EXPERIMENT NO 2

COLLECTING & SAVING TWITTER DATA

1. DOWNLOADING THE TWITTER DATA

In [1]:
```
# Importing Libraries

import tweepy as tw

import pandas as pd

import numpy as np
```

In [2]:
```
# gathering the consumer key and secret

api_key = 'iR8RR78dKB5Cji5RikrPacFYC'
api_key_secret = '3VnG6gZRPzP4EF2bEZAeUXFblLVXZWZlJ4HYb4JuWAAK5LMNit'
access_token = '797899482566369280-EERRBmQbu2H8EqTAfUQoDJSweO8U3Kb'
access_token_secret = 'IxEx6OW5d7MtOaW8ZW57bkAlFf5t70IPoE01yldyzPwlB'
```

In [3]:
```
# searching keyword name

search_words = "#Budget2022"

# searching number of tweets

number_posts = 1000
```

In [4]:
```
# using the Oauth for accessing tweets

auth = tw.OAuthHandler(api_key, api_key_secret)
auth.set_access_token(access_token, access_token_secret)
api = tw.API(auth, wait_on_rate_limit=True)
```

In [5]:
```
# downloading the tweets

tweets = tw.Cursor(api.search_tweets, q=search_words,
                   lang='en').items(number_posts)

# list to store downloaded data

tweets_text = []

# downloading diffrent attributes

for tweet in tweets:
    tweets_text.append([
        tweet.user.screen_name, tweet.user.name, tweet.user.description,
        tweet.text, tweet.user.location, tweet.user.followers_count,
        tweet.source, tweet.created_at, tweet.user.friends_count
    ])
```

2.STORING THE DOWNLOADED DATA

In [6]:
```
# storing tweet data

tweet_data = pd.DataFrame(tweets_text,
                          columns=[
                              "User_Name", "Name", "Description (Status)",
                              "Tweets Posted", "Location", "No of Followers",
                              "Device Used for Tweet",
                              "Date of Tweet Creation", "No of Friends"
                          ])

# starting few tweet data

tweet_data
```

Out[6]:

	User_Name	Name	Description (Status)	Tweets Posted	Location	No of Followers	Device Used for Tweet	Date of Tweet Creation	No of Friends
0	AdarshPallav	Adarsh Pallav		RT @satishacharya: Blueprint for 25 years! #Bu...		7	Twitter for Android	2022-02-07 07:11:43+00:00	38
1	bsindia	Business Standard	Latest news on the economy, companies, markets...	#BSMorningShow | Will #RBI go for a hike in re...	India	2178279	TweetDeck	2022-02-07 07:11:00+00:00	430
2	bsindia	Business Standard	Latest news on the economy, companies, markets...	#BSMornigShow | Will #RBI go for a hike in rev...	India	2178279	Twitter Web App	2022-02-07 07:10:43+00:00	430
3	bsindia	Business Standard	Latest news on the economy, companies, markets...	#BSMorningShow | Business Standard's @aruproyt...	India	2178279	TweetDeck	2022-02-07 07:10:00+00:00	430
4	orfonline	ORF	Non–partisan, independent analyses on security...	Tax concessions for co-operative societies is ...	India	106444	TweetDeck	2022-02-07 07:10:00+00:00	136
...	...	...	...	...	...	...	...	...	...
995	VikramB12Singh	विक्रम बहादुर सिंह	सहेयक आचार्य (Assistant Professor)	RT @IncomeTaxIndia: Budget Highlights:\nKey Di...	जांजगीरचांपा छत्तीसगढ, भरतखण्ड	467	Twitter for Android	2022-02-06 16:54:20+00:00	2205
996	TheRationalDesi	Rational One		#Budget2022 #Congress #BJPHataoDeshBachao #UPE...		0	Twitter Web App	2022-02-06 16:54:16+00:00	58
997	VikramB12Singh	विक्रम बहादुर सिंह	सहेयक आचार्य (Assistant Professor)	RT @cbic_india: Key highlights of Union Budget...	जांजगीरचांपा छत्तीसगढ, भरतखण्ड	467	Twitter for Android	2022-02-06 16:51:46+00:00	2205
998	Vishalk23312052	विशाल कुमार शाही	भारतीय जनता युवा मोर्चा (प्रदेश विशेष आमंत्रित...	RT @blsanthosh: 400 new generation Vande Bhara...	भिलाई छ.ग	78	Twitter for Android	2022-02-06 16:51:21+00:00	6
999	Vibhans28836256	Vibhanshu Mishra	Being sarcastic and truthful are deadly and cr...	RT @thebubblebuste1: My take on Financial Budg...	Seoni, India	79	Twitter for Android	2022-02-06 16:51:16+00:00	260

1000 rows × 9 columns

In [7]:
```
# collected data information

tweet_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   User_Name               1000 non-null   object
 1   Name                    1000 non-null   object
 2   Description (Status)    1000 non-null   object
 3   Tweets Posted           1000 non-null   object
 4   Location                1000 non-null   object
 5   No of Followers         1000 non-null   int64
 6   Device Used for Tweet   1000 non-null   object
 7   Date of Tweet Creation  1000 non-null   datetime64[ns, UTC]
 8   No of Friends           1000 non-null   int64
dtypes: datetime64[ns, UTC](1), int64(2), object(6)
memory usage: 70.4+ KB
```

In [8]:
```
# saving the data as csv file

tweet_data.to_csv(
    '/DOCUMENTS/COLLEGE/CLASSES/EXPERIMENT_NO_2/tweet_data_preprocess.csv')
```
