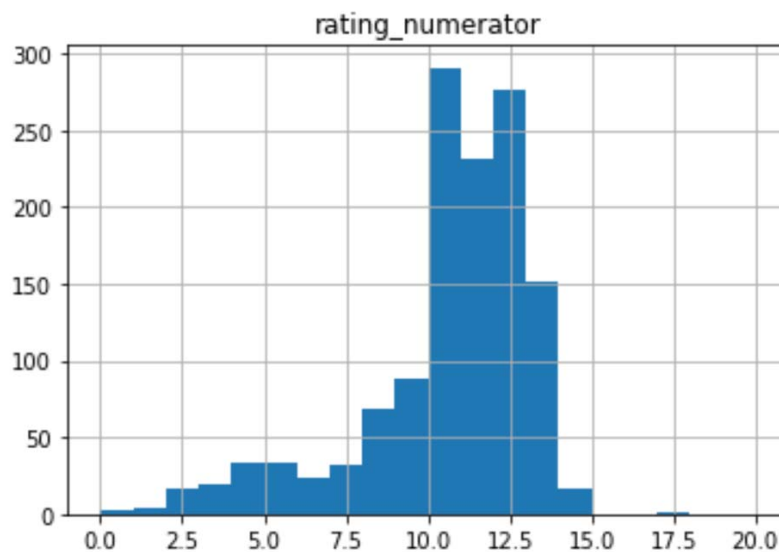


Analysis and insights from WeRateDogs

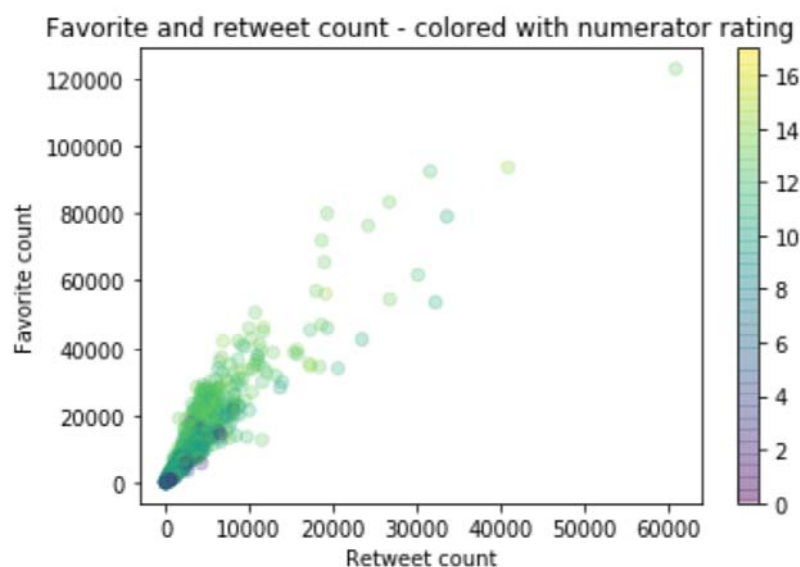
This analysis uses the cleaned dataset, which consists of tweets from the WeRateDogs group on Twitter. The tweets contains a lot of data but the cleaning process has significantly slimmed down the dataset. Furthermore, a machine learning algorithm has been used to identify the content of the pictures uploaded with the tweets.

The first insight produced from this analysis is the spread of dog ratings. Usually the dog will be rated on a scale up to 10, but it could be significantly higher as well. The denominator is usually 10 and the spread of the numerator for the investigated tweets are shown below.



The ratings given seems to behave in an unimodal behaviour with two peaks with a negative skew. One peak at 10 and one at 12. It seems that some people rates their dogs lower than the denominator of 10.

Another interesting part of the dataset is the amount of retweets and favorites that a tweet gets. A positive strong correlation was found between the count of retweets and favorites using Pearson's criteria. These two are plotted together below with the rating numerator as the color code.



From this analysis it seems that there is in fact a strong linear positive relationship between the two parameters. Another interesting fact is that the color coding could indicate that it is dogs with a very high rating that also gets a high favorite and retweet count.

Next the machine learning algorithm is tested to investigate how well it actually predicts the correct dog in the picture. The three predictions with the highest confidence are investigated below.

Machine Learning algorithm			
Dog type	Confidence	url	Actual dog type
Komondor	0.99995599999999996	https://t.co/EM5fDykrJg	Komondor
Chow	0.99983699999999998	https://t.co/E9jaU59bh9	Chow
Dalmatian	0.99982800000000005	https://t.co/01vpuRDXQ8	Dalmatian

The three predictions with the highest confidence were all spot on. The first in the table above, was identified as a *Komondor*, but when reading the tweet text it actually says it is a *Rigatoni Terrier*. However, after reading the comments on the tweet it becomes clear that the tweet text seems to be wrong and in fact it is a *Komondor* as the Machine Learning algorithm correctly predicted.