Act report

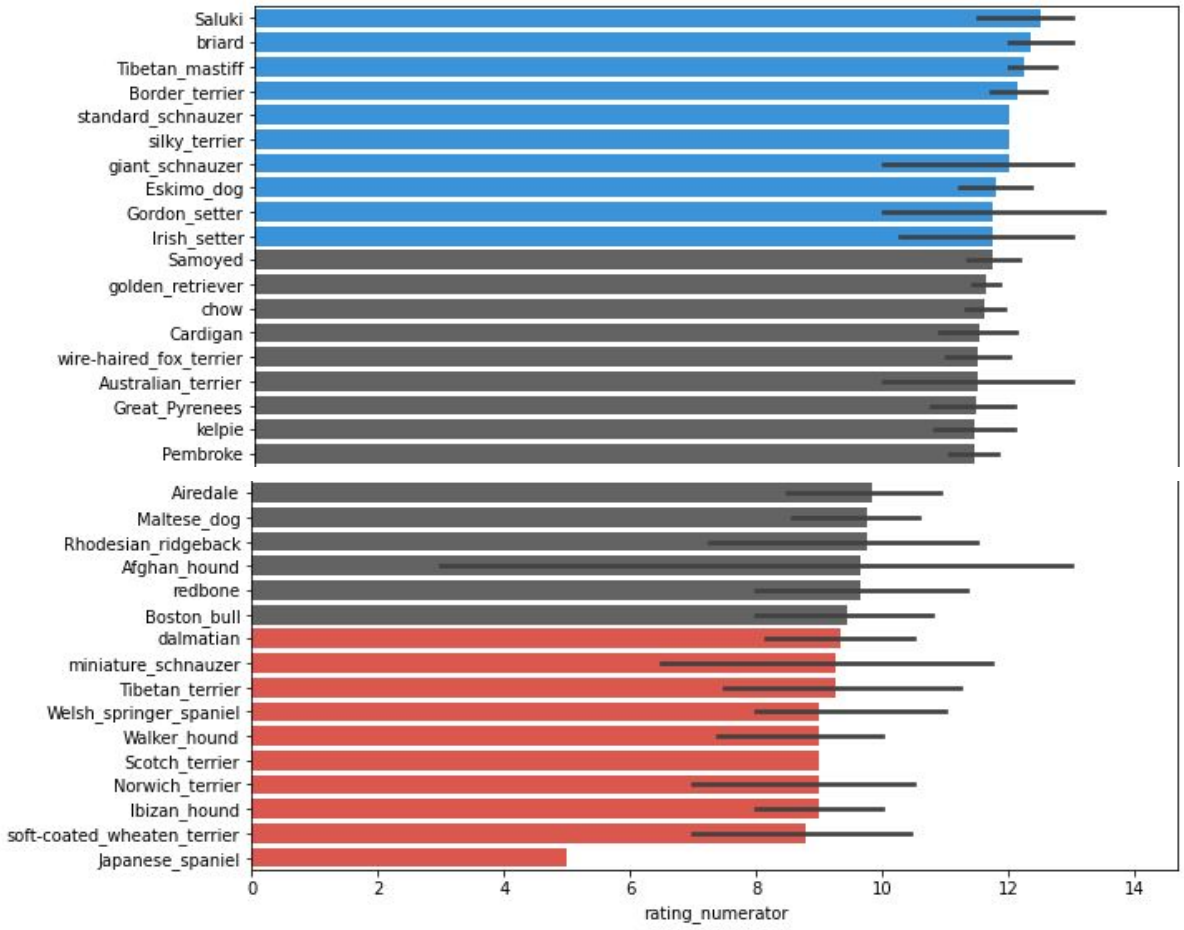
Insight 1. Which breeds are received good ratings?

Based on the average ratings of each breed, top 10 group and bottom 10 group are identified as below bar chart. I also investigate the distribution of ratings in the top 10 and bottom group dogs (histogram). Based on the histogram, the dogs in the top 10 group mostly get better grades than the dogs in the bottom 10 group.

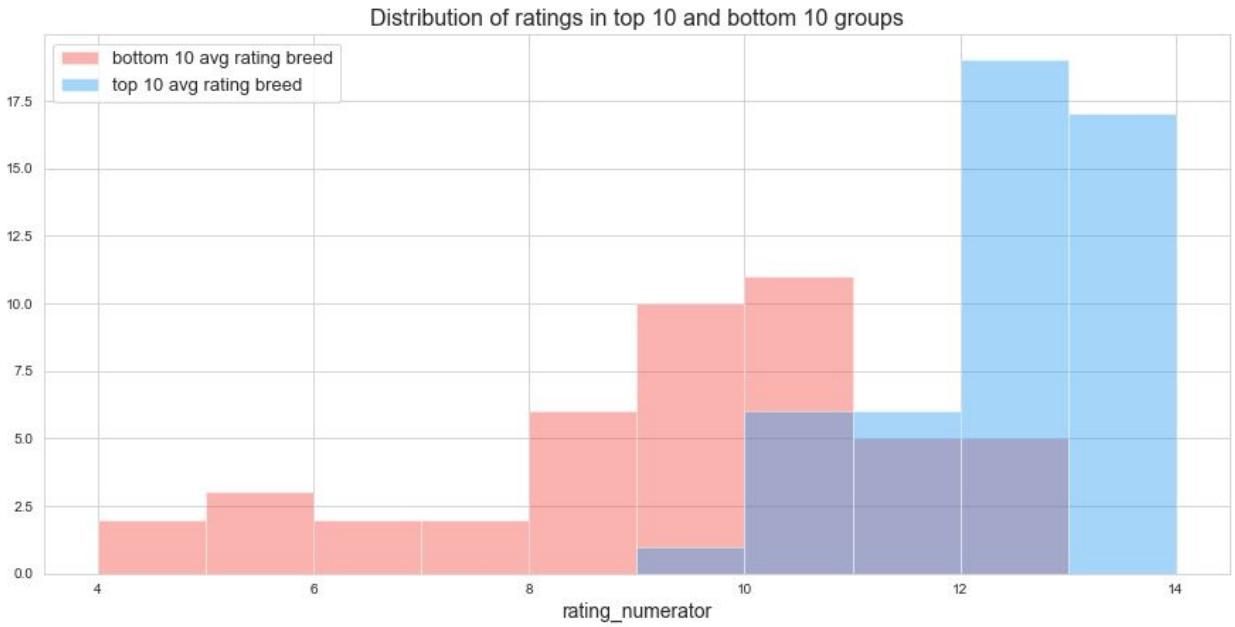
Top 10: 'Saluki', 'briard', 'Tibetan_mastiff', 'Border_terrier', 'standard_schnauzer', 'silky_terrier', 'giant_schnauzer', 'Eskimo_dog', 'Gordon_setter', 'Irish_setter'

Bottom 10: 'Japanese_spaniel', 'soft-coated_wheaten_terrier', 'Ibizan_hound', 'Norwich_terrier', 'Scotch_terrier', 'Walker_hound', 'Welsh_springer_spaniel', 'Tibetan_terrier', 'miniature_schnauzer', 'dalmatian'.

Picture 1. Average rating of each breed



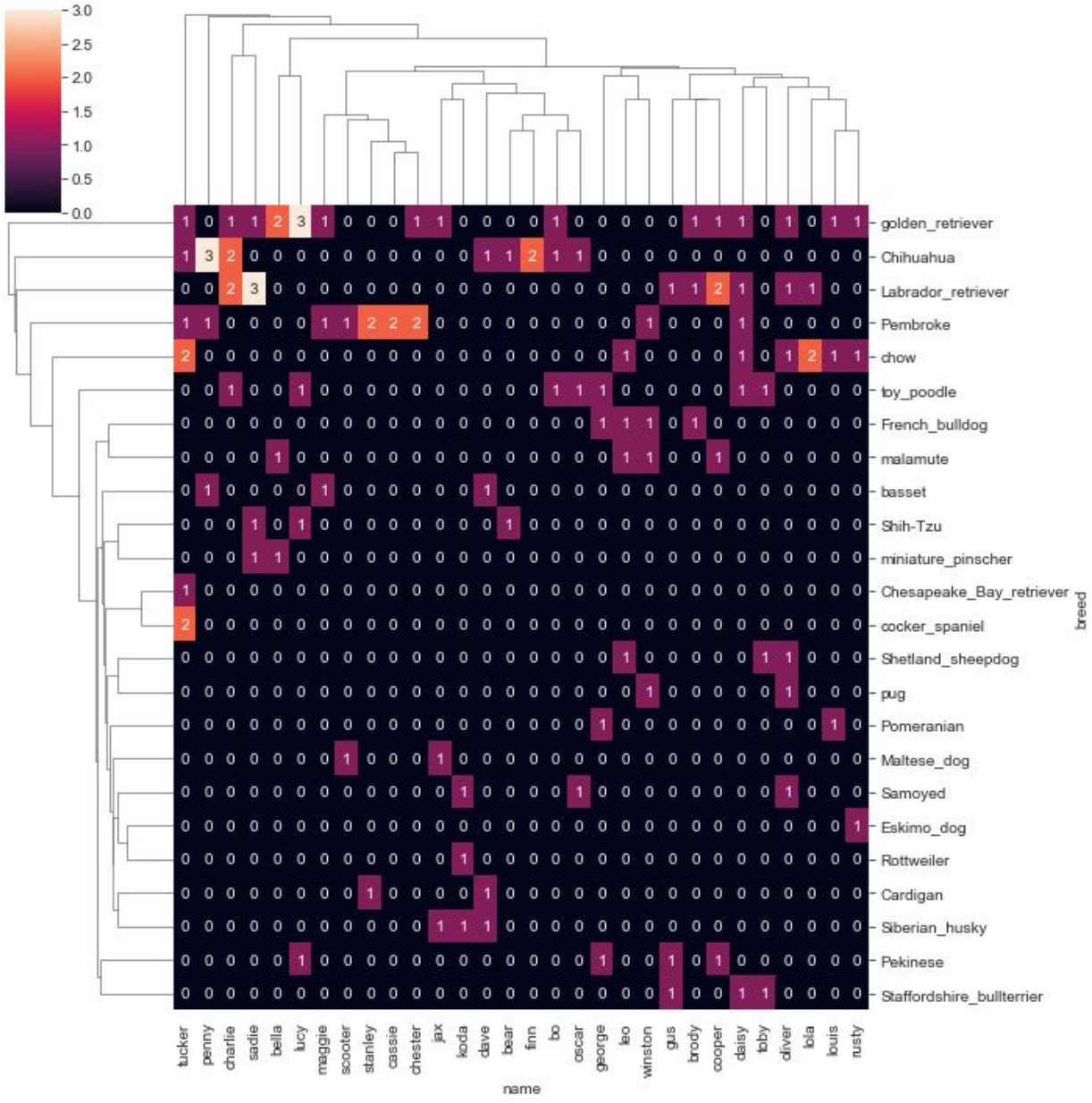
Picture 2. Distribution of ratings in the top 10 and bottom 10 groups



Insight 2. Is there any relationship between the name and breed of a dog?

To increase readability, I consider the most 30 popular breeds and names. As a below matrix, most of the popular breeds correspond to multiple names. Therefore, it is a little bit hard to say there is any relationship between the name and breed.

Picture 3. Matrix(heatmap) of the number of dogs between the name and breed

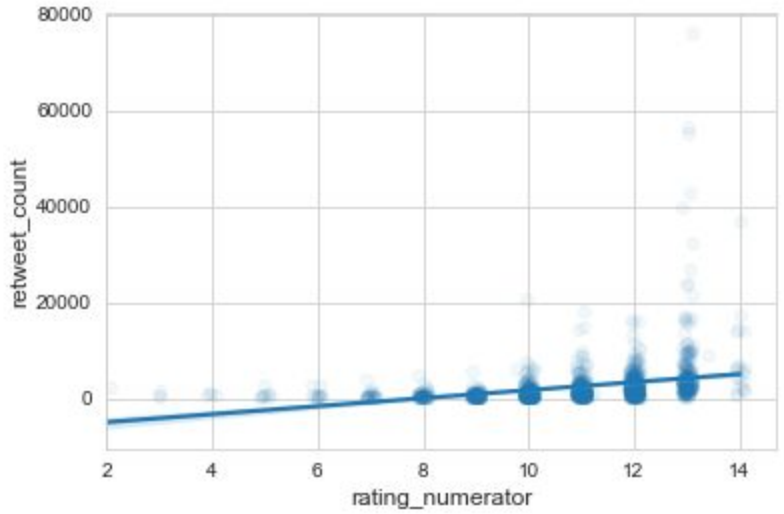


Insight 3. How are the ratings, retweet counts, and favorites counts correlated with each other?

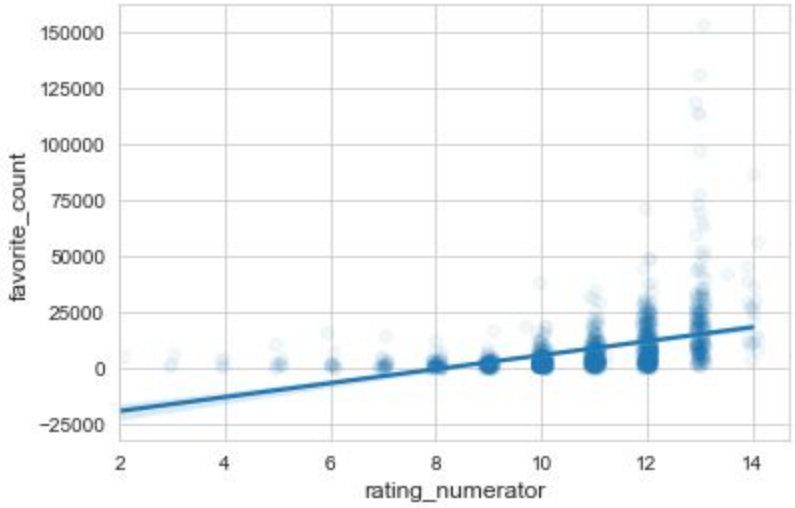
‘retweet_count’ and ‘favorite_count’ have weak positive correlations with ‘rating’. However, there is a strong correlation between ‘retweet_count’ and ‘favorite_count’,

Picture 4. Scatter plot between rating_numerator and retweet_count

Picture 5. Scatter plot between rating_numerator and retweet_count



Picture 6. Scatter plot between rating_numerator and favorite_count



Picture 7. Scatter plot between retweet_count and favorite_count