

# Twitter Data Vaccine 2

March 6, 2023

```
[1]: #libraries needed
!pip install snsrape
!pip install textblob
!pip install emot
!pip install dask
import pandas as pd
import snsrape.modules.twitter as sntwitter
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import time
from dask import dataframe as df1

import nltk
#nltk.download('stopwords')#run once and comment out
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.stem.porter import PorterStemmer

import string
import re
import textblob
from textblob import TextBlob
from wordcloud import WordCloud, STOPWORDS
from emot.emo_unicode import UNICODE_EMOJI

lemmatizer = WordNetLemmatizer()

from wordcloud import ImageColorGenerator
from PIL import Image

import warnings
%matplotlib inline
```

Requirement already satisfied: snsrape in ./opt/anaconda3/lib/python3.9/site-packages (0.5.0.20230113)

Requirement already satisfied: filelock in ./opt/anaconda3/lib/python3.9/site-

packages (from snsrape) (3.6.0)  
 Requirement already satisfied: lxml in ./opt/anaconda3/lib/python3.9/site-  
 packages (from snsrape) (4.8.0)  
 Requirement already satisfied: requests[socks] in  
 ./opt/anaconda3/lib/python3.9/site-packages (from snsrape) (2.27.1)  
 Requirement already satisfied: beautifulsoup4 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from snsrape) (4.11.1)  
 Requirement already satisfied: soupsieve>1.2 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from beautifulsoup4->snsrape)  
 (2.3.1)  
 Requirement already satisfied: certifi>=2017.4.17 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)  
 (2021.10.8)  
 Requirement already satisfied: urllib3<1.27,>=1.21.1 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)  
 (1.26.9)  
 Requirement already satisfied: charset-normalizer~2.0.0 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)  
 (2.0.4)  
 Requirement already satisfied: idna<4,>=2.5 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)  
 (3.3)  
 Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)  
 (1.7.1)  
 Requirement already satisfied: textblob in ./opt/anaconda3/lib/python3.9/site-  
 packages (0.17.1)  
 Requirement already satisfied: nltk>=3.1 in ./opt/anaconda3/lib/python3.9/site-  
 packages (from textblob) (3.7)  
 Requirement already satisfied: click in ./opt/anaconda3/lib/python3.9/site-  
 packages (from nltk>=3.1->textblob) (8.0.4)  
 Requirement already satisfied: joblib in ./opt/anaconda3/lib/python3.9/site-  
 packages (from nltk>=3.1->textblob) (1.1.0)  
 Requirement already satisfied: regex>=2021.8.3 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from nltk>=3.1->textblob)  
 (2022.3.15)  
 Requirement already satisfied: tqdm in ./opt/anaconda3/lib/python3.9/site-  
 packages (from nltk>=3.1->textblob) (4.64.0)  
 Requirement already satisfied: emot in ./opt/anaconda3/lib/python3.9/site-  
 packages (3.1)  
 Requirement already satisfied: dask in ./opt/anaconda3/lib/python3.9/site-  
 packages (2022.2.1)  
 Requirement already satisfied: pyyaml>=5.3.1 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from dask) (6.0)  
 Requirement already satisfied: fsspec>=0.6.0 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from dask) (2022.2.0)  
 Requirement already satisfied: partd>=0.3.10 in  
 ./opt/anaconda3/lib/python3.9/site-packages (from dask) (1.2.0)

```
Requirement already satisfied: cloudpickle>=1.1.1 in
./opt/anaconda3/lib/python3.9/site-packages (from dask) (2.0.0)
Requirement already satisfied: packaging>=20.0 in
./opt/anaconda3/lib/python3.9/site-packages (from dask) (21.3)
Requirement already satisfied: toolz>=0.8.2 in
./opt/anaconda3/lib/python3.9/site-packages (from dask) (0.11.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
./opt/anaconda3/lib/python3.9/site-packages (from packaging>=20.0->dask) (3.0.4)
Requirement already satisfied: locket in ./opt/anaconda3/lib/python3.9/site-
packages (from partd>=0.3.10->dask) (0.2.1)
```

```
[2]: # Get Tweets with the keyword Covid or pandemic or Covid-19
covid_scraper = sntwitter.TwitterSearchScraper('((Covid OR Covid-19 OR Covid19)
↳AND (vaccine OR booster OR shot OR Pfizer OR Moderna)) near:"Toronto" since:
↳2021-01-01 until:2022-12-31 lang:"en"')
print(type(covid_scraper))
# Display the data from the first tweet
for covid_tweet in covid_scraper.get_items():
    break
covid_tweet
```

```
<class 'snsrape.modules.twitter.TwitterSearchScraper'>
```

```
[2]: Tweet(url='https://twitter.com/AlokMukherjee45/status/1608322460285558784',
date=datetime.datetime(2022, 12, 29, 4, 42, 10, tzinfo=datetime.timezone.utc),
rawContent='What f_____g hypocrisy. 1.1 million people are dead in the US of
COVID. There is no masking, no testing, no enforcement of booster vaccinating
and no data collection. Where does @POTUS get off enforcing restrictions on
people from anywhere else?\nhhttps://t.co/9guvpE1DGG', renderedContent='What
f_____g hypocrisy. 1.1 million people are dead in the US of COVID. There is no
masking, no testing, no enforcement of booster vaccinating and no data
collection. Where does @POTUS get off enforcing restrictions on people from
anywhere else?\nabc.ca/news/world/cov...', id=1608322460285558784,
user=User(username='AlokMukherjee45', id=248891184, displayname='Alok
Mukherjee', rawDescription='Views my own. In the fight for a different normal.',
renderedDescription='Views my own. In the fight for a different normal.',
descriptionLinks=None, verified=False, created=datetime.datetime(2011, 2, 7, 23,
5, 2, tzinfo=datetime.timezone.utc), followersCount=1325, friendsCount=594,
statusesCount=11874, favouritesCount=11009, listedCount=22, mediaCount=129,
location='Toronto, Ontario', protected=False, link=None, profileImageUrl='https:
//pbs.twimg.com/profile_images/1367137795383447558/TkWWNgpV_normal.jpg',
profileBannerUrl='https://pbs.twimg.com/profile_banners/248891184/1614786117',
label=None), replyCount=0, retweetCount=0, likeCount=1, quoteCount=0,
conversationId=1608322460285558784, lang='en', source='<a
href="http://twitter.com/download/android" rel="nofollow">Twitter for
Android</a>', sourceUrl='http://twitter.com/download/android',
sourceLabel='Twitter for Android',
links=[TextLink(text='abc.ca/news/world/cov...',
```

```

url='https://www.cbc.ca/news/world/covid-china-travellers-u-s-canada-
test-1.6698952', tcourl='https://t.co/9guvpE1DGG', indices=(246, 269)],
media=None, retweetedTweet=None, quotedTweet=None, inReplyToTweetId=None,
inReplyToUser=None, mentionedUsers=[User(username='POTUS',
id=1349149096909668363, displayname='President Biden', rawDescription=None,
renderedDescription=None, descriptionLinks=None, verified=None, created=None,
followersCount=None, friendsCount=None, statusesCount=None,
favouritesCount=None, listedCount=None, mediaCount=None, location=None,
protected=None, link=None, profileImageUrl=None, profileBannerUrl=None,
label=None)], coordinates=Coordinates(longitude=-79.639319, latitude=43.403221),
place=Place(id='3797791ff9c0e4c6', fullName='Toronto, Ontario', name='Toronto',
type='city', country='Canada', countryCode='CA'), hashtags=None, cashtags=None,
card=SummaryCard(title='U.S. to require negative COVID-19 test for travellers
from China - but Canada not changing policy...',
url='https://www.cbc.ca/news/world/covid-china-travellers-u-s-canada-
test-1.6698952', description='The U.S. announced new COVID-19 testing
requirements Wednesday for all travellers from China, joining other nations that
have imposed restrictions because of a surge of infections.', thumbnailUrl='h
ttps://pbs.twimg.com/card_img/1628670463852290048/BN0SrGJO?format=jpg&name=orig'
, siteUser=User(username='CBC', id=16034244, displayname='CBC',
rawDescription='The official CBC Twitter. Bringing you the best of Canada
', renderedDescription='The official CBC Twitter. Bringing you the best
of Canada ', descriptionLinks=None, verified=True,
created=datetime.datetime(2008, 8, 28, 23, 8, 39, tzinfo=datetime.timezone.utc),
followersCount=709092, friendsCount=75204, statusesCount=151253,
favouritesCount=46106, listedCount=3765, mediaCount=15444, location='Canada',
protected=False, link=TextLink(text='cbc.ca', url='http://www.cbc.ca',
tcourl='https://t.co/wH8VYl3fRw', indices=(0, 23)), profileImageUrl='https://pbs
.twimg.com/profile_images/1576969187125256196/lw5LcJF1_normal.jpg',
profileBannerUrl='https://pbs.twimg.com/profile_banners/16034244/1678125419',
label=None), creatorUser=None), viewCount=96, vibe=None)

```

```

[4]: data = []
# Using TwitterSearchScrapper to scrape data and append tweets to list
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper
    ('((Covid OR pandemic OR Covid-19 OR Covid19) AND (vaccine OR booster OR_
    ↪shot OR dose OR Pfizer OR Moderna)) in: ("Toronto" OR "Montréal" OR_
    ↪"Vancouver" OR "Calgary" OR "Edmonton" OR "Ottawa" OR "Winnipeg" OR "Québec"_
    ↪OR "Hamilton" OR "Waterloo") since:2021-01-01 until:2022-12-31 lang:"en").
    ↪get_items()):
    if i>30000:
        break
    else:
        data.append([tweet.user.username, tweet.date, tweet.rawContent, tweet.
        ↪likeCount, tweet.user.followersCount, tweet.user.friendsCount, tweet.
        ↪replyCount, tweet.retweetCount, tweet.quoteCount, tweet.hashtags])

```

```
df = pd.DataFrame(data, columns = ["User", "Date Created", "Tweet", "Likes", "Followers", "Friends", "Reply", "Retweet", "Quote", "Hashtags"])
#write the resulting raw data in a csv
df.to_csv('~\Desktop\Twitter_New\raw_tweet_30K_vaccine_21_22.csv')
df
```

```
[4]:
```

	User	Date Created	\
0	Treep52	2022-12-30 22:51:15+00:00	
1	DrGurdeepParhar	2022-12-30 21:10:00+00:00	
2	stopthegaslight	2022-12-30 18:49:59+00:00	
3	Daisy381	2022-12-30 17:31:01+00:00	
4	RadarsScience	2022-12-30 16:20:06+00:00	
...	...	...	
21024	nika180479	2021-01-01 06:18:16+00:00	
21025	forjustminutes	2021-01-01 04:11:23+00:00	
21026	profamirrattaran	2021-01-01 02:05:23+00:00	
21027	EricBenninghoff	2021-01-01 01:45:26+00:00	
21028	NanouTheNomad	2021-01-01 00:24:34+00:00	

	Tweet	Likes	Followers	\
0	Per Alex: Approximately 95% of adults and 98% ...	1	3425	
1	Big Tobacco's divestment from Quebec's Medicag...	0	66607	
2	@brianlilley COVID Vaccinations are the most u...	0	17	
3	@MerlinofCanada @JustinTrudeau In Alberta it's...	1	331	
4	Big Tobacco&#039;s divestment from Quebec&...	0	29	
...	...	...	...	
21024	@Lin_Manuel It certainly been very hard year. ...	0	25	
21025	NOV/DEC \n-Covid cases shot back up as lead...	1	749	
21026	@DeNovo_Fatima @robyn_s_lee Translation: since...	3	14637	
21027	Adults with disabilities in Hamilton County ar...	2	602	
21028	In lieu of traditional vaccine selfie, I offer...	14	1965	

	Friends	Reply	Retweet	Quote	\
0	3949	0	0	0	
1	48	0	0	0	
2	189	0	0	0	
3	598	2	0	0	
4	45	0	0	0	
...	...	...	...	...	
21024	256	0	0	0	
21025	789	0	0	0	
21026	163	0	0	0	
21027	139	0	0	0	
21028	970	0	1	0	

	Hashtags
0	None

```

1                                     None
2                                     None
3                                     None
4                                     None
...                                  ...
21024                               None
21025                               None
21026                               None
21027                               None
21028 [Hamilton, COVID19, science, publichealth, Not...

```

[21029 rows x 10 columns]

```

[5]: #Data preprocessing
eng_stop_words = list(stopwords.words('english'))
emoji = list(UNICODE_EMOJI.keys())
#nltk.download('punkt') #comment out after download
#nltk.download('wordnet') #comment out after download
#nltk.download('omw-1.4') #comment out after download

#Function to preprocess tweets
def ProcessedTweets(text):
    text = text.lower() #convert words to lower case

    text = ' '.join(re.sub("(@[A-Za-z0-9]+)|(^0-9A-Za-z \t))|(\w+:\/\/\S+)", "",text).split()) #remove @ and links

    text = re.sub(r'\@w+|\#w+|\d+', '', text) #remove repeated characters

    punct = str.maketrans('','',string.punctuation+string.digits) #remove
    punctuation and numbers
    text = text.translate(punct)

    tokens = word_tokenize(text) # tokenizing words and removing stop words
    from teh tweet text
    filtered_words = [w for w in tokens if w not in eng_stop_words]
    filtered_words = [w for w in filtered_words if w not in emoji]

    lemmatizer = WordNetLemmatizer() # lemmmatizing words
    lemma_words = [lemmatizer.lemmatize(w) for w in filtered_words]
    text = " ".join(lemma_words)

    return text

```

```

[6]: #Generate a new column called 'Processed Tweets' by applying ProcessedTweets
    function to the 'Tweet' Column
df = pd.read_csv('~\Desktop\Twitter_New\raw_tweet_30K_vaccine_21_22.csv')

```

```
df['Processed Tweets'] = df['Tweet'].apply(ProcessedTweets)
df.to_csv('~\Desktop\Twitter_New\processed_tweet_30K_vaccine_21_22.csv')
df
```

```
[6]:
```

	Unnamed: 0	User	Date Created	\
0	0	Treep52	2022-12-30 22:51:15+00:00	
1	1	DrGurdeepParhar	2022-12-30 21:10:00+00:00	
2	2	stopthegaslight	2022-12-30 18:49:59+00:00	
3	3	Daisy381	2022-12-30 17:31:01+00:00	
4	4	RadarsScience	2022-12-30 16:20:06+00:00	
...	...	...	...	
21024	21024	nika180479	2021-01-01 06:18:16+00:00	
21025	21025	forjustminutes	2021-01-01 04:11:23+00:00	
21026	21026	profamirattaran	2021-01-01 02:05:23+00:00	
21027	21027	EricBenninghoff	2021-01-01 01:45:26+00:00	
21028	21028	NanouTheNomad	2021-01-01 00:24:34+00:00	

	Tweet	Likes	Followers	\
0	Per Alex: Approximately 95% of adults and 98% ...	1	3425	
1	Big Tobacco's divestment from Quebec's Medicag...	0	66607	
2	@brianlilley COVID Vaccinations are the most u...	0	17	
3	@MerlinofCanada @JustinTrudeau In Alberta it's...	1	331	
4	Big Tobacco&#039;s divestment from Quebec&...	0	29	
...	...	...	...	
21024	@Lin_Manuel It certainly been very hard year. ...	0	25	
21025	NOV/DEC \n-Covid cases shot back up as lead...	1	749	
21026	@DeNovo_Fatima @robyn_s_lee Translation: since...	3	14637	
21027	Adults with disabilities in Hamilton County ar...	2	602	
21028	In lieu of traditional vaccine selfie, I offer...	14	1965	

	Friends	Reply	Retweet	Quote	\
0	3949	0	0	0	
1	48	0	0	0	
2	189	0	0	0	
3	598	2	0	0	
4	45	0	0	0	
...	...	...	...	...	
21024	256	0	0	0	
21025	789	0	0	0	
21026	163	0	0	0	
21027	139	0	0	0	
21028	970	0	1	0	

	Hashtags	\
0	NaN	
1	NaN	
2	NaN	

```

3
4
...
21024
21025
21026
21027
21028 ['Hamilton', 'COVID19', 'science', 'publicheal...

```

```

                                Processed Tweets
0      per alex approximately adult adult quebec rece...
1      big tobacco divestment quebec medicago step ri...
2      covid vaccination unsafe ineffective vaccine e...
3      alberta ' ppl rollout edmonton walk pharmacy g...
4      big tobaccoamps divestment quebecamps medicago...
...
21024  manuel certainly hard year literally started d...
21025  novdec covid case shot back leadership drag...
21026  fatima slee translation since quebec ontario f...
21027  adult disability hamilton county eligible rece...
21028  lieu traditional vaccine selfie offer gif song...

```

[21029 rows x 12 columns]

```

[7]: #Add a column converting date and time to date only
from datetime import datetime
from email.utils import mktime_tz, parsedate_tz

# dtme = tweet['created_at']
dtme = '2022-01-01 00:05:15+00:00'
new_datetime = datetime.strptime(datetime.strptime(dtme, '%Y-%m-%d %H:%M:%S+00:
↳00'), '%Y-%m-%d')
print((new_datetime))

# Function to convert date
def ConvertDate(dt):
    new_date = datetime.strptime(datetime.strptime(str(dt), '%Y-%m-%d %H:%M:
↳%S+00:00'), '%Y-%m-%d')
    return new_date

```

2022-01-01

```

[8]: df['Date'] = df['Date Created'].apply(ConvertDate)
df.to_csv('~\Desktop\Twitter_New\prcessed_tweet_date_30K_vaccine_21_22.csv')

```

```

[9]: df2 = pd.read_csv('~\Desktop\Twitter_New\prcessed_tweet_date_30K_vaccine_21_22.
↳csv')

```



df2

```
[9]:      Unnamed: 0.1  Unnamed: 0      User      Date Created \
0              0      0      Treep52  2022-12-30 22:51:15+00:00
1              1      1  DrGurdeepParhar  2022-12-30 21:10:00+00:00
2              2      2  stopthegaslight  2022-12-30 18:49:59+00:00
3              3      3      Daisy381  2022-12-30 17:31:01+00:00
4              4      4    RadarsScience  2022-12-30 16:20:06+00:00
...
21024      21024      21024      nika180479  2021-01-01 06:18:16+00:00
21025      21025      21025    forjustminutes  2021-01-01 04:11:23+00:00
21026      21026      21026  profamirattaran  2021-01-01 02:05:23+00:00
21027      21027      21027  EricBenninghoff  2021-01-01 01:45:26+00:00
21028      21028      21028    NanouTheNomad  2021-01-01 00:24:34+00:00
```

```
      Tweet  Likes  Followers \
0  Per Alex: Approximately 95% of adults and 98% ...      1      3425
1  Big Tobacco's divestment from Quebec's Medicag...      0      66607
2  @brianlilley COVID Vaccinations are the most u...      0       17
3  @MerlinofCanada @JustinTrudeau In Alberta it's...      1      331
4  Big Tobacco&#039;s divestment from Quebec&...      0       29
...
21024  @Lin_Manuel It certainly been very hard year. ...      0       25
21025  NOV/DEC \n-Covid cases shot back up as lead...      1      749
21026  @DeNovo_Fatima @robyn_s_lee Translation: since...      3     14637
21027  Adults with disabilities in Hamilton County ar...      2      602
21028  In lieu of traditional vaccine selfie, I offer...     14     1965
```

```
      Friends  Reply  Retweet  Quote \
0      3949      0      0      0
1       48      0      0      0
2      189      0      0      0
3      598      2      0      0
4       45      0      0      0
...
21024      256      0      0      0
21025      789      0      0      0
21026      163      0      0      0
21027      139      0      0      0
21028      970      0      1      0
```

```
      Hashtags \
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
```

```
...
21024 NaN
21025 NaN
21026 NaN
21027 NaN
21028 ['Hamilton', 'COVID19', 'science', 'publicheal...
```

	Processed Tweets	Date
0	per alex approximately adult adult quebec rece...	2022-12-30
1	big tobacco divestment quebec medicago step ri...	2022-12-30
2	covid vaccination unsafe ineffective vaccine e...	2022-12-30
3	alberta ' ppl rollout edmonton walk pharmacy g...	2022-12-30
4	big tobaccoamps divestment quebecamps medicago...	2022-12-30
...	...	...
21024	manuel certainly hard year literally started d...	2021-01-01
21025	novdec covid case shot back leadership drag...	2021-01-01
21026	fatima slee translation since quebec ontario f...	2021-01-01
21027	adult disability hamilton county eligible rece...	2021-01-01
21028	lieu traditional vaccine selfie offer gif song...	2021-01-01

[21029 rows x 14 columns]

```
[10]: # Sentiment Analysis
import textblob
from textblob import TextBlob

# Function for polarity score
# Polarity is the output that lies between [-1,1],
# where -1 refers to negative sentiment and +1 refers to positive sentiment.
def polarity(text):
    return TextBlob(str(text)).sentiment.polarity

# Function for subjectivity score
# Subjectivity is the output that lies within [0,1]
# and refers to personal opinions and judgments.
# If a sentence has high subjectivity i.e. close to 1,
# It resembles that the text contains more personal opinion than factual
  ↪ information.
def subjectivity(text):
    return TextBlob(str(text)).sentiment.subjectivity
```

```
[11]: # Get polarity of processed text
df2['Polarity'] = df2['Processed Tweets'].apply(polarity)
df2['Subjectivity'] = df2['Processed Tweets'].apply(subjectivity)
df2
```

```
[11]:      Unnamed: 0.1  Unnamed: 0      User      Date Created \
0          0          0      Treep52  2022-12-30 22:51:15+00:00
1          1          1  DrGurdeepParhar  2022-12-30 21:10:00+00:00
2          2          2  stopthegaslight  2022-12-30 18:49:59+00:00
3          3          3      Daisy381  2022-12-30 17:31:01+00:00
4          4          4    RadarsScience  2022-12-30 16:20:06+00:00
...
21024      21024      21024      nika180479  2021-01-01 06:18:16+00:00
21025      21025      21025    forjustminutes  2021-01-01 04:11:23+00:00
21026      21026      21026  profamirattaran  2021-01-01 02:05:23+00:00
21027      21027      21027  EricBenninghoff  2021-01-01 01:45:26+00:00
21028      21028      21028    NanouTheNomad  2021-01-01 00:24:34+00:00
```

```
      Tweet  Likes  Followers \
0  Per Alex: Approximately 95% of adults and 98% ...      1      3425
1  Big Tobacco's divestment from Quebec's Medicag...      0      66607
2  @brianlilley COVID Vaccinations are the most u...      0         17
3  @MerlinofCanada @JustinTrudeau In Alberta it's...      1        331
4  Big Tobacco&#039;s divestment from Quebec&...      0         29
...
21024  @Lin_Manuel It certainly been very hard year. ...      0         25
21025  NOV/DEC \n-Covid cases shot back up as lead...      1        749
21026  @DeNovo_Fatima @robyn_s_lee Translation: since...      3      14637
21027  Adults with disabilities in Hamilton County ar...      2         602
21028  In lieu of traditional vaccine selfie, I offer...     14       1965
```

```
      Friends  Reply  Retweet  Quote \
0          3949      0         0      0
1           48      0         0      0
2          189      0         0      0
3          598      2         0      0
4           45      0         0      0
...
21024      256      0         0      0
21025      789      0         0      0
21026      163      0         0      0
21027      139      0         0      0
21028      970      0         1      0
```

```
      Hashtags \
0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
...
21024      NaN
```

```

21025 NaN
21026 NaN
21027 NaN
21028 ['Hamilton', 'COVID19', 'science', 'publicheal...

```

	Processed Tweets	Date \
0	per alex approximately adult adult quebec rece...	2022-12-30
1	big tobacco divestment quebec medicago step ri...	2022-12-30
2	covid vaccination unsafe ineffective vaccine e...	2022-12-30
3	alberta ' ppl rollout edmonton walk pharmacy g...	2022-12-30
4	big tobaccoamps divestment quebecamps medicago...	2022-12-30
...	...	...
21024	manuel certainly hard year literally started d...	2021-01-01
21025	novdec covid case shot back leadership drag...	2021-01-01
21026	fatima slee translation since quebec ontario f...	2021-01-01
21027	adult disability hamilton county eligible rece...	2021-01-01
21028	lieu traditional vaccine selfie offer gif song...	2021-01-01

	Polarity	Subjectivity
0	0.100000	0.300000
1	0.142857	0.317857
2	0.350000	0.575000
3	0.516667	0.916667
4	0.142857	0.317857
...	...	...
21024	-0.291667	0.541667
21025	0.000000	0.044231
21026	-0.033333	0.333333
21027	0.250000	0.650000
21028	-0.500000	0.875000

[21029 rows x 16 columns]

```
[12]: df2.to_csv('~\Desktop\Twitter_New\date_polarity_subjectivity_30K_vaccine_21_22.
      ↪csv')
```

```
[13]: # Individual influence score
      # It is computed based on the number of followers
      def IndividualInfluence(n_followers):
          if n_followers > 10000:
              x = 4
          elif n_followers > 1000:
              x = 3
          elif n_followers > 100:
              x = 2
          elif n_followers > 10:
              x = 1
```

```
else:
    x = 0
return x
```

```
[14]: # Response effect score
# It is computed based on the number of likes, replies, and retweets
def ResponseEffect(n_likes, n_replies, n_retweets):
    if n_likes >100:
        a=3
    elif n_likes >10:
        a=2
    elif n_likes >0:
        a=1
    else:
        a=0
    if n_replies >100:
        b=3
    elif n_replies >10:
        b=2
    elif n_replies >0:
        b=1
    else:
        b=0
    if n_retweets >100:
        c=3
    elif n_retweets >10:
        c=2
    elif n_retweets >0:
        c=1
    else:
        c=0
    return (a+b+c)/3

#Separate function for each category
def LikeEffect(n_likes):
    if n_likes >100:
        a=3
    elif n_likes >10:
        a=2
    elif n_likes >0:
        a=1
    else:
        a=0
    return a

def ReplyEffect(n_replies):
    if n_replies >100:
```

```

        b=3
    elif n_replies >10:
        b=2
    elif n_replies >0:
        b=1
    else:
        b=0
    return b

def RetweetEffect(n_retweets):
    if n_retweets >100:
        c=3
    elif n_retweets >10:
        c=2
    elif n_retweets >0:
        c=1
    else:
        c=0
    return c

```

```

[15]: # assign subjectivity level a category
def SubjectiveLevel(subjectivity_score):
    if subjectivity_score <0.25:
        d=1
    elif subjectivity_score <0.5:
        d=2
    elif subjectivity_score <0.75:
        d=3
    else:
        d=4
    return d

```

```

[16]: df3 = pd.read_csv('~/Desktop/Twitter_New/
    ↪date_polarity_subjectivity_30K_vaccine_21_22.csv')
df3['Individual Influence'] = df3['Followers'].apply(IndividualInfluence)
df3['Like Effect'] = df3["Likes"].apply(LikeEffect)
df3['Reply Effect'] = df3["Reply"].apply(ReplyEffect)
df3['Retweet Effect'] = df3["Retweet"].apply(RetweetEffect)
df3['Response Effect'] = df3.apply(lambda df3:
    ↪ResponseEffect(df3["Likes"],df3["Reply"],df3["Retweet"]),axis=1)
df3['Subjective Level'] = df3["Subjectivity"].apply(SubjectiveLevel)
df3
df3.to_csv('~/Desktop/Twitter_New/individualinf_repsonseeff_30K_vaccine_21_22.
    ↪csv')

```

```

[17]: df3

```

```
[17]:      Unnamed: 0.2  Unnamed: 0.1  Unnamed: 0      User \
0          0          0          0      Treep52
1          1          1          1  DrGurdeepParhar
2          2          2          2  stopthegaslight
3          3          3          3      Daisy381
4          4          4          4    RadarsScience
...
21024      21024      21024      21024      nika180479
21025      21025      21025      21025    forjustminutes
21026      21026      21026      21026  profamirattaran
21027      21027      21027      21027  EricBenninghoff
21028      21028      21028      21028    NanouTheNomad
```

```
      Date Created \
0      2022-12-30 22:51:15+00:00
1      2022-12-30 21:10:00+00:00
2      2022-12-30 18:49:59+00:00
3      2022-12-30 17:31:01+00:00
4      2022-12-30 16:20:06+00:00
...
21024  2021-01-01 06:18:16+00:00
21025  2021-01-01 04:11:23+00:00
21026  2021-01-01 02:05:23+00:00
21027  2021-01-01 01:45:26+00:00
21028  2021-01-01 00:24:34+00:00
```

```
      Tweet  Likes  Followers \
0      Per Alex: Approximately 95% of adults and 98% ...      1      3425
1      Big Tobacco's divestment from Quebec's Medicag...      0      66607
2      @brianlilley COVID Vaccinations are the most u...      0        17
3      @MerlinofCanada @JustinTrudeau In Alberta it's...      1       331
4      Big Tobacco&#039;s divestment from Quebec&...      0        29
...
21024  @Lin_Manuel It certainly been very hard year. ...      0        25
21025  NOV/DEC \n-Covid cases shot back up as lead...      1       749
21026  @DeNovo_Fatima @robyn_s_lee Translation: since...      3     14637
21027  Adults with disabilities in Hamilton County ar...      2       602
21028  In lieu of traditional vaccine selfie, I offer...     14     1965
```

```
      Friends  Reply  ...      Processed Tweets \
0      3949      0      ...  per alex approximately adult adult quebec rece...
1      48      0      ...  big tobacco divestment quebec medicago step ri...
2      189      0      ...  covid vaccination unsafe ineffective vaccine e...
3      598      2      ...  alberta ' ppl rollout edmonton walk pharmacy g...
4      45      0      ...  big tobacco&#039;s divestment quebec&#039;s medicago...
...
21024      256      0      ...  manuel certainly hard year literally started d...
```

21025	789	0	...	novdec covid case shot back leadership drag...
21026	163	0	...	fatima slee translation since quebec ontario f...
21027	139	0	...	adult disability hamilton county eligible rece...
21028	970	0	...	lieu traditional vaccine selfie offer gif song...

	Date	Polarity	Subjectivity	Individual Influence	Like Effect	\
0	2022-12-30	0.100000	0.300000		3	1
1	2022-12-30	0.142857	0.317857		4	0
2	2022-12-30	0.350000	0.575000		1	0
3	2022-12-30	0.516667	0.916667		2	1
4	2022-12-30	0.142857	0.317857		1	0
...	...	...	...	...	...	
21024	2021-01-01	-0.291667	0.541667		1	0
21025	2021-01-01	0.000000	0.044231		2	1
21026	2021-01-01	-0.033333	0.333333		4	1
21027	2021-01-01	0.250000	0.650000		2	1
21028	2021-01-01	-0.500000	0.875000		3	2

	Reply Effect	Retweet Effect	Response Effect	Subjective Level
0	0	0	0.333333	2
1	0	0	0.000000	2
2	0	0	0.000000	3
3	1	0	0.666667	4
4	0	0	0.000000	2
...	...	...	...	...
21024	0	0	0.000000	3
21025	0	0	0.333333	1
21026	0	0	0.333333	2
21027	0	0	0.333333	3
21028	0	1	1.000000	4

[21029 rows x 23 columns]

```
[18]: # Hashtag Analysis
hashtags = []
hashtag_pattern = re.compile(r"#[a-zA-Z]+")
hashtag_matches = list(df3['Tweet'].apply(hashtag_pattern.findall))
hashtag_dict = {}
for match in hashtag_matches:
    for singlematch in match:
        if singlematch not in hashtag_dict.keys():
            hashtag_dict[singlematch] = 1
        else:
            hashtag_dict[singlematch] = hashtag_dict[singlematch]+1

hashtag_ordered_list = sorted(hashtag_dict.items(), key=lambda x:x[1])
hashtag_ordered_list = hashtag_ordered_list[::-1]
```

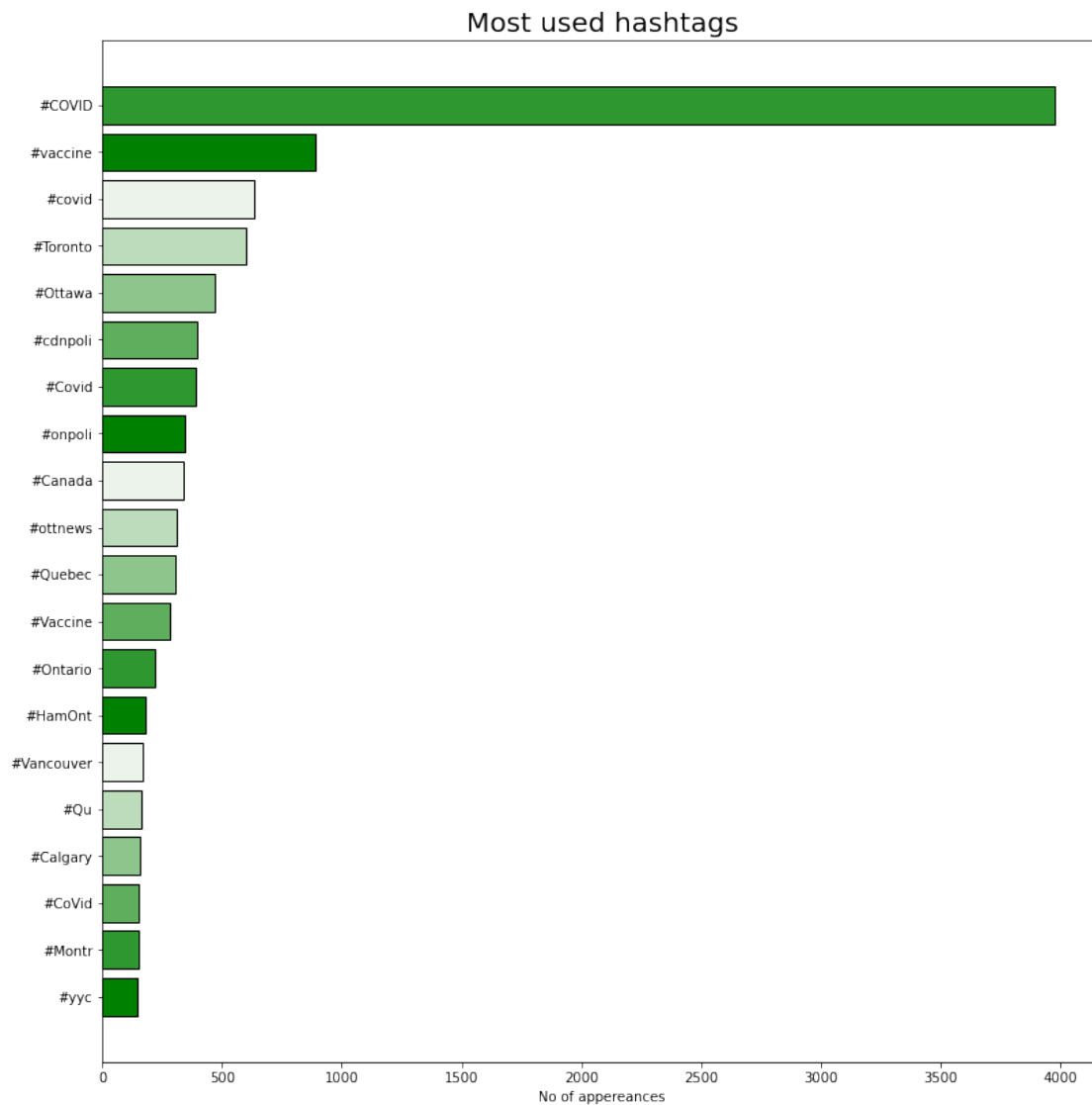


```

#Separating the hashtags and their values into two different lists
hashtag_ordered_values = []
hashtag_ordered_keys = []
#Pick the 20 most used hashtags to plot
for item in hashtag_ordered_list[0:20]:
    hashtag_ordered_keys.append(item[0])
    hashtag_ordered_values.append(item[1])

fig, ax = plt.subplots(figsize = (12,12))
y_pos = np.arange(len(hashtag_ordered_keys))
ax.barh(y_pos ,list(hashtag_ordered_values)[::-1], align='center', color = sns.
    light_palette("green", reverse=True), edgecolor = 'black', linewidth=1)
ax.set_yticks(y_pos)
ax.set_yticklabels(list(hashtag_ordered_keys)[::-1])
ax.set_xlabel("No of appereances")
ax.set_title("Most used hashtags", fontsize = 20)
plt.tight_layout(pad=3)
plt.show()

```



[ ]: