ACT REPORT

INSIGHTS:

1. The output from using the **‘.value_counts()’** function shows the top **20** most common names given to dogs in the dataset. **'Charlie'** is the most common name given of the dataset, with **'11'** occurances, **'Lucy'** , **'Cooper'**, **'Oliver'**, following after with **‘10’** occurrences each, **'Penny',** and **'Tucker',** with following after with **‘9’** occurrences each. Then, **'Winston'** and **'Sadie'**, with **'8'** occurrences each. There are a couple of "**names**", **'the'** and **'a'**, with ‘**7’;** and ‘**6’** occurances each, which indicates that there are still a couple of errors from the gathering process in the dataset.

   The following output from the **‘df.groupby(.count().reset_index())’** produced a table showing the popular names per the dog breeds, with **‘>= 3’** occurances in the dataframe. **'Lucy'** is the most common name used **'Golden Retriever'** breed owners, **'Penny'** by **'Chihuahua'** breed and **'Sadie'** by **'Labrador Retriever'** breed owners.

   These analyses can provide insight into naming trends of the Twitter users based on the dataset.

2. The output of the **‘df.loc[df[‘is_dog’] == True, ‘breed_prediction’].value_counts()’** function produced a table of the counts of each breed in the dataset. **‘.head(20)’** shows the most commonly predicted dog breed in the breed_prediction column with **'Golden Retriever'** as the most commonly uploaded breed, followed by **'Labrador Retriever'**, **'Pembroke'**, **'Chihuahua'** and **'Pug'**.

   This gives some insight on which dog breeds are especially popular and most uploaded among the twitter users based on this dataset.

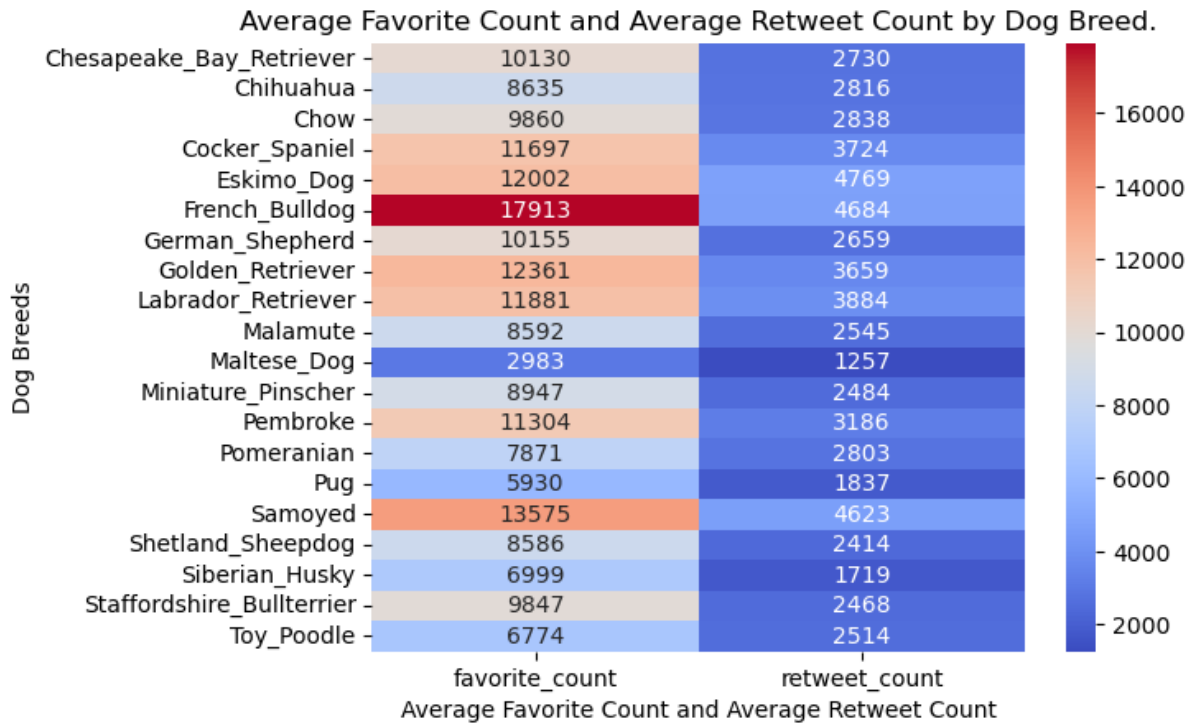
3. The **'df[''].sort_values()'** function sorted the **'rating_numerator'** column in descending order, displaying the **‘30’** highest values. This column contains the numerator ratings given to each tweet, with no typical range per se.

   Furthermore, from the **'.describe()'** function,the mean rating is **'12.22'**, which is higher than the median of **'11'**, suggesting that the distribution of the ratings are positively skewed, with some very high ratings increasing the mean. The minimum rating is **'0'**, possibly due to instances where no rating was given, and the maximum rating is an extreme value **'1776'**, which has no dog breed attached, and may have been assigned as a joke or as a way to draw attention to the tweet

   From the outputs above for the **'mean()'**, and **'.median'** functions, it indicated that the **‘Clumber’** breed had the premier of both mean and median ratings. With regards to the mean rating **'27.0'**, it indicates that tweets containing images of the **‘Clumber’** dogs were on average assigned very high scores by Twitter users. Additionally, the median rating **'27.0'** indicates that half of the tweets containing images of the breed have a score of **27** or higher.

   Additionally, the **‘Saluki’** and **‘Great Pyrenees’** breeds also appear on both tables, with the former at **‘9th’** place on the mean table, and **‘3rd’** place on the median table. The latter, however, was at **‘4th’** place on both tables.

VISUALIZATION:



The **‘seaborn’** library was imported, and used to produce the heatmap above. From the heatmap, there is an indication of a strong positive correlation between **'retweet_count'** and **'favorite_count'** per each breed. Where a particular breed is more popular, in terms of retweets, the more it is liked, in terms of favorites. Furthermore, accoring to the heatmap, the **‘French Bulldog’** has the highest average favorite count, whereas the ‘**Eskimo Dog’** has the highest average retweet count.

On the other hand, the **‘.groupby('breed_prediction')['retweet_count', 'favorite_count'].corr().iloc[0::2,-1]’** function showed the individual correlations for each breed in the subset. **‘Maltese Dog**’ breed showed the strongest correlation, although, having the lowest averages, per the heat map.

The Pearson correlation coefficient of **'0.913'** showed a strong positive correlation between these two variables, indicating that tweets with more retweets also tend to be more favored, and vice versa. This reinforces the observation from the heatmap that popularity on Twitter is measured by retweets and favorites.

The heatmap and correlation analysis provide an insightful look into the relationship between dog breeds, retweet counts and favorite counts in the dataset. Nonetheless,

it's important to remember that correlation does not necessarily imply causation, and further investigations can be done to determine the factors that also contribute to the popularity of the different dog breeds on Twitter.