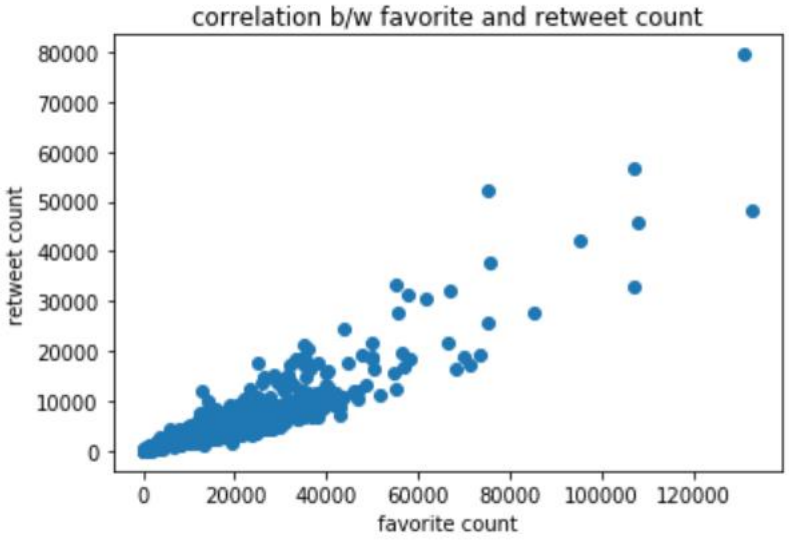
Report on Wrangling We rate dogs

We Rate Dogs rate pictures of dogs on their twitter account. This project was aimed at acquiring the datasets from different resources and clean them. Finally, to create a master dataframe which consists of relevant information on rated dogs. I was curious as to how ratings compared with people's response in the form of favorite count and retweet count. Also, is there any relationship b/w favorite count and retweet count. The final dataframe was analysed and following insights were found.

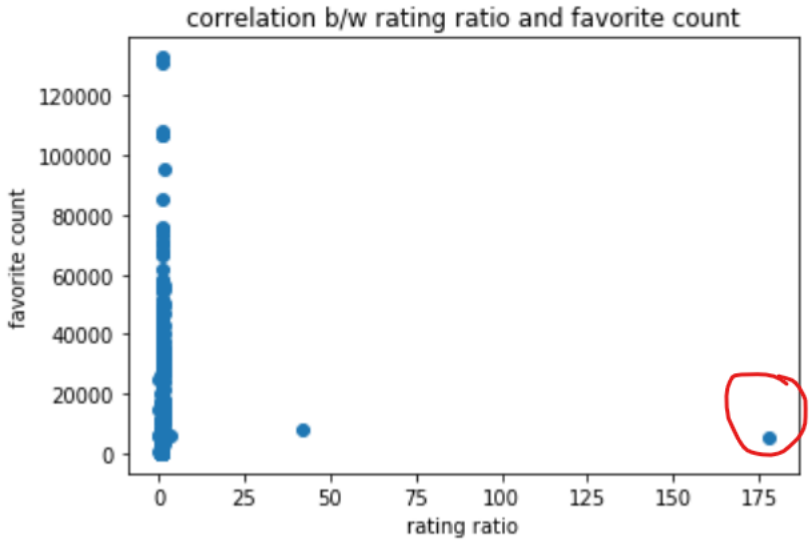
Insight 1

Looking at the correlation b/w retweet count and favorite count we see that there is a positive correlation and we can infer with some confidence that for high favorite counts there would be high retweet counts and vice versa.

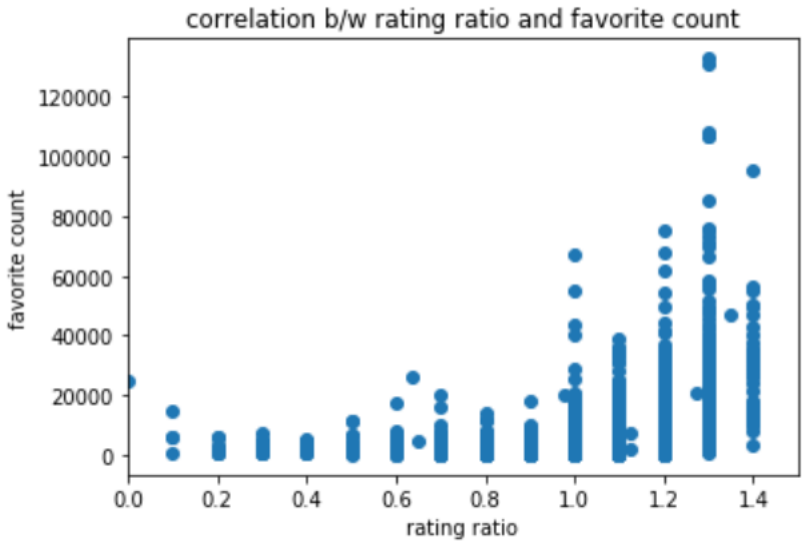


Insight 2

The rating ratio is the ratio between the rating numerator and denominator. It is used to normalize since we have multiple denominators. On plotting favorite count against rating ratio we see an outlier on the plot. This outlier is located at the extreme plotting which takes away the details of other ratings.



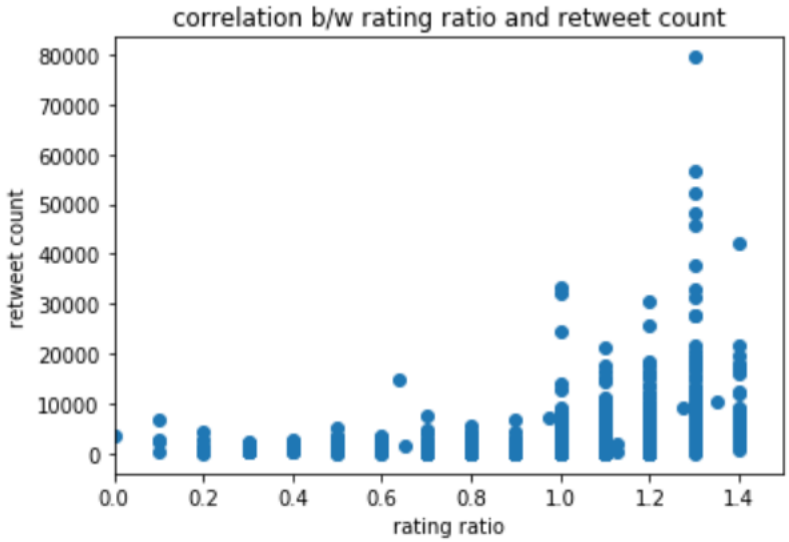
To get a detailed idea of the spread b/w the rating ratio and the favorite count, the x axis is limited to 1.5. This results in following plot.



We see a positive correlation b/w the 2 variables which means dogs with higher rating higher is often marked as favourite by public.

Insight 3

Similarly, on plotting the retweet count against rating ratio we see a positive correlation.



Hence, we can conclude with some confidence that in general dogs with higher rating also gets higher number of retweets and favorite counts.