**Exercise 02: Analyze tweets**

**Solve:**

1. Print the 5 most active senders

For this, we first consider the first column from the dataset. Then we counts which senders tweet more tweets, for doing maximum tweets have to active at tweeter for tweets more time. Here we consider the maximum tweet person that are more active. We use “data['Sender'].value\_counts()” for show the maximum tweeter person.

1. Print the 10 most re-tweeted tweets

Re-tweets can be determined by checking if the text of a tweet starts with “RT”. Here we use regex for determine which sentence are start off by RT. Then we counts which tweets are tweet maximum time based on this we count the maximum tweet and print most repeated tweets. We use “re\_tweets["Re-tweets"].value\_counts()” for determine the 10 most re-tweeted tweets.

1. Print the 5 most cited screen-names

Screen-names can be determined looking for words starting with “@” in the text of a tweet. Here we consider special character @, in the sentence we count those sentence, which are start off @. From the dataset it’s give nested list then we convert it from nested list to list. Then similarly we count the maximum word in the list after convert to the DataFrame and print the most 5 cited screen name.

1. Print the 10 most popular hashtags words

Hashtags words can be determined looking for words starting with “#” in the text of a tweet. For this we do the same thing like point 3.

1. Create a wordcloud with the words in the set. You will also print start and end time for each subset.

Wordclouds can be generated using the library "wordcloud". To eliminate stopwords, use the file “stopwords\_en.txt”. For this we import "wordcloud" library and use the “stopwords\_en.txt” file. Firstly, we create a function for whole subset so that we can represent "wordcloud" within less code. WordCloud mainly plot bold type word which word occur more time. Maximum bold word means maximum time the word has in the text file. We also count time for each subset during wordcloud. Also determine the execution time of each subset using import time library.