Reporting: act_report

After completing the data wrangling process, we need to analyze and visualize the data.
So, firstly I have to read the dataset in which we have stored our wrangled data by using the below function:

df_wrangled=pd.read_csv('twitter_archive_master.csv')

ANALYZING DATA: We need to provide 3 insights from the wrangled data . My questions for performing analysis and it's conclusions are listed below:

1. **The dog breed with highest number of favorite tweets**

I used the below function which displays the top 5 highest favorited tweets:

df_wrangled.sort_values(by = 'favorite_count', ascending = False).head(5)

Insight-1: Lakeland_terrier received the highest number of favorite tweets.

1. **The dog breed with highest number of retweets**

I used the below function which displays the top 5 highest retweets:

df_wrangled.sort_values(by = 'retweet_count', ascending = False).head(5)

Insight-2: Labrador_retriever received the highest number of retweets.

1. **Percentage of tweets with ratings>=10**

I used the below function to arrive at the % of tweets with the desired rating:

```
total_ratings = sum(df_wrangled.rating_numerator >= 10)
percentage_ratings = (total_ratings / df_wrangled.shape[0])*100
percentage_ratings
```

Insight-3: There are 85.43% tweets having ratings >=10.

VISUALIZING DATA: We need to provide 1 visualization and I choose to plot a 1D histogram showing the distribution of ratings(i.e., rating_numerator) by plotting the number of tweets on y axis and ratings on x axis. I have used plt.axvline() to represent the mean of ratings(represented by red line shown in histogram below) and to make it more readable.
The conclusion from my visualization is that it is a left skewed distribution and high number of tweets got ratings between 10-14.