

```
# API Key=ArSqGZOxJVnDPhM43YPLNVvvv
# Api key secret = 08EPsPa3LX2XrJLQORlHNxyHDSIk5rpqWCON5pfh3LNVuMDJUD
# Bearer Token = AAAAAAAAAAAAAAAAAAAEGMngEAAAAAUWLVqOPXGCeX%2F4qKeWTgegQJGv0%3DdFdQx5zv7w7k2URn48bF52y34VKZUq379MKR6nZjpp
# Access_token = 1065218660917264384-7U7cWCDC9lKm3yVtPFqRHwwVPqElqO
# Acces_token_secret= tIDPBxOtJGzq5xRs0Lak2IJFU4DercGqv9U3r2NyzuEK9
# client_id = d3A4NG5kV0xFblRXU25ZM2M2Ums6MTpjaQ
# client_secret = Wg3ih3xPK4E-L0e7bDu4B-L2BpV-RxDvQOkow7vCgnSlXh8e3l
!pip3 install git+https://github.com/JustAnotherArchivist/snsrape.git
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting git+https://github.com/JustAnotherArchivist/snsrape.git
  Cloning https://github.com/JustAnotherArchivist/snsrape.git to /tmp/pip-req-build-x3xkqv5e
  Running command git clone --filter=blob:none --quiet https://github.com/JustAnotherArchivist/snsrape.git /tmp/pip-
  Resolved https://github.com/JustAnotherArchivist/snsrape.git to commit 786815dd05681e2421cd03aa9acf5ab5c773bce9
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: requests[socks] in /usr/local/lib/python3.10/dist-packages (from snsrape==0.6.2.20230
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from snsrape==0.6.2.20230321.dev13+g
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from snsrape==0.6.2.202303
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from snsrape==0.6.2.20230321.dev
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->snsrap
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests[socks])
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests[socks]->s
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests[so
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests[socks]->snsrap
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.10/dist-packages (from requests[socks
```

```
import pandas as pd
from tqdm.notebook import tqdm
import snsrape.modules.twitter as sntwitter
```

```
# # Pulling the tweets
# scraper = sntwitter.TwitterSearchScraper("#ArtificialIntelligence")
```

```
import subprocess

# Define the hashtag you want to scrape
hashtag = "ArtificialIntelligence"

# Define the maximum number of results
max_results = 100

# Build the snsrape command
command = f"snsrape --max-results {max_results} twitter-hashtag {hashtag}"

# Execute the command and capture the output
output = subprocess.check_output(command, shell=True, text=True)

# Print the output
print(output)
```

[illegible]

```
!snscape --max-results 100000 --with-entity --jsonl twitter-hashtag ArtificialIntelligence >> output2.json
!snscape --max-results 100000 --with-entity --jsonl twitter-hashtag ArtificialIntelligence >> output2.json
```

```
    return self._sock.recv_into(b)
File "/usr/lib/python3.10/ssl.py", line 1274, in recv_into
    return self.read(nbytes, buffer)
File "/usr/lib/python3.10/ssl.py", line 1130, in read
    return self._sslobj.read(len, buffer)
KeyboardInterrupt
^C
2023-05-21 01:32:49.530 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-162491234'
2023-05-21 01:32:49.532 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-165811489'
Traceback (most recent call last):
  File "/usr/local/bin/snscape", line 8, in <module>
    sys.exit(main())
  File "/usr/local/lib/python3.10/dist-packages/snscape/_cli.py", line 320, in main
    for i, item in enumerate(scraper.get_items(), start = 1):
  File "/usr/local/lib/python3.10/dist-packages/snscape/modules/twitter.py", line 1704, in get_items
    for obj in self._iter_api_data('https://twitter.com/i/api/graphql/7jT5GT59P8IFjgxwqnEdOw/SearchTimeline', _Twitte
  File "/usr/local/lib/python3.10/dist-packages/snscape/modules/twitter.py", line 812, in _iter_api_data
    obj = self._get_api_data(endpoint, apiType, reqParams, instructionsPath = instructionsPath)
  File "/usr/local/lib/python3.10/dist-packages/snscape/modules/twitter.py", line 768, in _get_api_data
    r = self._get(endpoint, params = params, headers = self._apiHeaders, responseOkCallback = self._check_api_responses
  File "/usr/local/lib/python3.10/dist-packages/snscape/base.py", line 266, in _get
    return self._request('GET', *args, **kwargs)
  File "/usr/local/lib/python3.10/dist-packages/snscape/base.py", line 218, in _request
    r = self._session.send(req, allow_redirects = allowRedirects, timeout = timeout, **environmentSettings)
  File "/usr/local/lib/python3.10/dist-packages/requests/sessions.py", line 645, in send
    r = adapter.send(request, **kwargs)
  File "/usr/local/lib/python3.10/dist-packages/requests/adapters.py", line 440, in send
    resp = conn.urlopen(
  File "/usr/local/lib/python3.10/dist-packages/urllib3/connectionpool.py", line 703, in urlopen
    httplib_response = self._make_request(
  File "/usr/local/lib/python3.10/dist-packages/urllib3/connectionpool.py", line 449, in _make_request
    six.raise_from(e, None)
  File "<string>", line 3, in raise_from
  File "/usr/local/lib/python3.10/dist-packages/urllib3/connectionpool.py", line 444, in _make_request
    httplib_response = conn.getresponse()
  File "/usr/lib/python3.10/http/client.py", line 1375, in getresponse
    response.begin()
  File "/usr/lib/python3.10/http/client.py", line 318, in begin
    version, status, reason = self._read_status()
  File "/usr/lib/python3.10/http/client.py", line 279, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
  File "/usr/lib/python3.10/socket.py", line 705, in readinto
    return self._sock.recv_into(b)
  File "/usr/lib/python3.10/ssl.py", line 1274, in recv_into
    return self.read(nbytes, buffer)
  File "/usr/lib/python3.10/ssl.py", line 1130, in read
    return self._sslobj.read(len, buffer)
KeyboardInterrupt
^C
2023-05-21 01:32:50.887 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-165421280'
2023-05-21 01:32:50.892 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-162491234'
2023-05-21 01:32:51.385 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-162491234'
2023-05-21 01:32:51.387 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-165811489'
2023-05-21 01:32:51.388 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-163968863'
2023-05-21 01:32:51.848 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-165421280'
2023-05-21 01:32:51.850 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-162491234'
2023-05-21 01:32:51.853 WARNING snscape.modules.twitter Skipping unrecognised entry ID: 'promoted-tweet-165811489'
```

```
import pandas as pd
import json

# Read the JSON file into a pandas DataFrame
with open('output2.json') as file:
    data = [json.loads(line) for line in file]
df = pd.DataFrame(data)

# Print the column names
print(df.columns)
```

```
Index(['_type', 'url', 'date', 'rawContent', 'renderedContent', 'id', 'user',
      'replyCount', 'retweetCount', 'likeCount', 'quoteCount',
      'conversationId', 'lang', 'source', 'sourceUrl', 'sourceLabel', 'links',
      'media', 'retweetedTweet', 'quotedTweet', 'inReplyToTweetId',
```

```
'inReplyToUser', 'mentionedUsers', 'coordinates', 'place', 'hashtags',
'cashtags', 'card', 'viewCount', 'vibe', 'bookmarkCount', 'content',
'outlinks', 'outlinksss', 'tcooutlinks', 'tcooutlinksss', 'username'],
dtype='object')
```

```
df.head()
```

	_type	url	date	rawContent	renderedContent	
0	snscape.modules.twitter.Tweet	https://twitter.com/Spinneurfou/status/1660091...	2023-05-21T01:15:21+00:00	1/10: Are you familiar with SEI Network? This ...	1/10: Are you familiar with SEI Network? This ...	166
1	snscape.modules.twitter.Tweet	https://twitter.com/ExMuslimsMN/status/1660091...	2023-05-21T01:15:12+00:00	Does ChatGPT4 understand Winograd Schemas? \n\n...	Does ChatGPT4 understand Winograd Schemas?\n\n...	166
2	snscape.modules.twitter.Tweet	https://twitter.com/Salvoly/status/16600917966...	2023-05-21T01:15:01+00:00	Mia Khalifa's Thoughts On Artificial Intellige...	Mia Khalifa's Thoughts On Artificial Intellige...	166
3	snscape.modules.twitter.Tweet	https://twitter.com/marsh5300/status/166009171...	2023-05-21T01:14:42+00:00	@hamoncutta L	@hamoncutta L	166
4	snscape.modules.twitter.Tweet	https://twitter.com/TWEETPRESS/status/1660091...	2023-05-21T01:14:19+00:00	Can artificial intelligence write content for ...	Can artificial intelligence write content for ...	166

5 rows × 37 columns



```
import pandas as pd
import json

# Read the JSON file into a pandas DataFrame
with open('output2.json') as file:
    data = [json.loads(line) for line in file]
df = pd.DataFrame(data)

# Extract the desired columns
df_processed = df[['date', 'user', 'content', 'id', 'replyCount', 'retweetCount', 'likeCount', 'quoteCount', 'hashtags', '

# Perform data preprocessing on the 'content' column
df_processed['content'] = df_processed['content'].str.strip()

# Print the processed DataFrame
print(df_processed.head())
```

	date \		user \		content	id \
0	2023-05-21T01:15:21+00:00				1/10: Are you familiar with SEI Network? This ...	1660091880439181318
1	2023-05-21T01:15:12+00:00				Does ChatGPT4 understand Winograd Schemas?\n\n...	1660091839733547009
2	2023-05-21T01:15:01+00:00					
3	2023-05-21T01:14:42+00:00					
4	2023-05-21T01:14:19+00:00					

```
2 Mia Khalifa's Thoughts On Artificial Intellige... 1660091796641161217
3 @hamoncutta L 1660091716647661568
4 Can artificial intelligence write content for ... 1660091618655981568

replyCount  retweetCount  likeCount  quoteCount  hashtags  cashtags  \
0          1             0           0           0  [SEINetwork, AI]  None
1          0             0           0           0          None  None
2          0             0           0           0          None  None
3          0             0           0           0          None  None
4          0             0           0           0  [ContentCreat]  None

mentionedUsers  retweetedTweet  \
0          None          None
1          None          None
2          None          None
3  [{'_type': 'snsrape.modules.twitter.User', 'u...  None
4          None          None

quotedTweet  lang  links  \
0          None  en          None
1          None  en          None
2          None  en          None
3          None  und          None
4          None  en  [{'_type': 'snsrape.modules.twitter.TextLink'...
```

media

```
0          None
1  [{'_type': 'snsrape.modules.twitter.Photo', '...
2  [{'_type': 'snsrape.modules.twitter.Video', '...
3          None
4  [{'_type': 'snsrape.modules.twitter.Photo', '...
<ipython-input-3-e8e44ec1da36>:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning
df_processed['content'] = df_processed['content'].str.strip()
```

df\_processed.head()

	date	user	content	id	replyCount	retweetCount	likeCount	quote
0	2023-05-21T01:15:21+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	1/10: Are you familiar with SEI Network? This ...	1660091880439181318	1	0	0	
1	2023-05-21T01:15:12+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Does ChatGPT4 understand Winograd Schemas? \n\n...	1660091839733547009	0	0	0	
2	2023-05-21T01:15:01+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Mia Khalifa's Thoughts On Artificial Intellige...	1660091796641161217	0	0	0	
3	2023-05-21T01:14:42+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	@hamoncutta L	1660091716647661568	0	0	0	
4	2023-05-21T01:14:19+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Can artificial intelligence write content for ...	1660091618655981568	0	0	0	



```
column_data_types = df_processed.dtypes

print(column_data_types)
```

```
date           object
user           object
content        object
id             int64
replyCount     int64
retweetCount   int64
likeCount      int64
quoteCount     int64
hashtags       object
cashtags       object
mentionedUsers object
retweetedTweet object
quotedTweet    object
lang           object
links          object
media          object
dtype: object
```

```
df_processed['date'] = pd.to_datetime(df_processed['date'])
```

<ipython-input-6-dabd8c94808f>:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning)  
df\_processed['date'] = pd.to\_datetime(df\_processed['date'])

```
df_processed['date'].count()
```

```
22965
```

```
df_processed.count()
```

```
date           22965
user           22965
content        22965
id             22965
replyCount     22965
retweetCount   22965
likeCount      22965
quoteCount     22965
hashtags       12710
cashtags       382
mentionedUsers 6465
retweetedTweet 0
quotedTweet    992
lang           22965
links          12118
media          7982
dtype: int64
```

```
df_processed.head(100)
```

```
AI_count = df_processed[df_processed['hashtags'].apply(lambda x: x is not None and 'ArtificialIntelligence' in x)].shape[0]

print("Count of tweets with the hashtag #ArtificialIntelligence:", AI_count)
```

```
Count of tweets with the hashtag #ArtificialIntelligence: 6241
```

```
df_processed.head(100)
```

```
AI_count = df_processed[df_processed['content'].apply(lambda x: x is not None and 'ArtificialIntelligence' in x)].shape[0]
```

```
print("Count of tweets with the hashtag #ArtificialIntelligence:", AI_count)
```

Count of tweets with the hashtag #ArtificialIntelligence: 6248

```
df_processed.head(100)
# Filter the DataFrame based on the condition
df_filtered = df_processed[df_processed['hashtags'].apply(lambda x: x is not None and 'ArtificialIntelligence' in x)]

# Print the count of tweets with the hashtag '#ArtificialIntelligence'
print("Count of tweets with the hashtag #ArtificialIntelligence:", df_filtered)

# Continue with further analysis using the filtered DataFrame (df_filtered)

df_filtered.head()
```

```

Count of tweets with the hashtag #ArtificialIntelligence:
5      2023-05-21 01:14:02+00:00
19     2023-05-21 01:06:10+00:00
30     2023-05-21 01:02:38+00:00
33     2023-05-21 01:01:27+00:00
34     2023-05-21 01:00:48+00:00
...
22956  2023-05-20 23:11:15+00:00
22959  2023-05-20 23:09:42+00:00
22960  2023-05-20 23:09:31+00:00
22961  2023-05-20 23:09:02+00:00
22964  2023-05-20 23:07:00+00:00

user \
5      {'_type': 'snsrape.modules.twitter.User', 'us...
19     {'_type': 'snsrape.modules.twitter.User', 'us...
30     {'_type': 'snsrape.modules.twitter.User', 'us...
33     {'_type': 'snsrape.modules.twitter.User', 'us...
34     {'_type': 'snsrape.modules.twitter.User', 'us...
...
22956  {'_type': 'snsrape.modules.twitter.User', 'us...
22959  {'_type': 'snsrape.modules.twitter.User', 'us...
22960  {'_type': 'snsrape.modules.twitter.User', 'us...
22961  {'_type': 'snsrape.modules.twitter.User', 'us...
22964  {'_type': 'snsrape.modules.twitter.User', 'us...

content id \
5      'Heart wrenching': #AI #expert details #danger... 1660091546161627139
19     🐊 Gator #171\n\n🐊\n\n#AI #AIArt #AIイラスト #AIAr... 1660089566391603200
30     When the Default happens #DebtCeilingCrisis YE... 1660088677630976000
33     Checkout the latest updates from David Estrada... 1660088378816184321
34     Echa un vistazo a las últimas actualizaciones ... 1660088217385705473
...
22956  https://t.co/lFX97p6oMz https://t.co/kVNH12XM... 1660060648548573184
22959  From @AgileScrumGuide author @ScottGraffius .... 1660060256893042688
22960  7 Advantages and 7 Disadvantages of Artificial... 1660060212227653633
22961  #ChatGPT #ArtificialIntelligence \n\nYou Must ... 1660060088617299971
22964  These five data management trends will shape t... 1660059576262176771

replyCount retweetCount likeCount quoteCount \
5          0          0          0          0
19         0          0          2          0
30         0          0          0          0
33         0          0          0          0
34         0          0          0          0
...
22956      0          0          0          0
22959      0          1          7          0
22960      0          0          1          0
22961      0          0          0          0
22964      0          0          5          0

hashtags cashtags \
5      [AI, expert, dangers, deepfakes, detect, finte... None
19     [AI, AIArt, AIイラスト, AIArtSociety, AIGenerated,... None
30     [DebtCeilingCrisis, ArtificialIntelligence] None
33     [EZEEDBUY, SEE_OPEN_BUY, AIaaS, Innovative, Ret... None
34     [EZEEDBUY, SEE_OPEN_BUY, AIaaS, Innovative, Ret... None
...
22956  [AIart, GenerativeArt, ArtificialIntelligence,... None
22959  [AI, ArtificialIntelligence, MachineLearning, ... None
22960  [AI, MachineLearning, DataScience, ArtificialI... None
22961  [ChatGPT, ArtificialIntelligence] None
22964  [DataScience, ArtificialIntelligence] None

mentionedUsers retweetedTweet \
5      [{'_type': 'snsrape.modules.twitter.User', 'u... None
19     None None
30     None None
33     None None
34     None None
...
22956  [{'_type': 'snsrape.modules.twitter.User', 'u... None
22959  [{'_type': 'snsrape.modules.twitter.User', 'u... None
22960  None None
22961  None None
22964  None None

```




```
df_filtered.count()
```

```
date          6241
user          6241
content       6241
id            6241
replyCount    6241
retweetCount  6241
likeCount     6241
quoteCount    6241
hashtags      6241
cashtags       69
mentionedUsers 1590
retweetedTweet 0
quotedTweet   197
lang          6241
links         3962
media         3048
dtype: int64
#sns = sns.set_theme(style='whitegrid')
```

```
columns_to_drop = ['cashtags', 'mentionedUsers', 'retweetedTweet', 'quotedTweet']
df_filtered = df_filtered.drop(columns_to_drop, axis=1)
```

```
df_filtered.head()
```

	date	user	content	id	replyCount	retweetCount	likeCount	quot
5	2023-05-21 01:14:02+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	'Heart wrenching': #AI #expert details #danger...	1660091546161627139	0	0	0	
19	2023-05-21 01:06:10+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	 Gator #171\n\n AI #AI Art #AI イラスト #AI Ar...	1660089566391603200	0	0	2	
30	2023-05-21 01:02:38+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	When the Default happens #DebtCeilingCrisis YE...	1660088677630976000	0	0	0	
33	2023-05-21 01:01:27+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Checkout the latest updates from David Estrada...	1660088378816184321	0	0	0	
34	2023-05-21 01:00:48+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Echa un vistazo a las últimas actualizaciones ...	1660088217385705473	0	0	0	




```
df_filtered.count()
```

```
date          6241
user          6241
content       6241
id            6241
replyCount    6241
retweetCount  6241
likeCount     6241
quoteCount    6241
hashtags      6241
lang          6241
links         3962
media         3048
dtype: int64
```

```
df_filtered['links'] = df_filtered['links'].fillna('null')
df_filtered['media'] = df_filtered['media'].fillna('null')
df_filtered.count()
```

date	6241
user	6241
content	6241
id	6241
replyCount	6241
retweetCount	6241
likeCount	6241
quoteCount	6241
hashtags	6241
lang	6241
links	6241
media	6241
dtype: int64	

```
df_filtered.head()
```

	date	user	content	id	replyCount	retweetCount	likeCount	quot
5	2023-05-21 01:14:02+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	'Heart wrenching': #AI #expert details #danger...	1660091546161627139	0	0	0	
19	2023-05-21 01:06:10+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	 #171\n\n AI Art #AIイラスト #AIAr...	1660089566391603200	0	0	2	
30	2023-05-21 01:02:38+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	When the Default happens #DebtCeilingCrisis YE...	1660088677630976000	0	0	0	
33	2023-05-21 01:01:27+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Checkout the latest updates from David Estrada...	1660088378816184321	0	0	0	
34	2023-05-21 01:00:48+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Echa un vistazo a las últimas actualizaciones ...	1660088217385705473	0	0	0	



```
en_tweets = df_filtered[df_filtered['lang'] == 'en']
count_en_tweets = en_tweets.shape[0]
print("Number of English tweets:", count_en_tweets)
```

Number of English tweets: 5536

```
qme_tweets = df_filtered[df_filtered['lang'] == 'qme']
count_qme_tweets = qme_tweets.shape[0]
print("Number of qme tweets:", count_qme_tweets)
```

Number of qme tweets: 220

```
other_tweets = df_filtered[(df_filtered['lang'] != 'qme') & (df_filtered['lang'] != 'en')]
count_other_tweets = other_tweets.shape[0]
print("Number of tweets in other languages:", count_other_tweets)
```

Number of tweets in other languages: 485

```
other_languages = df_filtered[~df_filtered['lang'].isin(['en', 'qme'])]['lang'].unique()
print("Languages other than 'en' and 'qme':", other_languages)
```

```

Languages other than 'en' and 'gme': ['in' 'es' 'de' 'it' 'zh' 'ko' 'fr' 'ja' 'und' 'pt' 'da' 'qht' 'ht' 'hi'
'el' 'ro' 'ar' 'lt' 'bn' 'ca' 'ml']

```

```

df_filtered = df_filtered[df_filtered['lang'] == 'en']
df_filtered.count()

```

```

date          5536
user          5536
content       5536
id            5536
replyCount    5536
retweetCount  5536
likeCount     5536
quoteCount    5536
hashtags      5536
lang          5536
links         5536
media         5536
dtype: int64

```

```
df_filtered.head()
```

	user	content	id	replyCount	retweetCount	likeCount	quoteCount	hasht
	{'_type': ape.modules.twitter.User', 'us...	'Heart wrenching': #AI #expert details #danger...	1660091546161627139	0	0	0	0	[AI, expert, dan deepfakes, de fil
	{'_type': ape.modules.twitter.User', 'us...	When the Default happens #DebtCeilingCrisis YE...	1660088677630976000	0	0	0	0	[DebtCeilingC ArtificialIntellige
	{'_type': ape.modules.twitter.User', 'us...	Checkout the latest updates from David Estrada...	1660088378816184321	0	0	0	0	[EZEEL SEE_OPEN_] AlaaS, Innovat f
	{'_type': ape.modules.twitter.User', 'us...	Opinion: #CambridgeAnalytica on steroids: '\n\n...	1660087769895452672	0	0	0	0	[CambridgeAnaly ArtificialIntellige
	{'_type': ape.modules.twitter.User', 'us...	This nimble quadruped is designed to be a pet ...	1660087570141724675	0	1	4	1	[Robo ArtificialIntellige MI, Robo

```
df_filtered.to_csv('twitter_English_filtered_scrapper.csv', index=False)
```

```
df_filtered = df_filtered[df_filtered['likeCount'] >= 10]
```

```

# With 1 Lakh Tweet Scrapping got the output of 18 columns with all my filters
# With 10 Lakh Tweet Scrapping got the output of 18 columns with all my filters
# With 1 Cr Tweet Scrapping it got decreased to 16 Columns only with my filters

```

```

df_filtered.count()
# df_filtered.head(100)

```

```

date          269
user          269
content       269
id            269
replyCount    269
retweetCount  269
likeCount     269
quoteCount    269
hashtags      269
lang          269
links         269

```

```
media          269
dtype: int64
```

```
df_filtered.to_csv('twitter_filtered_scrapper.csv', index=False)
```

```
df_filtered.head(100)
```

	date	user	content	id	replyCount	retweetCount	likeCou
321	2023-05-20 23:06:04+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	On Instagram 🍷 https://t.co/HNYrOC0XGx #AI #AI...	1660059344107606017	0		1
424	2023-05-20 22:29:46+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	Chaotic.\n\nLove the clouds from the left\nBut...	1660050209039163392	3		1
558	2023-05-20 21:42:20+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	5/21/2023- Chuckie #AI #ArtificialIntelligence...	1660038270741364736	0		12
728	2023-05-20 20:38:31+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	After working tirelessly for months and months...	1660022212525535232	0		2
792	2023-05-20 20:19:06+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	#Robot playing Table Tennis\nvia @jblefevre60\...	1660017324307390468	2		6
...	...	...	...	...	...		...
8137	2023-05-20 14:23:55+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	The Optimist's #Guide \n\ninto #ArtificialIntell...	1659927940002373633	2		10
8140	2023-05-20 14:23:01+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	How #MachineLearning Can Prevent Credit Card F...	1659927713044475905	2		8
8291	2023-05-20 13:54:51+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	We will be hosting an AI Prompt Engineering Ma...	1659920622976131072	1		3
8344	2023-05-20 13:42:29+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	G7 leaders call for 'responsible' use of gener...	1659917513704996865	11		0
8383	2023-05-20 13:35:00+00:00	{ '_type': 'snsrape.modules.twitter.User', 'us...	#ArtificialIntelligence will drive massive dis...	1659915628663418880	5		22

100 rows × 12 columns



```
# Read the first CSV file
df1 = pd.read_csv('/content/twitter_scrapper.csv')

# Read the second CSV file
df2 = pd.read_csv('/content/twitter_filtered_scrapper.csv')

# Append the second file to the first file
merged_df = pd.concat([df1, df2])

# Save the merged data to a new CSV file
merged_df.to_csv('merged.csv', index=False)
```

```
final_data = pd.read_csv("/content/merged.csv")
final_data.head()
```

	date	user	content	id	replyCount	retweetCount	likeCount	quoteCo
0	2023-05-20 18:32:38+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	10 Algorithms for that dominate our world.\n\...	1659990533186494464	0	3	13	
1	2023-05-20 18:12:01+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	1/ What if LLM Hallucinations Were A Feature A...	1659985344836820993	1	7	11	
2	2023-05-20 17:23:01+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	What is reinforcement learning? How #AI trains...	1659973013264117760	0	5	15	
3	2023-05-20 16:55:13+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	Will be live on #pancakeswap SOON 🔥 \n\nWolf AI...	1659966014862372864	1	8	18	
4	2023-05-20 15:24:14+00:00	{'_type': 'snsrape.modules.twitter.User', 'us...	The volatility created by sophisticated AI-pow...	1659943119520165888	1	1	12	



```
final_data.count()
```

```
date          901
user          901
content       901
id            901
replyCount    901
retweetCount  901
likeCount     901
quoteCount    901
hashtags      901
lang          901
links         595
media         555
dtype: int64
```

```
import nltk
nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
True
```

```
#Finding the Compound Score, Positive & negative sentiments from the tweets
import pandas as pd
from nltk.sentiment import SentimentIntensityAnalyzer

# Read the CSV file into a DataFrame
final_data = pd.read_csv("/content/twitter_English_filtered_scrapper.csv")

# Initialize the sentiment analyzer
analyzer = SentimentIntensityAnalyzer()

# Define a function to get the sentiment scores using VADER
def get_sentiment_scores(text):
    return analyzer.polarity_scores(text)

# Apply sentiment analysis on the 'content' column and create a new column for sentiment scores
final_data['sentiment_scores'] = final_data['content'].apply(get_sentiment_scores)

# Extract the compound sentiment score from the sentiment scores dictionary
final_data['compound_score'] = final_data['sentiment_scores'].apply(lambda scores: scores['compound'])
```

```
print(final_data.head())
```

```

                                date \
0  2023-05-21 01:14:02+00:00
1  2023-05-21 01:02:38+00:00
2  2023-05-21 01:01:27+00:00
3  2023-05-21 00:59:01+00:00
4  2023-05-21 00:58:14+00:00

                                user \
0  {'_type': 'snsrape.modules.twitter.User', 'us...
1  {'_type': 'snsrape.modules.twitter.User', 'us...
2  {'_type': 'snsrape.modules.twitter.User', 'us...
3  {'_type': 'snsrape.modules.twitter.User', 'us...
4  {'_type': 'snsrape.modules.twitter.User', 'us...

                                content                                id \
0  'Heart wrenching': #AI #expert details #danger... 1660091546161627139
1  When the Default happens #DebtCeilingCrisis YE... 1660088677630976000
2  Checkout the latest updates from David Estrada... 1660088378816184321
3  Opinion: #CambridgeAnalytica on steroids: \n\n... 1660087769895452672
4  This nimble quadruped is designed to be a pet ... 1660087570141724675

replyCount  retweetCount  likeCount  quoteCount \
0           0             0           0           0
1           0             0           0           0
2           0             0           0           0
3           0             0           0           0
4           0             1           4           1

                                hashtags lang \
0  ['AI', 'expert', 'dangers', 'deepfakes', 'dete... en
1  ['DebtCeilingCrisis', 'ArtificialIntelligence'] en
2  ['EZEEDBUY', 'SEE_OPEN_BUY', 'AIaaS', 'Innovati... en
3  ['CambridgeAnalytica', 'ArtificialIntelligence... en
4  ['Robot', 'AI', 'ArtificialIntelligence', 'MI'... en

                                links \
0  [{'_type': 'snsrape.modules.twitter.TextLink'...
1  NaN
2  [{'_type': 'snsrape.modules.twitter.TextLink'...
3  [{'_type': 'snsrape.modules.twitter.TextLink'...
4  NaN

                                media \
0  [{'_type': 'snsrape.modules.twitter.Photo', '...'
1  NaN
2  NaN
3  [{'_type': 'snsrape.modules.twitter.Photo', '...'
4  [{'_type': 'snsrape.modules.twitter.Video', '...'

                                sentiment_scores  compound_score
0  {'neg': 0.106, 'neu': 0.894, 'pos': 0.0, 'comp... -0.3818
1  {'neg': 0.049, 'neu': 0.713, 'pos': 0.237, 'co... 0.8513
2  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... 0.0000
3  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... 0.0000
4  {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound... 0.0000

```

```
final_data.count()
```

```

date           5536
user           5536
content        5536
id             5536
replyCount     5536
retweetCount   5536
likeCount      5536
quoteCount     5536
hashtags       5536
lang           5536
links          3569
media          2674
sentiment_scores 5536
compound_score   5536
dtype: int64

```

```

#Finding what people are feeling. Count of people who feels optimistic, negative & neural about AI
positive_count = 0
negative_count = 0

```

```

neutral_count = 0

for score in final_data['compound_score']:
    if score > 0.2:
        positive_count += 1
    elif score < -0.2:
        negative_count += 1
    else:
        neutral_count += 1

print("Number of people feeling optimistic:", positive_count)
print("Number of people feeling negative:", negative_count)
print("Number of people feeling neutral:", neutral_count)

```

```

Number of people feeling optimistic: 2632
Number of people feeling negative: 621
Number of people feeling neutral: 2283

```

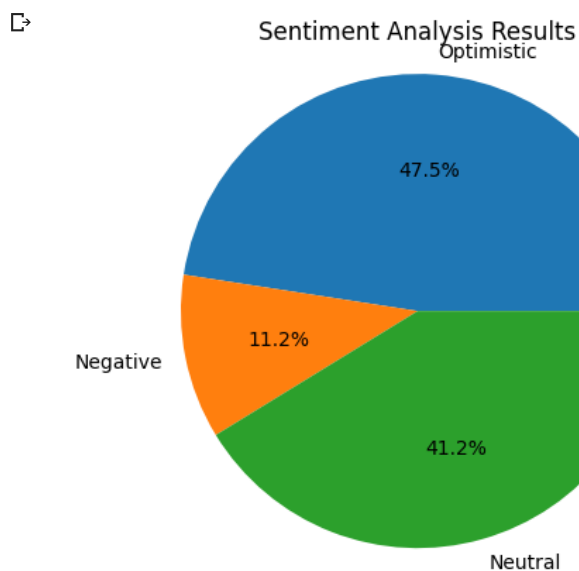
```

import matplotlib.pyplot as plt
from wordcloud import WordCloud

# Pie chart
sentiment_counts = [positive_count, negative_count, neutral_count]
sentiment_labels = ['Optimistic', 'Negative', 'Neutral']

plt.pie(sentiment_counts, labels=sentiment_labels, autopct='%1.1f%%')
plt.title('Sentiment Analysis Results')
plt.axis('equal')
plt.show()

```



```

import matplotlib.pyplot as plt

# Bar chart
sentiments = ['Optimistic', 'Negative', 'Neutral']
sentiment_counts = [positive_count, negative_count, neutral_count]

plt.bar(sentiments, sentiment_counts)
plt.title('Sentiment Analysis Results')
plt.xlabel('Sentiment')
plt.ylabel('Count')
plt.show()

```





✓ 0s completed at 10:09 PM

● ×