

**1) UNDERSTANDING:** To accomplish this, we will parse the dataset to identify the tweet with the highest number of retweets. We will track the user who authored that tweet.

**APPROACH:** We will iterate through the dataset, keeping track of the tweet with the highest retweet count. We will identify the user associated with this tweet. We can use `count()` and sort in descending order which counts the number of occurrences of each unique username in the 'username' column . We can use `sort_values(ascending=False)` which sorts the Series in descending order based on the counts. This means that usernames with the highest counts (most occurrences) will appear first. Then we'll extract the index (the usernames) of the sorted Series. Since the Series is sorted in descending order, `[0]` selects the first username from the sorted list, which corresponds to the user with the most tweets in the dataset.

**2) UNDERSTANDING:** Effectiveness can be defined based on various parameters such as retweets, time from the event (death of Queen Elizabeth II), likes, and potentially other factors. We will create a custom function that combines these factors to identify the most effective tweet.

**APPROACH:** We will create a function which takes 2 parameter one is dataset and given time period, then we will find the no. of tweets within the given time period by comparing the time at which the tweet were created which is provided to us in the given dataset with the given time period .Then we will find the username which had retweeted most number of times by counting the value of the tweets by using `count()` and sorting them in descending order by using `sort_values(ascending=False)`.

**3. UNDERSTANDING:** We will analyze the distribution of languages used in the tweets to determine the primary languages of the

users. Additionally, we will analyze the distribution of user locations (places) to identify regions where the tweets originated.

APPROACH: We use libraries like Pandas and pyplot of Matplotlib to create histograms showing the distribution of tweet languages and user locations. We use `.hist()` to plot the histogram.

4. UNDERSTANDING: We will create scatter plots and correlation analysis to understand the relationships between likes, retweets, and replies. We will interpret the visualizations to uncover insights into how these metrics are related.

APPROACH: We will use pyplot of matplotlib to create scatterplot between likes, retweets and replies. Here, We can use `cmap= 'viridis'` to use the viridis colormap to map the numbers to color where `c` which is mapped to colour is equal to the replies.

5. UNDERSTANDING: To determine the impact of videos on tweet engagement, we will compare the average number of likes for tweets with videos to those without. We will provide statistical evidence to support our conclusion.

APPROACH: We can create a new column based on whether the tweet contains a video or not. Then we will calculate the mean likes for tweets with and without videos by using `mean()` and then compare the mean likes for tweets with and without videos, if the result is yes then the tweet having videos are more likeable than tweets not having videos in them.

