# Act report

`-By imen bakir`

In this report I will be communicating the insights coming from gathering and wrangling the datasets in this project, as well as displaying visualizations produced from the wrangling data.

Firstly, I wanted to see first 10 rated dogs based on the 'rating_numerator' using 'groupby' function and 'sort_values' function, and as a result we have the following dataset :

	dog_breed	rating_numerator
0	clumber	27.000000
1	soft-coated wheaten terrier	25.454545
2	West Highland white terrier	15.642857
3	Great Pyrenees	14.928571
4	borzoi	14.444444
5	Labrador retriever	14.011236
6	golden retriever	13.250000
7	Siberian husky	13.250000
8	Pomeranian	12.868421
9	Saluki	12.500000

To visualize this, I made a bar plot using seaborn and rsParams, this was the resulting figure:


Top 10 most rated Dog breeds
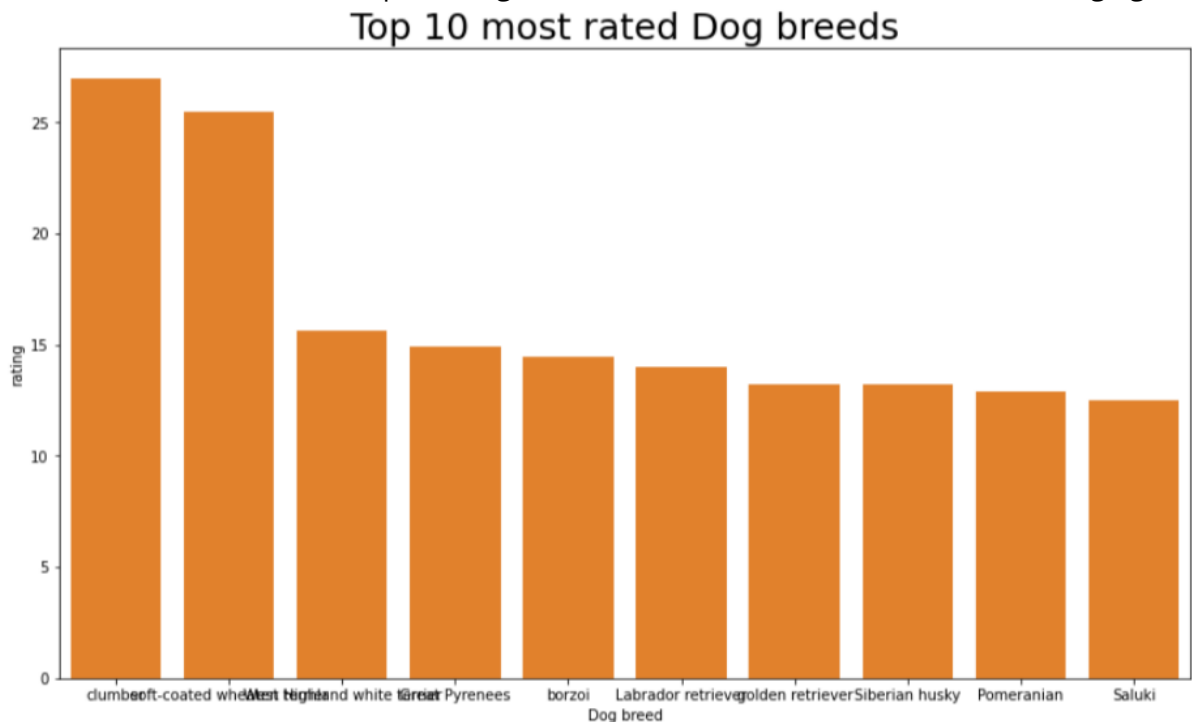
Then I wanted to display the image of the resulting most rated dog, using IPython's display library and by selecting the respective image url. This was the result :

```
from IPython import display
display.Image('https://pbs.twimg.com/media/Cswbc2yWcAAVsCJ.jpg')
```


Furthermore, I wanted to see the least 10 rated dogs based on the rating numerator, and this was the resulting dataset:

	dog_breed	rating_numerator
0	Japanese spaniel	5.000000
1	miniature schnauzer	8.333333
2	Ibizan hound	9.000000
3	Norwich terrier	9.000000
4	Scotch terrier	9.000000
5	Walker hound	9.000000
6	Welsh springer spaniel	9.000000
7	dalmatian	9.214286
8	Tibetan terrier	9.250000
9	Maltese dog	9.277778

As for the visualization, I made a barplot using seaborn and rsParams.

Top 10 least rated Dog breeds
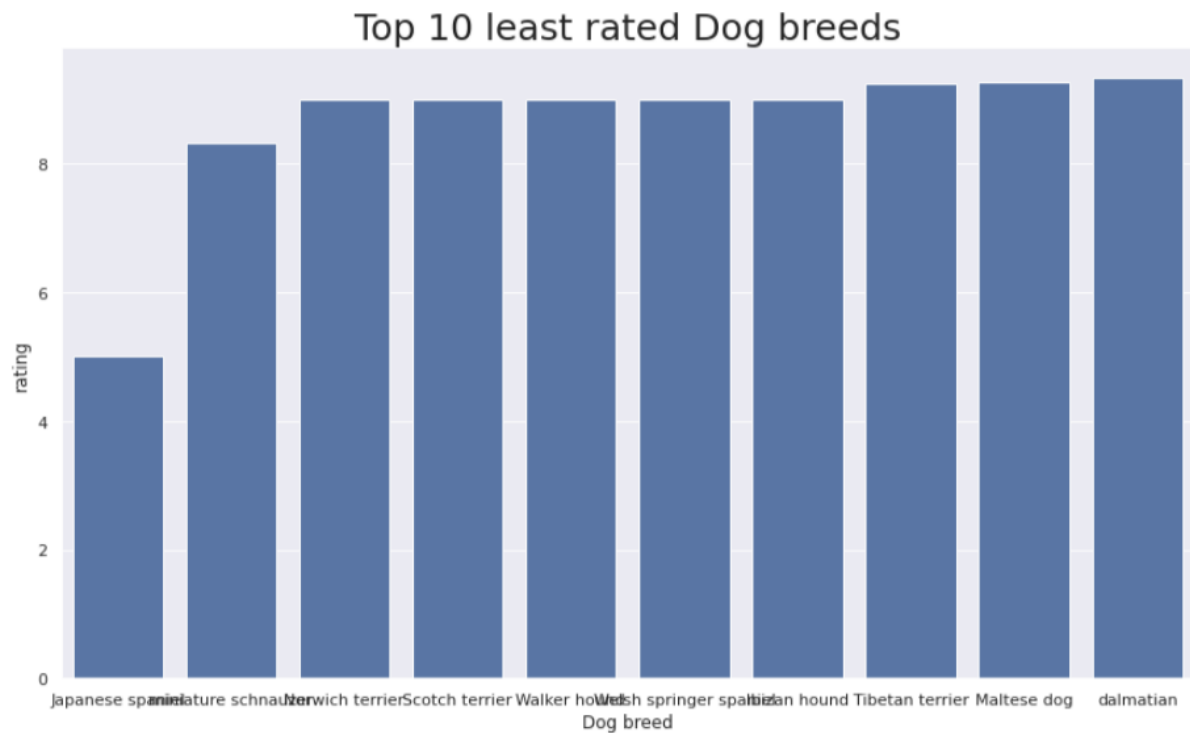
I wanted to assess the linear relationship between favorite count, retweet count and rating numerator, therefore, I have made a pairplot visualizing the relationship between each two of the variables, using seaborn's 'pairplot' function. This was the result :
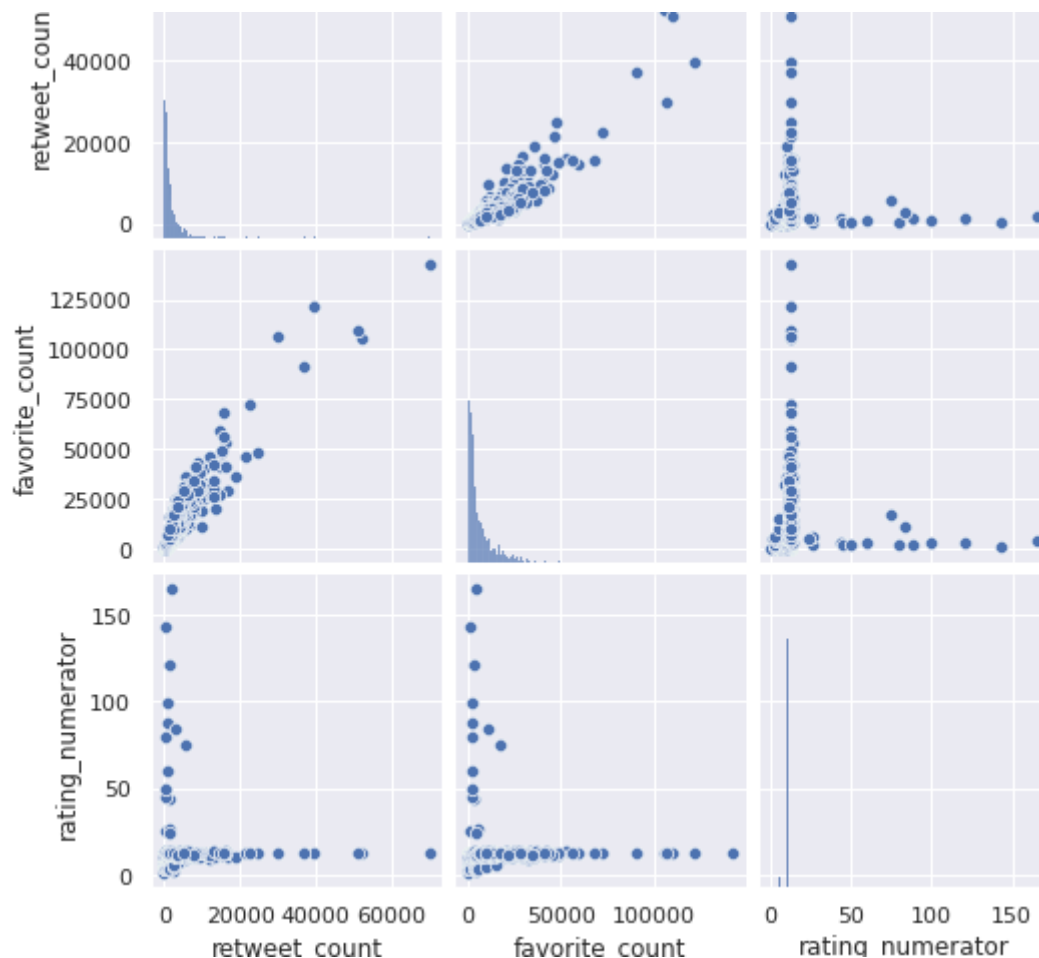
Also, I wanted to plot a figure of the correlation matrix for better understanding of the relationship between each pair of variables:

	retweet_count	favorite_count	rating_numerator
retweet_count	1.00	0.93	0.05
favorite_count	0.93	1.00	0.06
rating_numerator	0.05	0.06	1.00

We can conclude that there is a strong positive linear relationship between retweet count and favorite count, and that would help us later if we ever wanted to use a regression model. With the help of 'scipy.stats.pearsonr' function, I also calculated the correlation between these two variables : correlation = 0.928.