identified in the tweets images



From the above graph , The most common dog breed in the images where the algorithm has identified a dog in them , is the 'golden_retriever' followed by the 'Labrador_retriever' then 'Pembroke' then 'Chihuahua'

Question 2

The below scatter plot shows the relation between the retweet count and number of likes

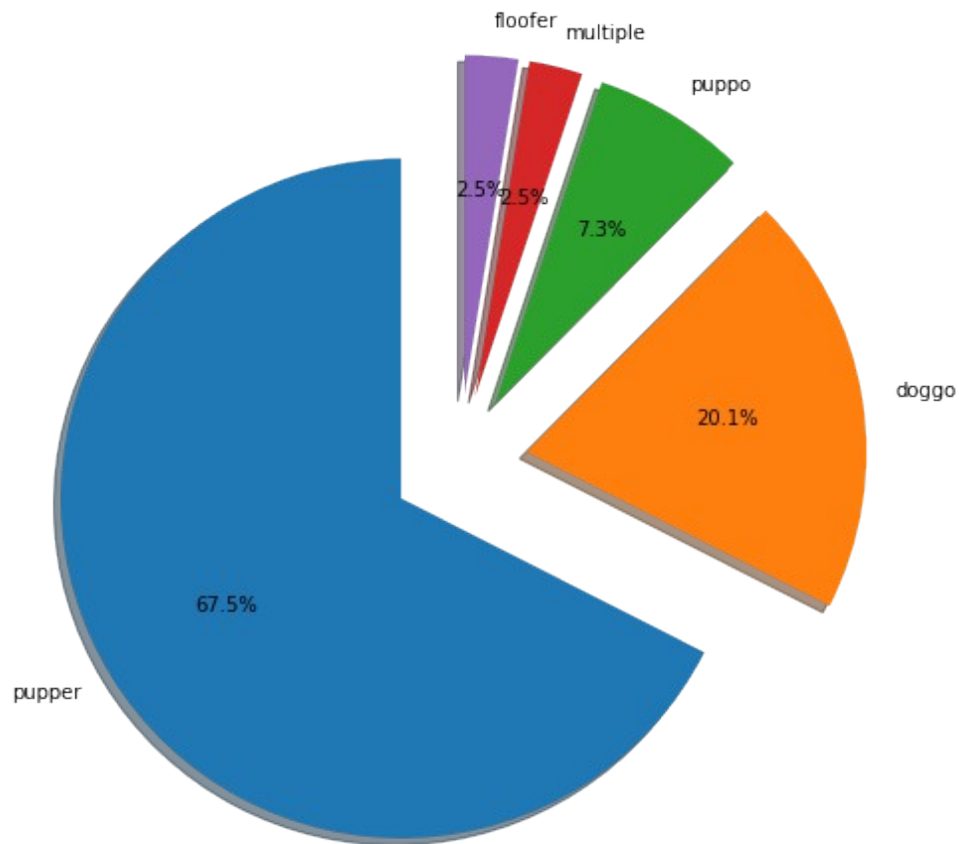


From the above graph , It seems that there is a strong positive correlation between the number of retweets and count of likes

By using the calculations, the correlation is computed to be 0.912

Question 3

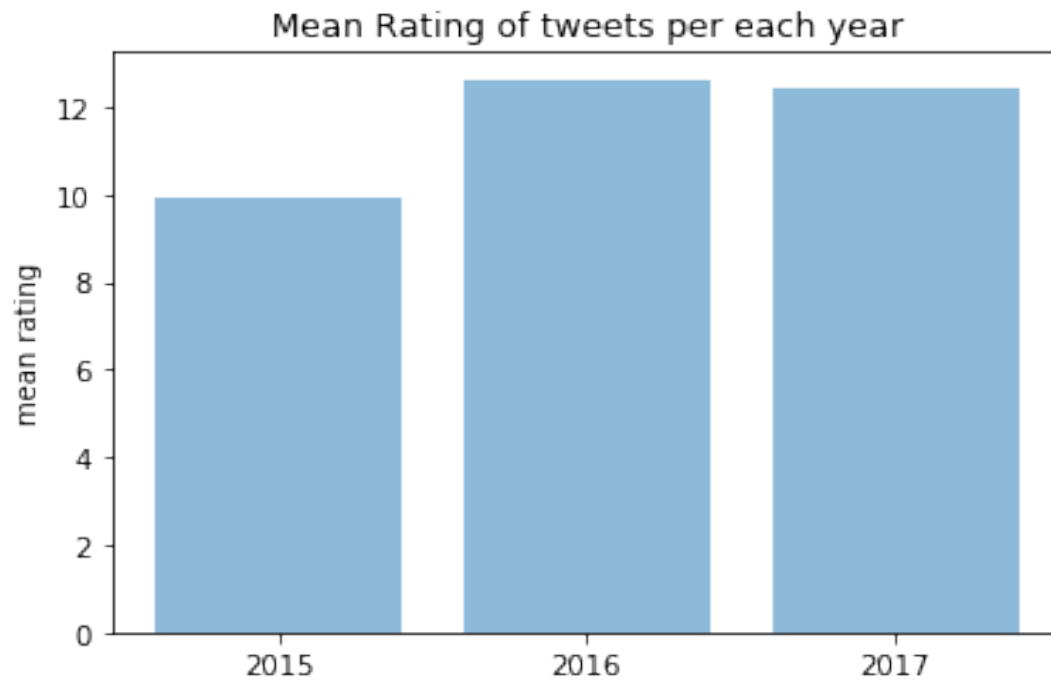
The below pie chart , show the percentage of each of the dog stages in the tweets data



From the above graph , pupper has the highest percentage in the tweets almost 70 % while the lowest percentage is 2.6 % for floofer and also for multiple dog stages

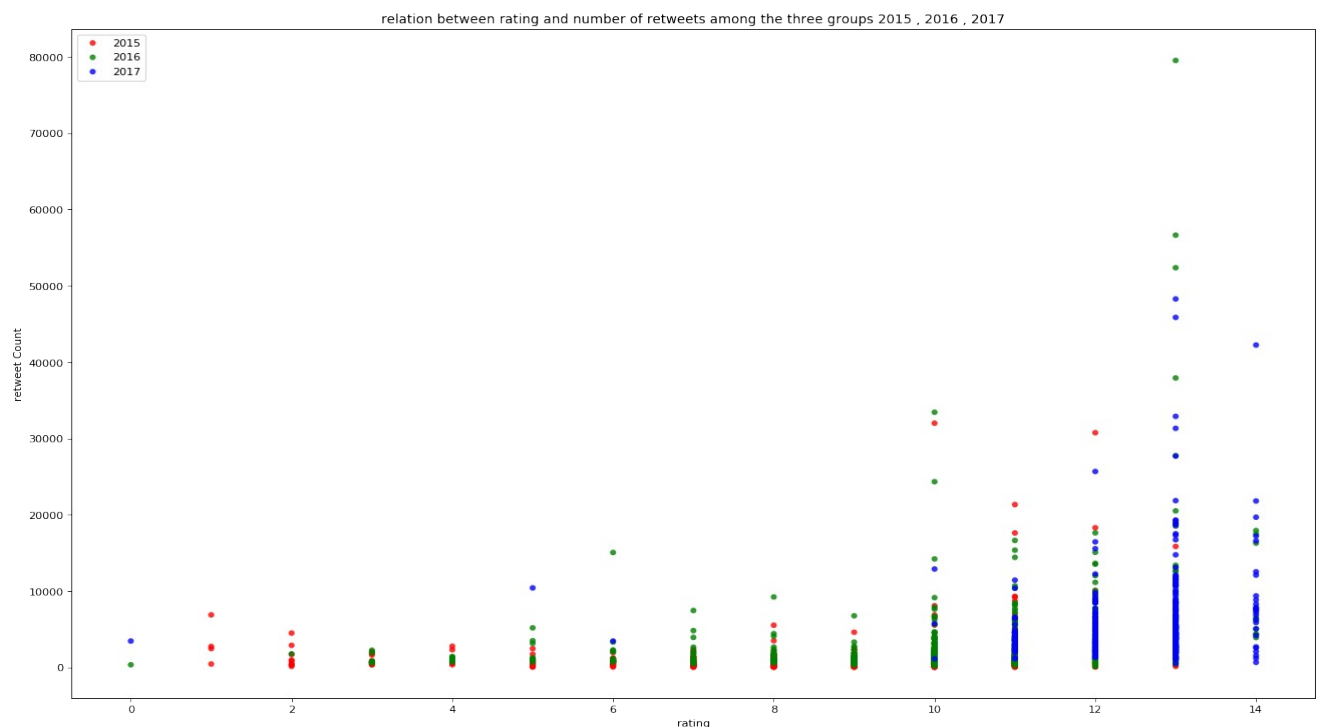
Question 4

The below bar chart shows the variation of mean rating among the tweets that were published in 2015 , 2016 and 2017



From the above graph, 2015 year has the lowest mean rating , while 2016 has the highest mean rating which is over 12.

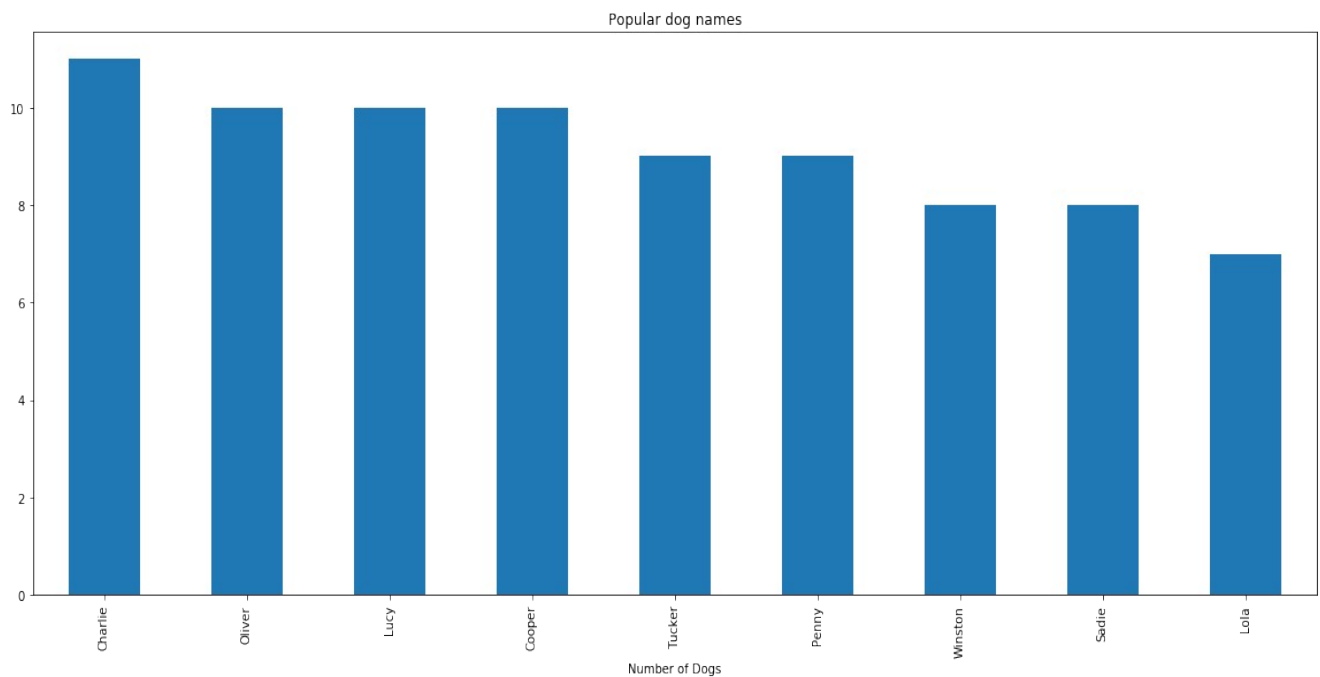
The below scatter plot is showing the relation between the rating values and its effect on the number of retweets among the 3 periods 2015 , 2016 , 2017



From the above graph, There is a slight direct relation between the rating and the number of retweets , and this appears especially in tweets written in 2017

Question 5

The below bar chart shows the top 10 names count of the dogs in the tweets



From the above graph, Charlie is the most popular name with 11 times repetition , followed by oliver , Lucy and cooper each of 10times repetition.