**Student Name: Sanjana Bale**

**Student Number:C0800565**

**COLLECTING DATA FROM TWITTER**

With the help of the Twitter Developer account, I retrieve tweets related to covid from the timeline. According to the Home Timeline, there were many tweets related to Covid Posted by TIMES NOW. Before retrieving the tweets, I made a project Name CBD\_Analysis\_Inclass on my Twitter developer account, from where I fetched the tokens necessary for the analysis.

**SOME NLP PREPROCESSING:**

**NLTK WORD TOKENIZE:**

NLTK has a method named word tokenize() (nominally words) for breaking strings into tokens. It divides tickets based on punctuation and white space. I have work on applying NLP methods to tweet extracted.

Len(tweet.text):

It resulted in the length of the first tweet on the timeline

**NLTK FreqDist:**

The FreqDist function in NLTK in Python calculates the frequency of words in a text. I found the frequency of words in tweet extracted from the timeline

**NLTK Bigrams, TriGrams, NGrams:**

* text\_trigrams = list(trigrams(tweet.text))
* print(text\_trigrams)

N-grams are a set of co-occurring or continuous sequences of n elements from a series of enormous text or sentences, and they are one of the essential ideas in text mining.

**NLTK STOPWORDS:**

NLTK supports stop word removal, and the list of stop words may be found in the corpus module. I tried to eliminate the stopwords from a tweet by defining the function. To eliminate stop words from a sentence, I divided the text into words and then checked to see if the word is in the NLTK list of stop words.

**RE:**

Used regular expression to compare the text by first retrieving the entire tweet, including text and numbers, and then tried to eliminate the tweet text. Also Focused on experience with the removal of few punctations and then compared to see the changes made.

1st Output: 'Lutyens' poster group' under lens: 'Arms training camp' tracked down, explosives &amp; ammunition recovered. \nFunds all… <https://t.co/Rr2ePvKL2j'>

2nd Output: 'Lutyens' poster group' under lens: 'Arms training camp' tracked down, explosives &amp; ammunition recovered. \nFunds all… '

3rd Output: 'Lutyens' poster group' under lens: 'Arms training camp’ tracked down, explosives &amp ammunition recovered. \nFunds all… <https://t.co/Rr2ePvKL2j'>