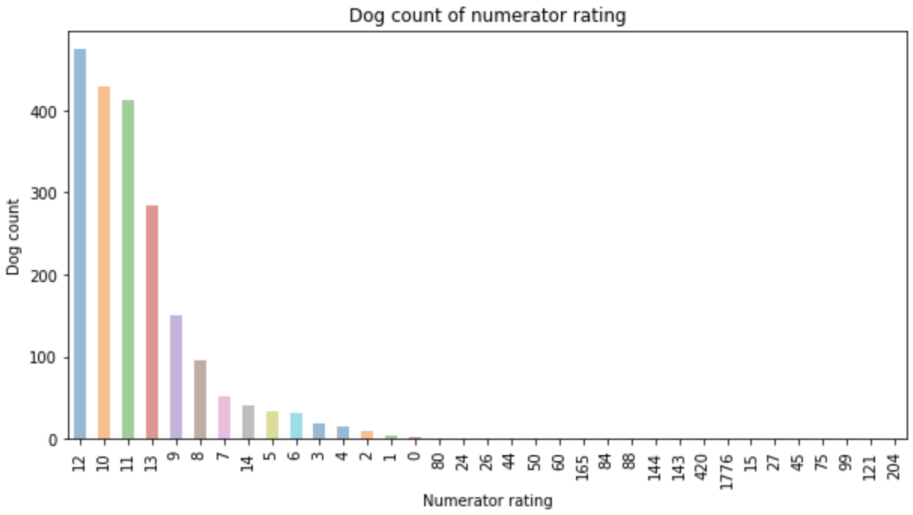
Report: act_report

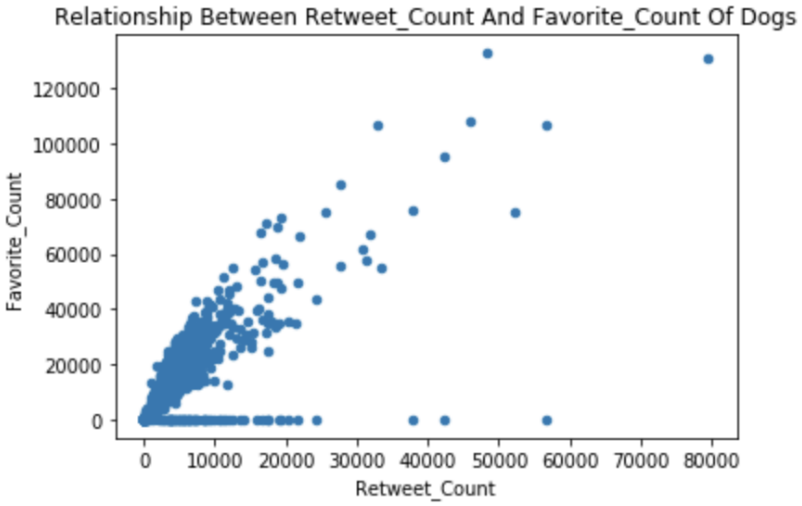
From the trends that reigned years back on twitter, the ‘WeRateDogs’ trend stood out with over 5000+ tweets on dog ratings. Three datasets from the ‘WeRateDogs’ archive were gathered, assessed, analysed, cleaned, and merged to form a master dataset. This merged dataset was insightful and generated visuals that can be easily understood and appreciated.

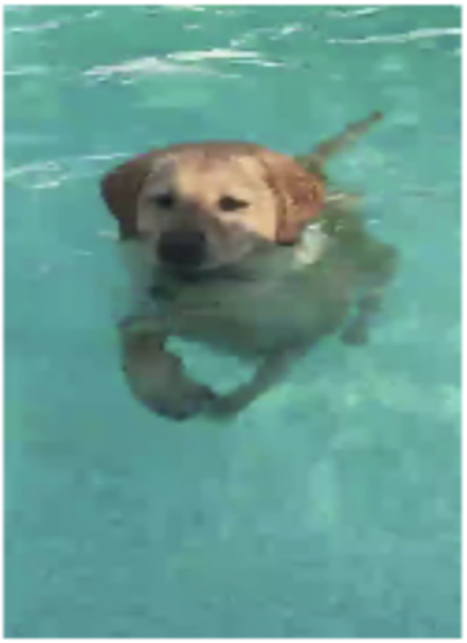
The insights that could be seen in, and visual created from the data are below;

1. The first insight is on the dog rating, we see that most dogs have rating values above the denominator value of 10, from the introductory article we understand that this is because some of the raters simply loved the dogs and believed they deserved more that 10 points. Using the method .describe() to generate statistical information, 25% of the rates are at 10points, 50% at 11points and 75% at 12points. The image displayed below is a bar chart displaying the number of dogs that have certain rating value, therefore it can be assumed that 400+ dogs have a rating of 12/10.



2. Two types of responses given to a tweet are by retweeting such a tweet or by liking(favorite) the tweet, and so another insight can be deduced from the retweet and favourite count to see the relationship between the two variables. Also the data can be analysed to show which dogs using their tweet_id have the highest retweet_count and the highest favorite_count. The first visual seen below shows the relationship between the retweet_count and favourite_count of dogs. While the second visual is a picture of the cute dog with tweet_id 744234799360020481 which is in the top five of the highest retweet_count and the highest favorite_count. Unfortunately this dog's name was not included in the database.





3. A third insight can be seen after the number of different dog types are displayed. The majority of dogs are pupper with a count of 231, then there is doggo with 75 counts, puppo with 29 counts, and floofer with only 3 counts. The other dogs in the data were not categorised.