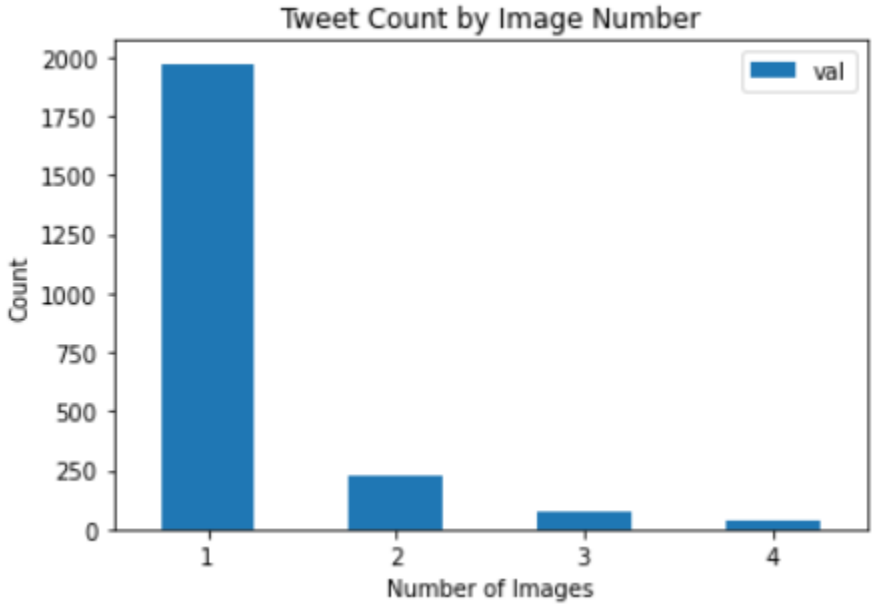
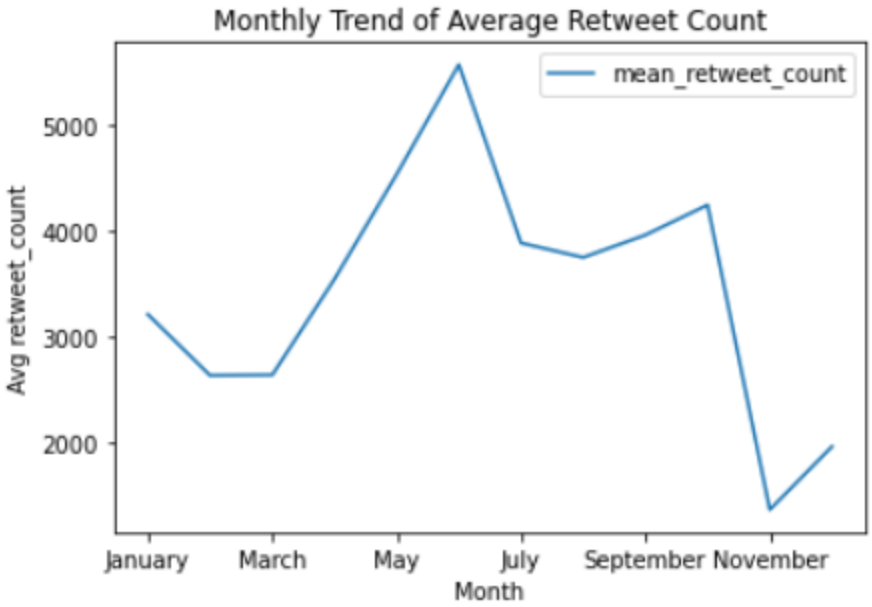
Background
After data wrangling, there is a final dataframe which combines the three datasets provided for this exercise. This documents the visualization and insights generated from the final dataset. The dataframe named ‘final’ has 2311 rows and 23 columns.

Visualization
I plotted a bar chart showing the count of tweets for each image number.
To achieve this, I used the value_counts() function to get the count for each image number and extracted the index to a variable and the value to another variable. These two variables were then used to create a dataframe. The plot.bar() function was used to create the below bar chart, adding parameters like xlabel to name the x-axis, ylabel to name the y-axis, title to give the chart a heading and rot.



This visualization shows that the majority of the tweets has one image, taking about 85% of the total tweet count.

Visualization 2
I plotted a line graph showing the monthly trend of the average retweet count.
To achieve this, I extracted the month name from the date column and stored it in a variable. I also stored the mean of the retweet_count column for each month in a variable. I created a dataframe from the two variables and used the CategoricalIndex function to order the month. The dataframe with sorted month is used to plot the line graph using plot.line() along with parameters like xlabel to name the x-axis, ylabel to name the y-axis, title to give the chart a heading.



This visualization shows that the average retweet_count is the highest in the month of June and lowest in November. There is also an increase pattern between March to June and a decrease pattern from October to November.