ACT REPORT

The data was first assessed and few issues were discovered in the rating data. The rating column had wrong few wrongly inputted records. From side research, it shows that twitter dog rating is not usually more than 15/10 or 1.5. These records were then replaced by the rating mean using the pandas .replace() function.

Act Steps

1. The pandas `groupby()` function was used to categorize the dog_stage based on the dog_rating mean values to get the first insight and to know if there are differences in the rating means based on the dog_stage.
2. To draw the second insight a scatter plot was displayed to show the relationship between favorite_count and dog_rating. This was done using the pandas `plot.scatter()` function.
3. The third insight was drawn by using the `grouby()` function to group the number of dog image number (img_num) based on the dog_rating mean value. Afterwards, a barchart was displayed for visualization.

Insights

1. It was evident that the 'duggo, puppo' stage have the highest mean. Generally speaking, dogs in the 'puppo' only, and 'doggo' only dog_stage have high ratings mean.
2. The scatter plot indicates that there is a low but positive relationship between dog_ratings and favorite counts. This might imply that high dog ratings received more favorite counts. However, this cannot be established using a scatterplot.
3. The bar chart shows that dogs with more number of images tend to have higher mean rating than those with lesser number of images. In other words, dogs that have 4 numbers of images have better ratings compared to those that have only 1 image.