Analyzing, and Visualizing WeRateDogs Data

-Gathring and Wrangling Data:

The WeRateDogs Data gathered from 3 resources:

Enhanced Twitter Archive file which contains basic tweet data for all 5000+ of their tweets and after filtering tweets with ratings only the ratings now contains 2356 tweets.

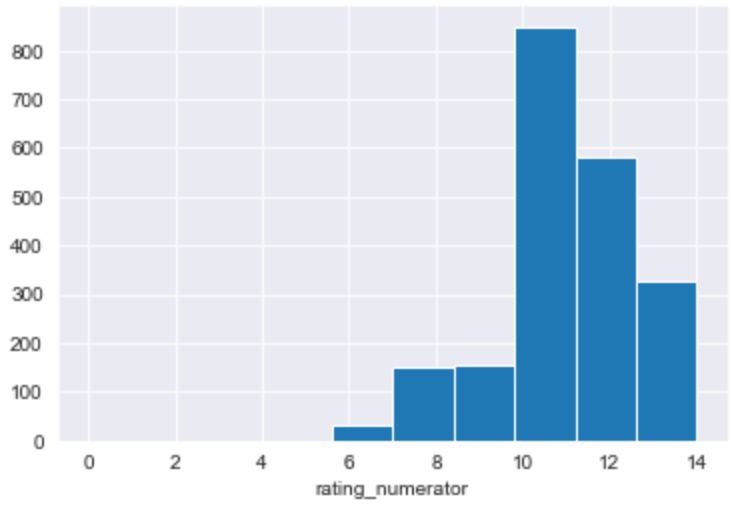
Image Predictions File which contains algorithms for prediction of the dog breed based on the image url provided in tweets .

Additional Data via the Twitter API which contain retweet count and favorite count and both are very important to our analzing

After gathering these resources the data collected went through a Data Assessment processing that was done visually and programmatically based on quality of data and the data tidiness key points , afterward programmatically cleaning data process to make sure the analysis which will be shown below to be accurate as possible and provide useful information.

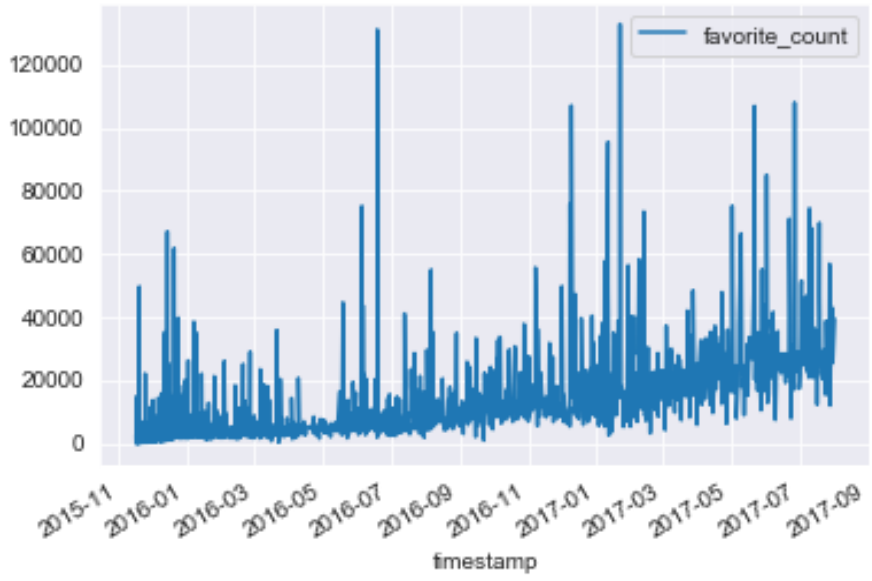
-Insights :

1- Getting an idea about the distribution of ratings given.

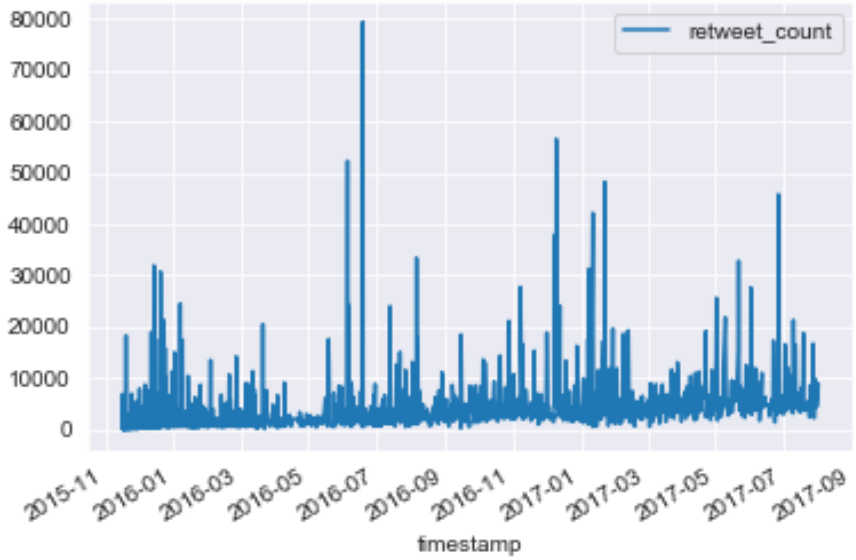


As shown above the distribution is left skewed , and the average of ratings givin is 10.9 and the most of ratings are between 10 to 12.

2-Find the correlation between timestamp and the retweets and favorite¶

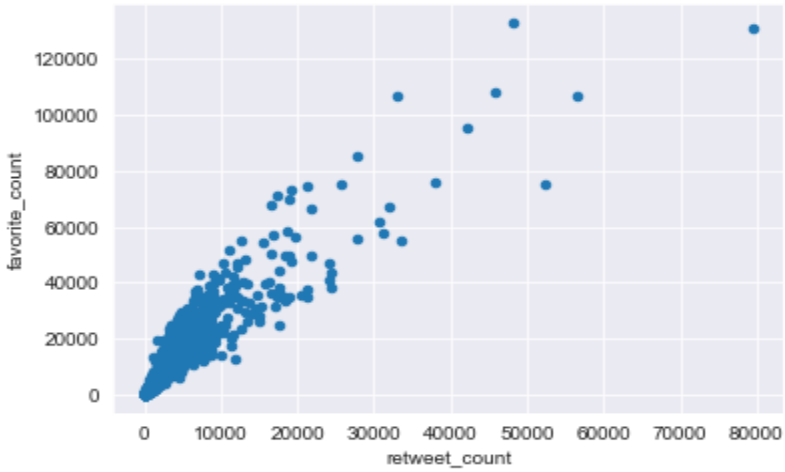


This shows an increase in the numbers of favorite counts with time , and notice here the peak of favorites at july16 and between feb17 and mar17.



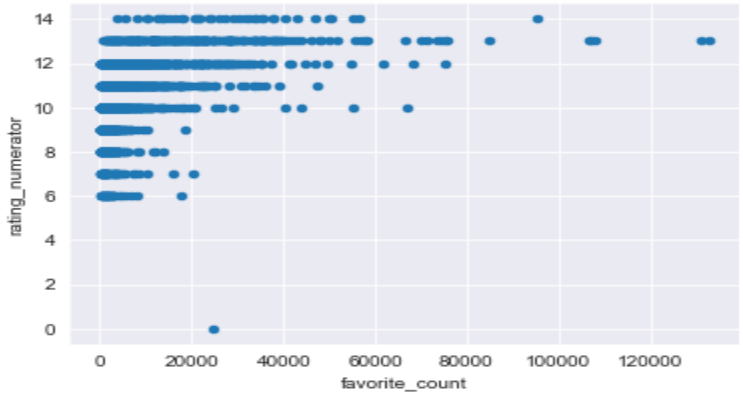
Once again we still have slightly increase of retweets over time , the first peak still around july16 ,but the second peak not as powerful as the one occured in favorites between feb 17 and march

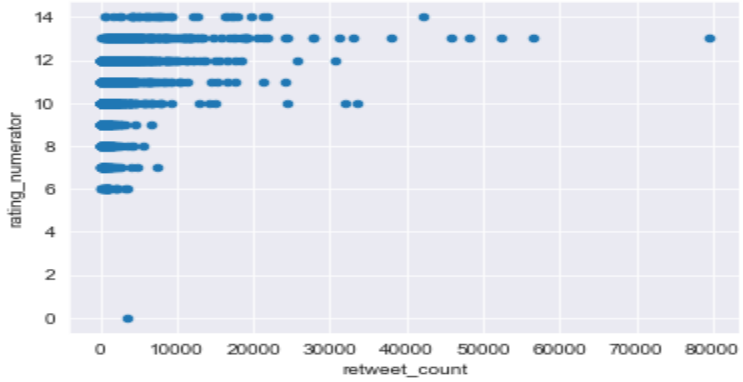
3-Find the correlation between retweets and favorite



As shown that there are positive correlation between retweets and favorite.

4-Find the correlation between ratings , favorite count , retweet count.





As Shown above the corrlation between ratings ,retweets and favorites are positve, and the ratings given more slightly above the mode have the most retweets and favorites.