

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

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

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

## ▼ Search List

```
search_term = ['aap', 'bsp', 'bjp', 'sp', 'congress']
from_date = '2022-02-24'
```

## ▼ Find the number of tweets

```
for i in search_term:
    os.system(f"snsrape --since {from_date} twitter-search '{i} until:{end_date}' >{i}.txt")
    if os.stat(f"{i}.txt").st_size == 0:
        counter = 0
    else:
        df = pd.read_csv(f'{i}.txt', names=['link'])
        counter = df.size

print(f'Number Of Tweets {i}: ' + str(counter))

Number Of Tweets aap: 668
Number Of Tweets bsp: 743
Number Of Tweets bjp: 669
Number Of Tweets sp: 331
Number Of Tweets congress: 200

max_results = 100
```

## ▼ Extract Data from tweets

```
for i in search_term:
    extracted_tweets = "snsrape --format '{content!r}'"+ f" --max-results {max_results} --since {from_date}"
    os.system(extracted_tweets)
    if os.stat(f"extracted-tweets{i}.txt").st_size == 0:
        print('No Tweets found')
    else:
        df = pd.read_csv(f'extracted-tweets{i}.txt', names=['content'])
        for row in df['content'].iteritems():
```

print(row)

```
(44, "@RepGosar You truly are worthless in Congress and I believe you are instrumental in b  
(45, "'@sobifty @rssurjewala @INCIndia Sahi kha'")  
  
(46, "@Erwin_Joe And this comes with a valuable lesson: You too can be elected to Congress  
(47, "'Lawyers ask US justices not to block new NC Congress map #election2022 #ncpol https:  
\(48, "I'm supportive of having a minimum AND MAXIMUM age for the president and any member o  
\(49, "@B52Malmst @joncoopertweets For someone who lives in Colorado but not in Lauren Boebe  
\(50, "'@tedcruz @SenSasse @SenMikeLee @SenateGOP can someone explain why Congress can't joi  
\(51, "'@gggreenwald @matthewstoller Imagine thinking the Chinese regime "representing" the C  
\(52, "'@PhilArballo Lauren Boebert and MTG are a disgrace to Congress. #RepublicansHateAmer  
\(53, "'Thanks for sharing the struggles of affordable housing opportunities in Huntington"\)  
\(54, "'@jeffttimmer These women need to be held responsible for their horrible behavior! The  
\(55, "@AnyaSonder There's a USA Congress candidate who has written a really good article, s  
\(56, "'@sawhill @DeAngelisCorey let it just be the start. If I am elected to Congress I vow  
\(57, "'साक्षात्कार : जेपी नड्डा बोले- विपक्ष 'घातक गठजोड़' बना पहले चुनाव लड़ता है...फिर हारने पर आपस में ल  
\(58, "'@80sretro215 @RepMTG @RepBoebert @GOP Oh wow...she actually got all the spelling right  
\(59, "@AdamKinzinger Classic deflection for someone who isn't qualified to serve biscuits t  
\(60, "'It's CPOC - Congressional Parasites of Congress. https://t.co/nKECdda000'"\)  
\(61, "Yup. Hire for ingenuity people. #COVID cracked old school structures, #Ukraine has b  
\(62, "'Can you pitch in today to help send me back to Congress to keep fighting for Arizona  
\(63, "'Over the last few years"\)  
\(64, "'@DrSJaishankar Gr8 job. India is proud of our MEA\\nOpposition Plz shut up! \\n@Rahu  
\(65, "'I'm sorry"\)  
\(66, "'#ExpelLaurenBoebert #ExpelMarjorieTaylorGreene Congress has expelled Reps. for much  
\(67, "@Dragone30031440 @TheZeebs @BadLegalTakes Get it? I'm a liberal, and I am in 100% ag  
\(68, "'@Jim\_Jordan You are either an idiot or a liar"\)  
\(69, "Congress. That's why I need your help. Can \\nyou chip in whatever you can to help gr  
\(70, "'@Angry\_Staffer Who"\)  
\(71, "'@americai Sir"\)  
\(72, "'@laurenboebert You're an a\$\$\$. An embarrassment. A joke. A misfit. You do not belong  
\(73, "@KariLake @ScottPresler As passionate and hard working as he is, Scott should totally  
\(74, "'Ripple Proposes Collab With Congress / Joe Lubin Hot Water / ODL Growth https://t.co  
\\(75, "'@thehill Well hill"\\)  
\\(76, "'Oh wow! Someone in Congress is actually thinking! https://t.co/ajdIwDqLMH'"\\)  
\\(77, "'@DineshDSouza Repubs will only take back senate and congress if they find a way to s  
\\(78, "'@AbiBagabaldo Kung nasasakupan mo lang ako"\\)  
\\(79, "'@passivec @izquierdo\\_isis @YepItsMilfy @cav\\_pac @delysmig14 @JoubertSonya @JohnHayes  
\\(80, "@eegarim @atrubar I look at them as an example of equality. For years we've put men i  
\\(81, "'Will someone flush this #Sewage! How about it @GOPLLeader. Or do you condone this typ  
\\(82, "'@laurenboebert You're a disrespectful person! Calling out a president during his sta  
\\(83, "STOP 🙌 GIVING 🙌 ATTENTION 🙌 TO THESE--oh fuck it, lets' just get to the good part  
\\(84, "'@RahulGandhi He has to have figures to understand.He gets confused about strategy is  
\\(85, "@alex\\_marie\\_z @BetteMidler How dumb are these two?\\nThey're so dumb, they can't tell  
\\(86, "'Nuevos teléfonos"\\)  
\\(87, "'@New\\_Narrative @MalcolmNance CNP"\\)  
\\(88, "@benshapiro Arguably the ODDEST thing I've ever seen from a member of Congress"\\)  
\\(89, "'@JamesTate121 @curtisc5105 What a great combination of photos. If only we could get  
\\(90, "'@nathaliejacoby1 And removed from congress"\\)  
\\(91, "'@DonJBacon Do you have any comments on the pariah members of Congress who voted agai  
\\(92, "'@JWParks\\_ America deserves better than this Representative in Congress.'"\\)  
\\(93, "'Biden seems to have pretty much unanimous support from Congress as he goes after Put  
\\(94, "'@razzli\\_ @auntmaggiep Damn autocorrect it should be Ass Clowns in Congress!'""\\)  
\\(95, "'@MysterySolvent Or should I have said the "members" of congress🤔🤔🤔🤔'""\\)  
\\(96, "@AdamKinzinger If you don't understand that caring about the Ukranian border more tha  
\\(97, "'@laurenboebert As a member of Congress \\(unfortunate for Congress\\) you should learn a  
\\(98, "'@TeamTurner2022 @ninaturner Nina Your Voice Is Needed In Congress.\\nYou Are A Rare  
\\(99, "@Dragone30031440 @TheZeebs @BadLegalTakes You seem extremely confused about how this
```

```

filenames = ["extracted-tweetsaap.txt", "extracted-tweetsbjp.txt", "extracted-tweetsbsp.txt","extra

with open("output_file.txt", "w") as outfile:
    for filename in filenames:
        with open(filename) as infile:
            contents = infile.read()
            outfile.write(contents)


```

### ▼ Converted into CSV with column name

```

df = pd.read_csv('output_file.txt', sep='delimiter', header=None, engine='python', index_col=False,n
df.to_csv('output_file.csv')
df.head()

```

	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'

### ▼ Cleaning Data


#### ▼ Remove username & punctuations

```

def remove_username_no_punctuation(test_str):
    test_str = re.sub('@[^\s]+',' ',test_str)
    for char in test_str:
        if char in string.punctuation:
            test_str = test_str.replace(char, "")
    return test_str

df['text'] = df['text'].apply(remove_username_no_punctuation)
df.head()

```

	text 
0	Bhai aap ko dm karna he
1	Bismillah Hir Rahma Nir Rahim MashaAllah Mash...
2	Bhai aap jaise logo ko sirf kamiya hi dikht...
3	AUS SE BARRA NAWAZ AKA BAKWAAS SHARIF FUNDI...
4	Aap aaj aayiye thik hai

▼ Remove Emojis

```
def remove_emoji(test_str):
    regex_pattern = re.compile(pattern = "["
        u"\U0001F600-\U0001F64F"  # emoticons
        u"\U0001F300-\U0001F5FF"  # symbols & pictographs
        u"\U0001F680-\U0001F6FF"  # transport & map symbols
        u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
        u"\U00002500-\U00002BEF"  # chinese char
        u"\U00002702-\U000027B0"
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
        u"\U0001f926-\U0001f937"
        u"\U00010000-\U0010ffff"
        u"\u2640-\u2642"
        u"\u2600-\u2B55"
        u"\u200d"
        u"\u23cf"
        u"\u23e9"
        u"\u231a"
        u"\ufe0f"  # dingbats
        u"\u3030"
        "]" +, flags = re.UNICODE)
    return regex_pattern.sub(r'',test_str)

df['text'] = df['text'].apply(remove_emoji)
df.head()
```

	text
0	Bhai aap ko dm karna he
1	Bismillah Hir Rahma Nir Rahim MashaAllah Mash...
2	Bhai aap jaise logo ko sirf kamiya hi dikht...
3	AUS SE BARRA NAWAZ AKA BAKWAAS SHARIF FUNDI...
4	Aap aaj aayiye thik hai

▼ Find the null values

```
df.isnull().sum()

text      0
dtype: int64
```

▼ Labeling the data in search term data

```
def labeling(new_string):
    for char in new_string.lower().split():
```

```
        if char in search_term:
            return 1
    else:
        return 0
```

```
df['Label'] = df['text'].apply(labeling)
df.to_csv('output_file.csv')
```

## ▼ Count the values

```
df['Label'].value_counts()
```

```
1    277
0    223
Name: Label, dtype: int64
```

```
df.to_csv('output_file.csv')
```