```
!pip install -q snscrape==0.3.4
```

Importing required liabraries

```
import os
import pandas as pd
from datetime import date
```

```
today = date.today()
end_date = today
print(end_date)
```

```
     2022-03-03
```

```
#giving search term list text files
search_term = ['aap','bsp','bjp','sp','congress']
from_date = '2021-01-01'
max_results = 1000
```

Total Number of Tweets for Search Terms Extracting

Importing required liabraries

```
import time
from datetime import date
from random import uniform
start = time.time()
```

```
start = time.time()

def Extract_tweets(search_term):
  '''
    This function is to Extracting Tweets from Twitter and create 5 file of Every Element
  '''
  os.system("snscrape --format '{content!r}'"+ f" --max-results {max_results} --since {fro

for i in range(len(search_term)):
  Extract_tweets(search_term[i])

print('\n', f'Time: {time.time() - start}')
```

```
      Time: 79.18754124641418
```

```
 start = time.time()

text_files=['aap.txt','bjp.txt','bsp.txt','congress.txt','sp.txt']
```

```
def merge_text_files(text_files):
  '''
    Function to Merge all 5 Txt Extracted File
  '''
  with open('merged_file.txt', 'w') as outfile:#merge the files
    for i in text_files:
      with open(i) as infile:
        for line in infile:
          outfile.write(line)
merge_text_files(text_files)

print('\n', f'Time: {time.time() - start}')
```

```
      Time: 0.015130043029785156
```

```
import pandas as pd
df = pd.read_csv("merged_file.txt", header=None,  #read the .csv file
                 names=["Text"])  #under the header text
```

Convert Merge .txt File To merge.csv

```
df.to_csv("merged_file.csv")
```

```
#Reading the data from converted .csv file
df
```

	Text
0	'@Rizwank3315 Bhai aap ko dm karna he'
1	'@teamimrankhanpm Bismillah Hir Rahma Nir Rahi...
2	'@Shailes47245787 @Alienkumar1 @TeamKangana5 B...
3	'@mehmoodkakar604 @ShakilRai2 @YouTube AUS SE ...
4	'@ShubhiM42016011 Aap aaj aayiye thik hai'
...	...
4990	'@fffffffrd pse acho q em sp n tem n kkkk'
4991	'@laurent703611LP おつありです〜\nお腹すいた〜🦈'
4992	'@USPatriotSerena I call my #2s “taking a 45”!!!'
4993	'https://t.co/WGbYjZH0zr https://t.co/r0QhR5TVZz'
4994	'F4780FE3 :参戦ID\n参加者募集！\nLv200 リンドヴルム\nhttps:/...

4995 rows × 1 columns

Cleaning of Data

1.Remove Usernames @

```
start = time.time()
df['Text'] = df['Text'].str.replace(r'\s*@\w+', '', regex=True) #Remove username in @....
df['Text'] = df['Text'].str.replace(r'\s*\B@\w+', '', regex=True)
df['Text'] = df['Text'].str.replace(r'\s*@\S+', '', regex=True)
df['Text'] = df['Text'].str.replace(r'\s*@\S+\b', '', regex=True)
print('\n', f'Time: {time.time() - start}')
```

```
 Time: 0.08468270301818848
```

```
df.head(30)
```

	Text
0	Rizwank3315 Bhai aap ko dm karna he
1	teamimrankhanpm Bismillah Hir Rahma Nir Rahim ...
2	Shailes47245787 Alienkumar1 TeamKangana5 Bhai ...
3	mehmoodkakar604 ShakilRai2 YouTube AUS SE BARR...
4	ShubhiM42016011 Aap aaj aayiye thik hai
5	ajaydevgn DisneyPlusHS sir vimal khana kab cho...
6	Kitne hurdles aae lekin apna kaam karte rahe A...

2.Remove Punctuation

8 AamAadmiParty You are good leader

```
start = time.time()
punctuation = '''!()-[]{};:'"\,<>./?@#$%^&*_~'''
# Removing punctuations in string
# Using loop + punctuation string
def remov_punct(text):
      for i in text:
          if i in punctuation:
                text = text.replace(i, "")
      return text
print('\n', f'Time: {time.time() - start}')
```

```
      Time: 0.00010323524475097656
```

17 ArvindKejriwal raghavchadha msisodia AamAadmiP...

```
df['Text']=df['Text'].apply(remov_punct)
```

19 Sangeet51434930 abhishek3588 AadeshRawal Aap b...

```
df.head(20)
```

	Text
0	Rizwank3315 Bhai aap ko dm karna he
1	teamimrankhanpm Bismillah Hir Rahma Nir Rahim ...
2	Shailes47245787 Alienkumar1 TeamKangana5 Bhai ...
3	mehmoodkakar604 ShakilRai2 YouTube AUS SE BARR...
4	ShubhiM42016011 Aap aaj aayiye thik hai
5	ajaydevgn DisneyPlusHS sir vimal khana kab cho...
6	Kitne hurdles aae lekin apna kaam karte rahe A...
7	Aap yadi kisi aesi cheez par kabja karke baith...
8	AamAadmiParty You are good leader
9	RajatSharmaLive ap bhi apni reporting mein iss

3.Remove Emojis

```
start = time.time()
df = df.astype(str).apply(lambda x: x.str.encode('ascii', 'ignore').str.decode('ascii'))
print('\n', f'Time: {time.time() - start}')
```

```
 Time: 0.013239860534667969
```

15 Aap ki baar modi charkar httpstcoXhLT7gFuch

```
df.head(50)
```

```
18    HeijnekampJaap Dit is gewoon een samenvatting ...
19    Sangeet51434930 abhishek3588 AadeshRawal Aap b...
20    AAgharkar Iqrasaysthat Meri wajah se aap logon...
21          Parveen40918171 Bilkul sahi Pharmaya aap ne
22    Main to corrt pol se yehi kahunga ki aap apne ...
23                SubhasiniMaurya Sahi kha rahi hai aap
24                                                     
25                                 AamAadmiParty Salute
26                           SyyedSuhail Aree Iska data
27    RJDforIndia yadavtejashwi RJD apna time yaad k...
28    RockingNainaa accha ji agar koi raji ho jaye t...
29    DharendraDr DeependerSHooda KirenRijiju Gajab ...
30                             stichtingaap ZEMBLA Ziek
31                             AamAadmiParty httpstc...
32    AAPnnNew Insider Filing on ADVANCE AUTO PARTS ...
33    Paytm Very bad service no support i called man...
34    RadhikaKhera Wah WahMere Dil ki baat likh di a...
35    ImSatyayadav Gobar ko khaane se aur mooth pina...
36    DelhiPolice dtptraffic AamAadmiParty PIBHomeAf...
37                  kyabataye iMemeStore Admins aur aap
38    Zohi786 HamzyKhanx iamqadirkhawaja In SHA ALLA...
39    DiyaGhosh ParmarVeera baesakhiiz guys ye kya b...
40    MessiTheKing19 lakshayhere Ok bro toh aap ka n...
41    AribaShahid FootballPak BabarEnthusiast Aap lo...
42    AAgharkar Iqrasaysthat Ye Kab ka description d...
43    Around 500000 people across NSW have been orde...
44                                      AamAadmiParty 1
45        ReenaRoy8587 Reena ji aap kya job karte ho ji
46    fearlessssoul Madam chay me kefin hota hai jo ...
47    reply Aap bhi Scenes outsides a Bandra restaur...
48                 DrUditraj Deshhit me ek kaam kiya he
49    TeamKangana5 Agar aap negative comments padh k...
```

```
df.tail(20)
```

	Text
4975	excited nako maka kuhag sp huhuhu
4976	cirogomes E a ideia de pagar a dvida do carto ...
4977	nishisp
4978	awawasp
4979	isislimax Akakakakakakak mas eu ganhei a conqu...
4980	iiyokinisinaide nnnspu200dnnnnnnn
4981	BIGBANG1nnn
4982	Timberflakes Whodat4lyfe1 spsalas StanLew24950...
4983	mi compaera me regal una pulcera uwu
4984	imrrkt GitanjaliMoha20 SP
4985	Video | 020322 4K WONHO Eye On You I Sho...
4986	jinoias MAS O GOSTO EH HORRIVEL
4987	ZDROWCO TSzulc TashWitk ireneuszkulesza Kazimi...
4988	AnaliseVerdao mariapmihok homenagem
4989	Another 20 tweet Zacktivation I vote for Zack ...
4990	fffffffrd pse acho q em sp n tem n kkkk
4991	laurent703611LP n
4992	USPatriotSerena I call my 2s taking a 45
4993	httpstcoWGbYjZH0zr httpstcor0QhR5TVZz
4994	F4780FE3 IDnnLv200 nhttpstcoL0MPuOHPQj

4.Remove Null Rows data

```
df['Text'] = df['Text'].str.lower() #Lower all the data
```

```
df.head() #Display top 5
```

	Text
0	rizwank3315 bhai aap ko dm karna he
1	teamimrankhanpm bismillah hir rahma nir rahim ...
2	shailes47245787 alienkumar1 teamkangana5 bhai ...
3	mehmoodkakar604 shakilrai2 youtube aus se barr...

```
df.dropna() #Drop the NUll values
```

	Text
0	rizwank3315 bhai aap ko dm karna he
1	teamimrankhanpm bismillah hir rahma nir rahim ...
2	shailes47245787 alienkumar1 teamkangana5 bhai ...
3	mehmoodkakar604 shakilrai2 youtube aus se barr...
4	shubhim42016011 aap aaj aayiye thik hai
...	...
4990	ffffffffrd pse acho q em sp n tem n kkkk
4991	laurent703611lp n
4992	uspatriotserena i call my 2s taking a 45
4993	httpstcowgbyjzh0zr httpstcor0qhr5tvzz
4994	f4780fe3 idnnlv200 nhttpstcol0mpuohpqj

4995 rows × 1 columns

```
df.isnull().sum()  #Return Sum of NULL Values if exists
```

```
Text    0
dtype: int64
```

```
df.info()  #Extract the information
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4995 entries, 0 to 4994
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   Text    4995 non-null   object
dtypes: object(1)
memory usage: 39.1+ KB
```

```
df.describe() #Describe the dataset
```

	Text
count	4995
unique	4763
top	

Give label to the data

```
start = time.time()
df['label'] = df.Text.str.extract('(aap|bsp|bjp|sp|congress)') # if list aap|bsp|bjp|sp|co
df['label'] = df['label'].replace(['aap','bsp','bjp','sp','congress'],'1') # replaced all
df['label'] = df['label'].fillna(0) #fill 0 remaining all
print('\n', f'Time: {time.time() - start}')
```

```
 Time: 0.018356800079345703
```

```
df.head(40)
```

8	aamaadmiparty you are good leader	0
9	rajatsharmalive ap bhi apni reporting mein iss...	0
10	aryan11626940 aap ki din mangalmay ho	1
11	yadavakhilesh tonti bhaiya didi ne bengal me h...	1
12	brickmack cries in aap	1
13	ptshekhardixit pmoindia aap bhi indian ho cab ...	1
14	aryan11626940 aap ki baat me dam hai dear	1
15	aap ki baar modi charkar httpstcoxnly7gfuch	1
16	manishakimball svidyasagar arre aap hamesha it...	1
17	arvindkejriwal raghavchadha msisodia aamaadmip...	0
18	heijnekampjaap dit is gewoon een samenvatting ...	1
19	sangeet51434930 abhishek3588 aadeshrawal aap b...	1
20	aagharkar iqrasaysthat meri wajah se aap logon...	1
21	parveen40918171 bilkul sahi pharmaya aap ne	1
22	main to corrt pol se yehi kahunga ki aap apne ...	1
23	subhasinimaurya sahi kha rahi hai aap	1
24		0
25	aamaadmiparty salute	0
26	syyedsuhail aree iska data	0
27	rjdforindia yadavtejashwi rjd apna time yaad k...	1
28	rockingnainaa accha ji agar koi raji ho jaye t...	1
29	dharendradr deependershooda kirenriiiju gajab ...	1

```
df['label'].isnull().sum()
```

```
0
```

30 aonnnnew insider filing on advance auto parts ... 1

count the data i.e values of 0 and 1 in the dataset

34 radhikakhera wah wahmere dil ki baat likh di a... 1

```
(df['label'] == 0).sum()  #0 labled data
```

```
1623
```

37 kyabataye imemestore admins aur aap 1

```
(df['label'] == '1').sum()  #1 labled data
```

```
3372
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4995 entries, 0 to 4994
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   Text    4995 non-null   object
 1   label   4995 non-null   object
dtypes: object(2)
memory usage: 78.2+ KB
```

```
df
```

	Text	label
0	rizwank3315 bhai aap ko dm karna he	1
1	teamimrankhanpm bismillah hir rahma nir rahim ...	1
2	shailes47245787 alienkumar1 teamkangana5 bhai ...	1
3	mehmoodkakar604 shakilrai2 youtube aus se barr...	1
4	shubhim42016011 aap aaj aayiye thik hai	1
...	...	...
4990	fffffffrd pse acho q em sp n tem n kkkk	1
4991	laurent703611lp n	0
4992	uspatriotserena i call my 2s taking a 45	1
4993	httpstcowgbyjzh0zr httpstcor0qhr5tvzz	0
4994	f4780fe3 idnnlv200 nhttpstcol0mpuohpqj	0

4995 rows × 2 columns

```
df.describe()
```

	Text	label
count	4995	4995
unique	4763	2
top		1
freq	51	3372