

Twitter Data Other Interventions 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: lxml in ./opt/anaconda3/lib/python3.9/site-

packages (from snsrape) (4.8.0)
 Requirement already satisfied: beautifulsoup4 in
 ./opt/anaconda3/lib/python3.9/site-packages (from snsrape) (4.11.1)
 Requirement already satisfied: filelock in ./opt/anaconda3/lib/python3.9/site-
 packages (from snsrape) (3.6.0)
 Requirement already satisfied: requests[socks] in
 ./opt/anaconda3/lib/python3.9/site-packages (from snsrape) (2.27.1)
 Requirement already satisfied: soupsieve>1.2 in
 ./opt/anaconda3/lib/python3.9/site-packages (from beautifulsoup4->snsrape)
 (2.3.1)
 Requirement already satisfied: idna<4,>=2.5 in
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)
 (3.3)
 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: urllib3<1.27,>=1.21.1 in
 ./opt/anaconda3/lib/python3.9/site-packages (from requests[socks]->snsrape)
 (1.26.9)
 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: 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: 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: click in ./opt/anaconda3/lib/python3.9/site-
 packages (from nltk>=3.1->textblob) (8.0.4)
 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: 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: pyyaml>=5.3.1 in
 ./opt/anaconda3/lib/python3.9/site-packages (from dask) (6.0)

```
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: cloudpickle>=1.1.1 in
./opt/anaconda3/lib/python3.9/site-packages (from dask) (2.0.0)
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 (mask OR social distancing OR isolation OR quarantine)) 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 'snscrate.modules.twitter.TwitterSearchScraper'>
```

```
[2]: Tweet(url='https://twitter.com/pouchbaby/status/1608748144480878597',
date=datetime.datetime(2022, 12, 30, 8, 53, 41, tzinfo=datetime.timezone.utc),
rawContent='Nice plain language explanation for the risk of not masking. In my
house now each family member (except me) has had Covid at different times.
Nobody else got it from them here or anywhere we know of. We do mask. We
quarantine. We rapid test. Out After 13 days when negative again.',
renderedContent='Nice plain language explanation for the risk of not masking. In
my house now each family member (except me) has had Covid at different times.
Nobody else got it from them here or anywhere we know of. We do mask. We
quarantine. We rapid test. Out After 13 days when negative again.',
id=1608748144480878597, user=User(username='pouchbaby', id=22322274,
displayname='Anita Szigeti', rawDescription="I'm a lawyer and a Mom. Obsessed
with coffee and baseball. Superpowers: fluent in Hungarian & dog breeds. Also
fluent in mental health law, civil and criminal.", renderedDescription="I'm a
lawyer and a Mom. Obsessed with coffee and baseball. Superpowers: fluent in
Hungarian & dog breeds. Also fluent in mental health law, civil and criminal.",
descriptionLinks=None, verified=False, created=datetime.datetime(2009, 3, 1, 2,
18, 30, tzinfo=datetime.timezone.utc), followersCount=3313, friendsCount=90,
statusesCount=12809, favouritesCount=16739, listedCount=17, mediaCount=1111,
location='Toronto, Ontario, Canada', protected=False,
link=TextLink(text='anitaszigeti.wordpress.com',
url='https://anitaszigeti.wordpress.com/', tcourl='https://t.co/QONpvWvFYe',
indices=(0, 23)), profileImageUrl='https://pbs.twimg.com/profile_images/16269223
49151895553/-U3E9geM_normal.jpg',
profileBannerUrl='https://pbs.twimg.com/profile_banners/22322274/1593958647',
```

```

label=None), replyCount=0, retweetCount=0, likeCount=3, quoteCount=0,
conversationId=1608748144480878597, lang='en', source='<a
href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>', sourceUrl='http://twitter.com/download/iphone',
sourceLabel='Twitter for iPhone', links=None, media=None, retweetedTweet=None, q
uotedTweet=Tweet(url='https://twitter.com/JohnBrownsBody2/status/160853025187706
0608', date=datetime.datetime(2022, 12, 29, 18, 27, 51,
tzinfo=datetime.timezone.utc), rawContent='Say you go to a party with your
friends. No one masks. \n\nYou go about your life for the next few days.
Maskless, of course.\n\nYou wake up 5 days later with a sore throat. You test
positive. You bounce back in a week. \n\nThat's your experience...but it goes far
beyond you. 1/10', renderedContent='Say you go to a party with your friends. No
one masks. \n\nYou go about your life for the next few days. Maskless, of
course.\n\nYou wake up 5 days later with a sore throat. You test positive. You
bounce back in a week. \n\nThat's your experience...but it goes far beyond you.
1/10', id=1608530251877060608, user=User(username='JohnBrownsBody2',
id=231525540, displayname='#SaveSilwan #SaveSheikhJarrah', rawDescription='The
views espoused here do not represent those of my employer. ML. he/him
#FreePalestine', renderedDescription='The views espoused here do not represent
those of my employer. ML. he/him #FreePalestine', descriptionLinks=None,
verified=False, created=datetime.datetime(2010, 12, 28, 19, 34, 22,
tzinfo=datetime.timezone.utc), followersCount=4779, friendsCount=4935,
statusesCount=48664, favouritesCount=77566, listedCount=84, mediaCount=2560,
location='Stolen Land', protected=False, link=None, profileImageUrl='https://pbs
.twimg.com/profile_images/1625746337886085121/pzC3ugJh_normal.jpg',
profileBannerUrl='https://pbs.twimg.com/profile_banners/231525540/1548288831',
label=None), replyCount=120, retweetCount=2265, likeCount=7363, quoteCount=284,
conversationId=1608530251877060608, lang='en', source='<a
href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>', sourceUrl='http://twitter.com/download/iphone',
sourceLabel='Twitter for iPhone', links=None, media=None, retweetedTweet=None,
quotedTweet=None, inReplyToTweetId=None, inReplyToUser=None,
mentionedUsers=None, coordinates=None, place=None, hashtags=None, cashtags=None,
card=None, viewCount=732128, vibe=None), inReplyToTweetId=None,
inReplyToUser=None, mentionedUsers=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=None, viewCount=528, vibe=None)

```

```

[3]: data = []
# Using TwitterSearchScrapper to scrape data and append tweets to list
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper
    ('(mask OR distancing OR isolation) 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>45000:
    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_interventions_21_22.csv')
df

```

```

[3]:

```

	User	Date Created	\
0	HSans	2022-12-30 23:42:46+00:00	
1	Randi_RLB	2022-12-30 23:11:27+00:00	
2	Golden_Pup	2022-12-30 22:51:45+00:00	
3	EnigmaMachine8	2022-12-30 19:14:24+00:00	
4	Cognisant2000	2022-12-30 18:59:47+00:00	
...	
35627	SanjayB_yyz	2021-01-01 01:28:41+00:00	
35628	PeteJolicoeur	2021-01-01 01:04:25+00:00	
35629	OttawaDreamCafe	2021-01-01 00:43:30+00:00	
35630	UberTumbleweed	2021-01-01 00:23:29+00:00	
35631	kool1051	2021-01-01 00:05:29+00:00	

	Tweet	Likes	Followers	\
0	Here in #Québec, #Canada we've had this rule i...	0	2335	
1	@DrAGrace_cyhtt "Alberta's education minister ...	0	1531	
2	November 18\n\nJust three days after defying a...	196	39775	
3	@shogun17761 @Katheri58113861 @kirkJK04 @jeanp...	0	2957	
4	@GMMalliet Agree. We always wear a mask in enc...	0	2037	
...	
35627	#CNNNYE Great way to end what has been a rough...	0	5	
35628	To this day I cannot tell you what Rip Hamilto...	0	596	
35629	BREAKING NEWS: #bonneannee2021 #NewYear ELVIS ...	3	583	
35630	Hey @sobey's how come employees at your Newcast...	0	160	
35631	@big_catsf @PattyArquette Nor anywhere in Queb...	0	562	

	Friends	Reply	Retweet	Quote	\
0	4527	0	0	0	
1	2204	0	0	0	
2	584	3	23	1	
3	3137	0	0	0	
4	1713	0	0	0	
...	
35627	47	0	0	0	
35628	1198	0	0	0	

35629	2354	0	3	0
35630	202	1	0	0
35631	4955	0	0	0

	Hashtags
0	[Québec, Canada, SpecialCircumstances, WeShall...
1	None
2	None
3	None
4	None
...	...
35627	[CNNNYE]
35628	None
35629	[bonneannee2021, NewYear, Ottawa, mask]
35630	None
35631	None

[35632 rows x 10 columns]

```
[6]: #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+:\/\/\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
```

```
[8]: #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_interventions_21_22.csv')
df['Processed Tweets'] = df['Tweet'].apply(ProcessedTweets)
df.to_csv('~\Desktop\Twitter_New\prcessed_tweet_30K_interventions_21_22.csv')
df
```

```
[8]:
```

	Unnamed: 0	User	Date Created	\
0	0	HSans	2022-12-30 23:42:46+00:00	
1	1	Randi_RLB	2022-12-30 23:11:27+00:00	
2	2	Golden_Pup	2022-12-30 22:51:45+00:00	
3	3	EnigmaMachine8	2022-12-30 19:14:24+00:00	
4	4	Cognisant2000	2022-12-30 18:59:47+00:00	
...	
35627	35627	SanjayB_yyz	2021-01-01 01:28:41+00:00	
35628	35628	PeteJolicoeur	2021-01-01 01:04:25+00:00	
35629	35629	OttawaDreamCafe	2021-01-01 00:43:30+00:00	
35630	35630	UberTumbleweed	2021-01-01 00:23:29+00:00	
35631	35631	kool1051	2021-01-01 00:05:29+00:00	

	Tweet	Likes	Followers	\
0	Here in #Québec, #Canada we've had this rule i...	0	2335	
1	@DrAGrace_cyhtt "Alberta's education minister ...	0	1531	
2	November 18\n\nJust three days after defying a...	196	39775	
3	@shogun17761 @Katheri58113861 @kirkJK04 @jeanp...	0	2957	
4	@GMMalliet Agree. We always wear a mask in enc...	0	2037	
...	
35627	#CNNNYE Great way to end what has been a rough...	0	5	
35628	To this day I cannot tell you what Rip Hamilto...	0	596	
35629	BREAKING NEWS: #bonneannee2021 #NewYear ELVIS ...	3	583	
35630	Hey @sobey's how come employees at your Newcast...	0	160	
35631	@big_catsf @PattyArquette Nor anywhere in Queb...	0	562	

	Friends	Reply	Retweet	Quote	\
0	4527	0	0	0	
1	2204	0	0	0	
2	584	3	23	1	
3	3137	0	0	0	
4	1713	0	0	0	
...	
35627	47	0	0	0	
35628	1198	0	0	0	
35629	2354	0	3	0	
35630	202	1	0	0	

```
35631      4955      0      0      0
```

```

                                Hashtags \
0      ['Québec', 'Canada', 'SpecialCircumstances', '...'
1                                NaN
2                                NaN
3                                NaN
4                                NaN
...
35627                                ...
                                ['CNNNYE']
35628                                NaN
35629      ['bonneannee2021', 'NewYear', 'Ottawa', 'mask']
35630                                NaN
35631                                NaN

```

```

                                Processed Tweets
0      weve rule place couple month like almost every...
1      cyhtt alberta education minister intervene w m...
2      november three day defying sanity pleading ont...
3      youre trying convince person car take mask eve...
4      agree always wear mask enclosed space one time...
...
35627      great way end rough year hanging amp guy aweso...
35628      day tell rip hamilton look like without includ...
35629      breaking news elvis seen walking social distan...
35630      hey come employee newcastle store edmonton wor...
35631      catsf anywhere quebec im sure even get door li...

```

```
[35632 rows x 12 columns]
```

```

[9]: #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
```



```
[10]: df['Date'] = df['Date Created'].apply(ConvertDate)
df.to_csv('~/Desktop/Twitter_New/prcessed_tweet_date_30K_interventions_21_22.
↳csv')
```

```
[11]: df2 = pd.read_csv('~/Desktop/Twitter_New/
↳prcessed_tweet_date_30K_interventions_21_22.csv')
df2
```

```
[11]:      Unnamed: 0.1  Unnamed: 0      User      Date Created \
0                0          0      HSans  2022-12-30 23:42:46+00:00
1                1          1    Randi_RLB  2022-12-30 23:11:27+00:00
2                2          2    Golden_Pup  2022-12-30 22:51:45+00:00
3                3          3  EnigmaMachine8  2022-12-30 19:14:24+00:00
4                4          4   Cognisant2000  2022-12-30 18:59:47+00:00
...
35627          35627      35627    SanjayB_yyz  2021-01-01 01:28:41+00:00
35628          35628      35628   PeteJolicoeur  2021-01-01 01:04:25+00:00
35629          35629      35629  OttawaDreamCafe  2021-01-01 00:43:30+00:00
35630          35630      35630   UberTumbleweed  2021-01-01 00:23:29+00:00
35631          35631      35631      kool1051  2021-01-01 00:05:29+00:00
```

```
      Tweet  Likes  Followers \
0  Here in #Québec, #Canada we've had this rule i...      0      2335
1  @DrAGrace_cyhtt "Alberta's education minister ...      0      1531
2  November 18\n\nJust three days after defying a...    196     39775
3  @shogun17761 @Katheri58113861 @kirkJK04 @jeanp...      0      2957
4  @GMMalliet Agree. We always wear a mask in enc...      0      2037
...
35627  #CNNNYE Great way to end what has been a rough...      0          5
35628  To this day I cannot tell you what Rip Hamilto...      0      596
35629  BREAKING NEWS: #bonneannee2021 #NewYear ELVIS ...      3      583
35630  Hey @sobey's how come employees at your Newcast...      0      160
35631  @big_catsf @PattyArquette Nor anywhere in Queb...      0      562
```

```
      Friends  Reply  Retweet  Quote \
0          4527      0          0      0
1          2204      0          0      0
2           584      3         23      1
3          3137      0          0      0
4          1713      0          0      0
...
35627         47      0          0      0
35628        1198      0          0      0
35629        2354      0          3      0
35630         202      1          0      0
35631        4955      0          0      0
```

	Hashtags \
0	['Québec', 'Canada', 'SpecialCircumstances', '...
1	NaN
2	NaN
3	NaN
4	NaN
...	...
35627	['CNNNYE']
35628	NaN
35629	['bonneannee2021', 'NewYear', 'Ottawa', 'mask']
35630	NaN
35631	NaN

	Processed Tweets	Date
0	weve rule place couple month like almost every...	2022-12-30
1	cyhtt alberta education minister intervene w m...	2022-12-30
2	november three day defying sanity pleading ont...	2022-12-30
3	youre trying convince person car take mask eve...	2022-12-30
4	agree always wear mask enclosed space one time...	2022-12-30
...
35627	great way end rough year hanging amp guy aweso...	2021-01-01
35628	day tell rip hamilton look like without includ...	2021-01-01
35629	breaking news elvis seen walking social distan...	2021-01-01
35630	hey come employee newcastle store edmonton wor...	2021-01-01
35631	catsf anywhere quebec im sure even get door li...	2021-01-01

[35632 rows x 14 columns]

```
[12]: # 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
```

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

```
[13]:      Unnamed: 0.1  Unnamed: 0      User      Date Created \
0              0          0      HSans  2022-12-30 23:42:46+00:00
1              1          1    Randi_RLB  2022-12-30 23:11:27+00:00
2              2          2   Golden_Pup  2022-12-30 22:51:45+00:00
3              3          3 EnigmaMachine8  2022-12-30 19:14:24+00:00
4              4          4  Cognisant2000  2022-12-30 18:59:47+00:00
...
35627          35627      35627    SanjayB_yyz  2021-01-01 01:28:41+00:00
35628          35628      35628  PeteJolicoeur  2021-01-01 01:04:25+00:00
35629          35629      35629 OttawaDreamCafe  2021-01-01 00:43:30+00:00
35630          35630      35630  UberTumbleweed  2021-01-01 00:23:29+00:00
35631          35631      35631     kool1051  2021-01-01 00:05:29+00:00
```

```
      Tweet  Likes  Followers \
0  Here in #Québec, #Canada we've had this rule i...      0      2335
1  @DrAGrace_cyhtt "Alberta's education minister ...      0      1531
2  November 18\n\nJust three days after defying a...    196     39775
3  @shogun17761 @Katheri58113861 @kirkJK04 @jeanp...      0      2957
4  @GMMalliet Agree. We always wear a mask in enc...      0      2037
...
35627 #CNNNYE Great way to end what has been a rough...      0         5
35628 To this day I cannot tell you what Rip Hamilto...      0         596
35629 BREAKING NEWS: #bonneannee2021 #NewYear ELVIS ...      3         583
35630 Hey @sobey's how come employees at your Newcast...      0         160
35631 @big_catsf @PattyArquette Nor anywhere in Queb...      0         562
```

```
      Friends  Reply  Retweet  Quote \
0         4527      0         0      0
1         2204      0         0      0
2          584      3         23      1
3         3137      0         0      0
4         1713      0         0      0
...
35627         47      0         0      0
35628        1198      0         0      0
35629        2354      0         3      0
35630         202      1         0      0
35631        4955      0         0      0
```

```
      Hashtags \
0  ['Québec', 'Canada', 'SpecialCircumstances', '...'
1                                     NaN
```

```

2
3
4
...
35627      ['CNNNYE']
35628      NaN
35629      ['bonneannee2021', 'NewYear', 'Ottawa', 'mask']
35630      NaN
35631      NaN

```

```

                                Processed Tweets      Date \
0      weve rule place couple month like almost every... 2022-12-30
1      cyhtt alberta education minister intervene w m... 2022-12-30
2      november three day defying sanity pleading ont... 2022-12-30
3      youre trying convince person car take mask eve... 2022-12-30
4      agree always wear mask enclosed space one time... 2022-12-30
...
35627 great way end rough year hanging amp guy aweso... 2021-01-01
35628 day tell rip hamilton look like without includ... 2021-01-01
35629 breaking news elvis seen walking social distan... 2021-01-01
35630 hey come employee newcastle store edmonton wor... 2021-01-01
35631 catsf anywhere quebec im sure even get door li... 2021-01-01

```

```

      Polarity  Subjectivity
0      0.500000      0.500000
1     -0.117262      0.577381
2      0.000000      0.000000
3      0.000000      1.000000
4      0.000000      0.000000
...
35627  0.600000      0.830000
35628  0.000000      0.000000
35629  0.116667      0.133333
35630  0.000000      0.000000
35631  0.500000      0.888889

```

[35632 rows x 16 columns]

```
[14]: df2.to_csv('~\Desktop\Twitter_New/
         ↳date_polarity_subjectivity_30K_interventions_21_22.csv')
```

```
[15]: # 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

```

```

[16]: # 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

```

```

[17]: # 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

```

```

[18]: df3 = pd.read_csv('~\Desktop\Twitter_New/
    ↪date_polarity_subjectivity_30K_interventions_21_22.csv')
df3['Like Effect'] = df3["Likes"].apply(LikeEffect)
df3['Reply Effect'] = df3["Reply"].apply(ReplyEffect)
df3['Retweet Effect'] = df3["Retweet"].apply(RetweetEffect)
df3['Individual Influence'] = df3['Followers'].apply(IndividualInfluence)
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_interventions_21_22.csv')
```

```
[19]: df3
```

```
[19]:      Unnamed: 0.2  Unnamed: 0.1  Unnamed: 0      User \
0              0              0              0      HSans
1              1              1              1      Randi_RLB
2              2              2              2      Golden_Pup
3              3              3              3      EnigmaMachine8
4              4              4              4      Cognisant2000
...
35627          35627          35627          35627      SanjayB_yyz
35628          35628          35628          35628      PeteJolicoeur
35629          35629          35629          35629      OttawaDreamCafe
35630          35630          35630          35630      UberTumbleweed
35631          35631          35631          35631      kool1051

      Date Created \
0      2022-12-30 23:42:46+00:00
1      2022-12-30 23:11:27+00:00
2      2022-12-30 22:51:45+00:00
3      2022-12-30 19:14:24+00:00
4      2022-12-30 18:59:47+00:00
...
35627      2021-01-01 01:28:41+00:00
35628      2021-01-01 01:04:25+00:00
35629      2021-01-01 00:43:30+00:00
35630      2021-01-01 00:23:29+00:00
35631      2021-01-01 00:05:29+00:00

      Tweet  Likes  Followers \
0      Here in #Québec, #Canada we've had this rule i...      0      2335
1      @DrAGrace_cyhtt "Alberta's education minister ...      0      1531
2      November 18\n\nJust three days after defying a...      196      39775
3      @shogun17761 @Katheri58113861 @kirkJK04 @jeanp...      0      2957
4      @GMMalliet Agree. We always wear a mask in enc...      0      2037
...
35627      #CNNNYE Great way to end what has been a rough...      0      5
35628      To this day I cannot tell you what Rip Hamilto...      0      596
35629      BREAKING NEWS: #bonneannee2021 #NewYear ELVIS ...      3      583
35630      Hey @sobey's how come employees at your Newcast...      0      160
35631      @big_catsf @PattyArquette Nor anywhere in Queb...      0      562

      Friends  Reply  ...      Processed Tweets \
0      4527      0      ...      weve rule place couple month like almost every...
1      2204      0      ...      cyhtt alberta education minister intervene w m...
```

2	584	3	...	november three day defying sanity pleading ont...
3	3137	0	...	youre trying convince person car take mask eve...
4	1713	0	...	agree always wear mask enclosed space one time...
...
35627	47	0	...	great way end rough year hanging amp guy aweso...
35628	1198	0	...	day tell rip hamilton look like without includ...
35629	2354	0	...	breaking news elvis seen walking social distan...
35630	202	1	...	hey come employee newcastle store edmonton wor...
35631	4955	0	...	catsf anywhere quebec im sure even get door li...

	Date	Polarity	Subjectivity	Like	Effect	Reply	Effect \
0	2022-12-30	0.500000	0.500000		0		0
1	2022-12-30	-0.117262	0.577381		0		0
2	2022-12-30	0.000000	0.000000		3		1
3	2022-12-30	0.000000	1.000000		0		0
4	2022-12-30	0.000000	0.000000		0		0
...
35627	2021-01-01	0.600000	0.830000		0		0
35628	2021-01-01	0.000000	0.000000		0		0
35629	2021-01-01	0.116667	0.133333		1		0
35630	2021-01-01	0.000000	0.000000		0		1
35631	2021-01-01	0.500000	0.888889		0		0

	Retweet	Effect	Individual	Influence	Response	Effect	Subjective	Level
0		0			3	0.000000		3
1		0			3	0.000000		3
2		2			4	2.000000		1
3		0			3	0.000000		4
4		0			3	0.000000		1
...
35627		0			0	0.000000		4
35628		0			2	0.000000		1
35629		1			2	0.666667		1
35630		0			2	0.333333		1
35631		0			2	0.000000		4

[35632 rows x 23 columns]

```
[20]: # 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
```



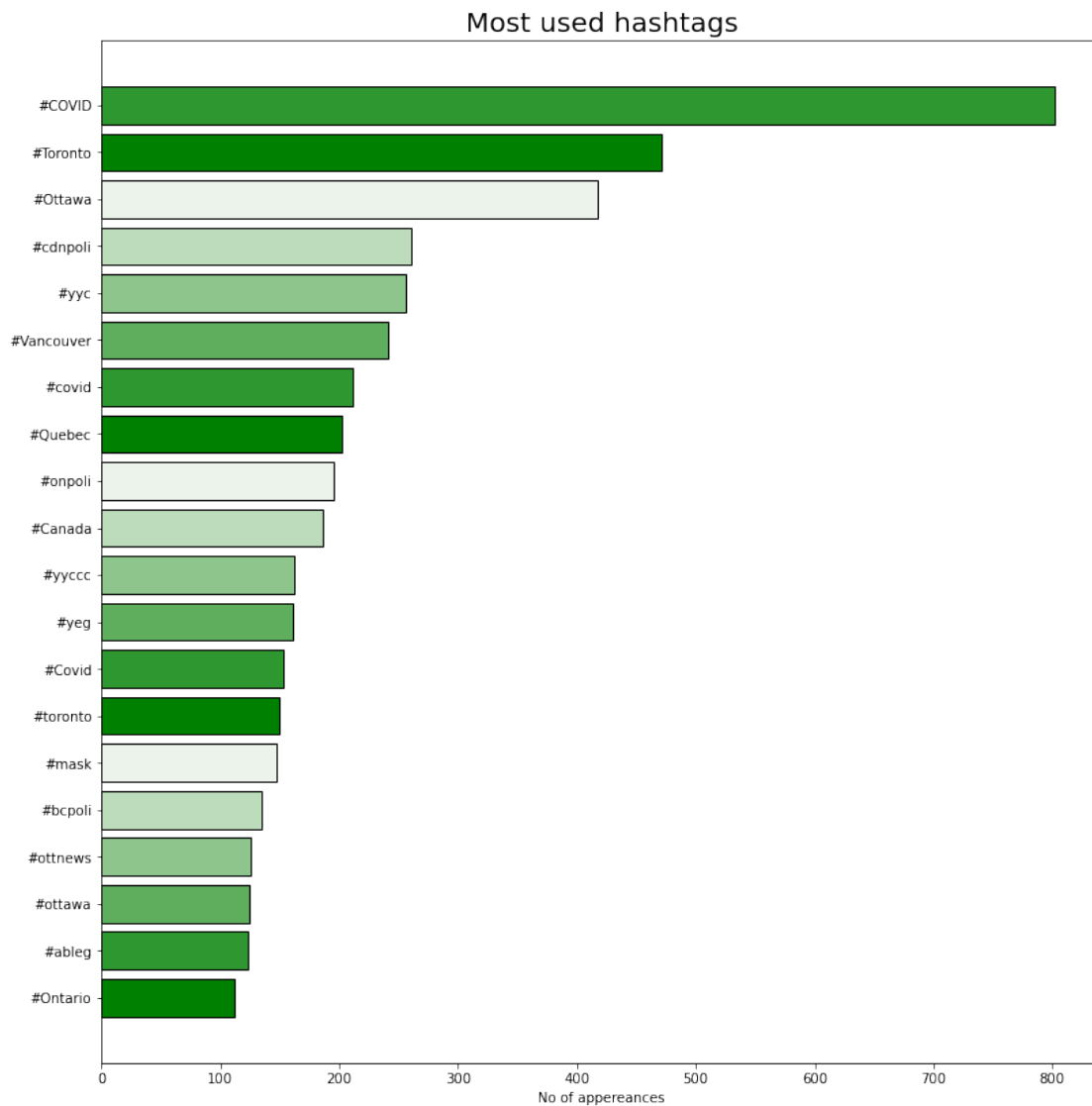
```

        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()

```



[]: