By: MUSKAN GUPTA , SUMMER INTERNSHIP PROJECT

Covid 19 Tweets Data Exploration and Sentiment analysis

Task

Text Preprocessing

Sentiment Analysis

Keyword Extraction

Entity Extraction

In [5]:

```python
### EDA Pkgs
import pandas as pd
```

In [7]:

```python
# Data Viz Pkg
import matplotlib.pyplot as plt
import seaborn as sns
```

In [8]:

```python
# Hide warnings
import warnings
warnings.filterwarnings('ignore')
```

In [11]:

```python
# Load Dataset
df = pd.read_csv("/home/muskangupta/Downloads/covid19_tweets.csv")
```

```
# Preview
df.head()
```

Out[12]:

	user_name	user_location	user_description	user_created	user_followers	user_friends	us
0	ᐯˀꙇ🌑ｌℂ↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	
1	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	
2	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[]_[] #Cavs ...	2019-03-07 01:45:06	197	987	
4	DIPR-J&K	Jammu and Kashmir	🖊Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	

Text preprocessing

In [13]:

```
# Check Columns
df.columns
```

Out[13]:

```
Index(['user_name', 'user_location', 'user_description', 'user_create
d',
       'user_followers', 'user_friends', 'user_favourites', 'user_veri
fied',
       'date', 'text', 'hashtags', 'source', 'is_retweet'],
      dtype='object')
```

```python
# Datatype
df.dtypes
```

```
user_name          object
user_location      object
user_description   object
user_created       object
user_followers      int64
user_friends        int64
user_favourites     int64
user_verified        bool
date               object
text               object
hashtags           object
source             object
is_retweet           bool
dtype: object
```

```python
# Source/ Value Count/Distribution of the Sources
df['source'].unique()
```

```
array(['Twitter for iPhone', 'Twitter for Android', 'Twitter Web Ap
p',
       'Buffer', 'TweetDeck', 'Twitter for iPad', 'Africa Newsroom',
       'Blood Donors India', 'TweetCaster for Android',
       'Alexander Higgins', 'IFTTT', 'Hootsuite Inc.', 'Sprout Socia
l',
       'Sprinklr', 'assarofficial', 'IAMBLOG2TWITTER', 'CrowdControl
HQ',
       'COVID19-Updates', 'EveryoneSocial', 'Dynamic Signal', 'Insta
gram',
       'TweetCaster for iOS', 'GlobalPandemic.NET', 'Venrap Radio',
       'HeyOrca', 'Twitter for Advertisers', 'Paper.li',
       'Twitter Media Studio', 'Twitter for Mac', 'dlvr.it',
       'Cheap Bots, Done Quick!', 'Prof. Shanku', 'LaterMedia',
       'SEMrush Social Media Tool', 'Twitterrific for iOS',
       "Sebastian's Twitter Bot", 'Threader_client', 'COVID19Factoid
Bot',
       'PwC UK SMART', 'tweet pro stiff', 'UK COVID-19 Alerts',
```

```python
# Source/ Value Count/Distribution of the Sources
df['source'].value_counts()
```

```
Twitter Web App             56891
Twitter for Android         40179
Twitter for iPhone          35472
TweetDeck                    8543
Hootsuite Inc.               7321
                            ...
TtwTimes Top News               1
PRNewswire                      1
CIO Tech Asia Twitter           1
whatSaoCarlos                   1
InfoBlaze Southeast Asia        1
Name: source, Length: 610, dtype: int64
```

```python
# Plot the top value_counts
df['source'].value_counts().nlargest(30)
```

Out[23]:

```
Twitter Web App            56891
Twitter for Android        40179
Twitter for iPhone         35472
TweetDeck                   8543
Hootsuite Inc.              7321
Twitter for iPad            4336
Buffer                      2728
Sprout Social               1833
Instagram                   1759
IFTTT                       1545
dlvr.it                      681
GlobalPandemic.NET           679
LinkedIn                     628
COVID19-Updates              625
Twitter Media Studio         596
WordPress.com                537
HubSpot                      513
Sprinklr                     497
FS Poster                    414
Dynamic Signal               391
Resistbot Open Letters       390
Cheap Bots, Done Quick!      335
Paper.li                     322
Twitter for Advertisers      317
Blood Donors India           281
IAMBLOG2TWITTER              266
Alexander Higgins            225
Zoho Social                  224
The Social Jukebox           216
Fabrik.fm                    209
Name: source, dtype: int64
```
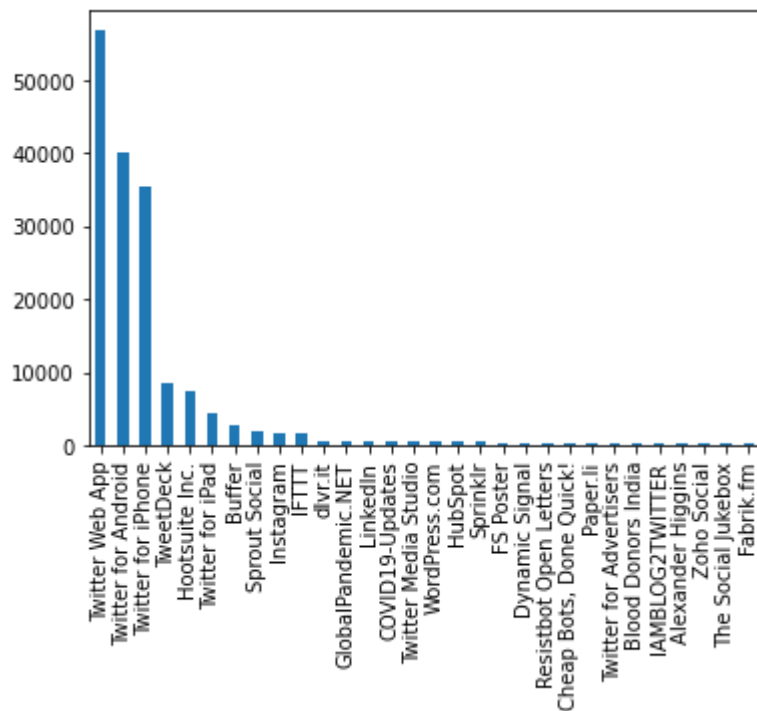
In [24]:

```python
# Plot the top value_counts
df['source'].value_counts().nlargest(30).plot(kind='bar')
```
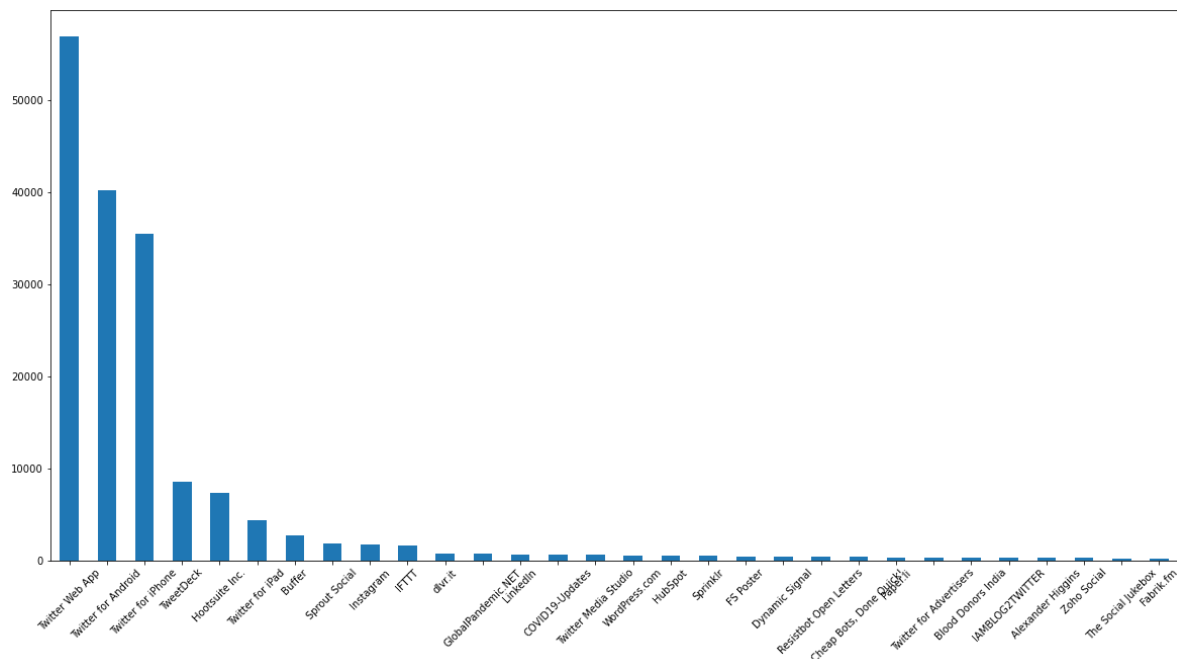
Out[24]:

<AxesSubplot:>

```python
# Plot the top value_counts
plt.figure(figsize=(20,10))
df['source'].value_counts().nlargest(30).plot(kind='bar')
plt.xticks(rotation=45)
plt.show()
```


Text Analysis of tweet

```
pip install neattext
```

```
Collecting neattext
  Downloading neattext-0.1.2-py3-none-any.whl (114 kB)
      |████████████████████████████████| 114 kB 1.2 MB/s eta 0:00:01
Installing collected packages: neattext
Successfully installed neattext-0.1.2
Note: you may need to restart the kernel to use updated packages.
```

In [27]:

```python
# Load Text Cleaning Package
import neattext.functions as nfx
```

In [28]:

```python
# Methods/Attrib
dir(nfx)
```

Out[28]:

```
['BTC_ADDRESS_REGEX',
 'CURRENCY_REGEX',
 'CURRENCY_SYMB_REGEX',
 'Counter',
 'DATE_REGEX',
 'EMAIL_REGEX',
 'EMOJI_REGEX',
 'HASTAG_REGEX',
 'MASTERCard_REGEX',
 'MD5_SHA_REGEX',
 'MOST_COMMON_PUNCT_REGEX',
 'NUMBERS_REGEX',
 'PHONE_REGEX',
 'PoBOX_REGEX',
 'SPECIAL_CHARACTERS_REGEX',
 'STOPWORDS',
 'STOPWORDS_de',
 'STOPWORDS_en',
```

In [29]:

```python
df['text'].iloc[2]
```

Out[29]:

'@diane3443 @wdunlap @realDonaldTrump Trump never once claimed #COVID1
9 was a hoax. We all claim that this effort to… https://t.co/Jkk8vHWHb
3' (https://t.co/Jkk8vHWHb3')

Noise

    remove mentions/userhandles
    remove hashtags
    urls
    emojis
    special char

```
df.head()
```

	user_name	user_location	user_description	user_created	user_followers	user_friends	us
0	ᐯ˘ἰ◉ʟℂ↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	
1	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	
2	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[]_[] #Cavs ...	2019-03-07 01:45:06	197	987	
4	DIPR-J&K	Jammu and Kashmir	🖊️Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	

```
df['text'].apply(nfx.extract_hashtags)
```

```
0                                      []
1                                      []
2                               [#COVID19]
3                               [#COVID19]
4          [#CoronaVirusUpdates, #COVID19]
                     ...
179103                       [#WearAMask]
179104                        [#COVID19]
179105                                []
179106                        [#COVID19]
179107                                []
Name: text, Length: 179108, dtype: object
```

```
df['extracted_hashtags'] = df['text'].apply(nfx.extract_hashtags)
```

```
df[['extracted_hashtags','hashtags']]
```

Out[33]:

	extracted_hashtags	hashtags
0	[]	NaN
1	[]	NaN
2	[#COVID19]	['COVID19']
3	[#COVID19]	['COVID19']
4	[#CoronaVirusUpdates, #COVID19]	['CoronaVirusUpdates', 'COVID19']
...	...	...
179103	[#WearAMask]	['WearAMask']
179104	[#COVID19]	['COVID19']
179105	[]	NaN
179106	[#COVID19]	['COVID19']
179107	[]	NaN

179108 rows × 2 columns

In [34]:

```
# Cleaning Text
df['clean_tweet'] = df['text'].apply(nfx.remove_hashtags)
```

```
df[['text','clean_tweet']]
```

	text	clean_tweet
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey @Yankees @YankeesPR and @MLB - wouldn't it...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	@diane3443 @wdunlap @realDonaldTrump Trump nev...
3	@brookbanktv The one gift #COVID19 has give me...	@brookbanktv The one gift has give me is an ...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel \n@kansa...
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks @IamOhmai for nominating me for the @WH...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol! https://t.c...
179105	@CTVNews A powerful painting by Juan Lucena. I...	@CTVNews A powerful painting by Juan Lucena. I...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for a...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop\n\n@SABCNews\n@Izinda...

179108 rows × 2 columns

```
df['clean_tweet'] = df['clean_tweet'].apply(lambda x: nfx.remove_userhandles(x))
```

In [37]:

```
df[['text','clean_tweet']]
```

Out[37]:

	text	clean_tweet
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and - wouldn't it have made more sen...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax. W...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciatio...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel \n ...
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challen...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol! https://t.c...
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena. It's a t...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for a...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop\n\n \n \n \n \n \n \n...

179108 rows × 2 columns

In [38]:

```
df['clean_tweet'].iloc[10]
```

Out[38]:

'49K+ Covid19 cases still no response from \n   cancel the compartment exa… https://t.co/kV2ZKmumu1' (https://t.co/kV2ZKmumu1')

In [39]:

```
# Cleaning Text: Multiple WhiteSpaces
df['clean_tweet'] = df['clean_tweet'].apply(nfx.remove_multiple_spaces)
```

In [40]:

```python
df['clean_tweet'].iloc[10]
```

Out[40]:

'49K+ Covid19 cases still no response from cancel the compartment exa…
https://t.co/kV2ZKmumu1' (https://t.co/kV2ZKmumu1')

In [41]:

```python
# Cleaning Text : Remove urls
df['clean_tweet'] = df['clean_tweet'].apply(nfx.remove_urls)
```

In [42]:

```python
# Cleaning Text: Punctuations
df['clean_tweet'] = df['clean_tweet'].apply(nfx.remove_puncts)
```

In [43]:

```python
df[['text','clean_tweet']]
```

Out[43]:

	text	clean_tweet
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and wouldnt it have made more sense to ha...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax We all cl...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciation f...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challenge I n...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020 The year of insanity Lol
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena Its a trib...
179106	More than 1,200 students test positive for #CO...	More than 1200 students test positive for at m...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop

179108 rows × 2 columns

SENTIMENT ANALYSIS

In [45]:

```
pip install textblob
```

```
Collecting textblob
  Downloading textblob-0.15.3-py2.py3-none-any.whl (636 kB)
     |████████████████████████████████| 636 kB 355 kB/s eta 0:00:01
Requirement already satisfied: nltk>=3.1 in /home/muskangupta/anaconda
3/lib/python3.8/site-packages (from textblob) (3.6.1)
Requirement already satisfied: click in /home/muskangupta/anaconda3/li
b/python3.8/site-packages (from nltk>=3.1->textblob) (7.1.2)
Requirement already satisfied: regex in /home/muskangupta/anaconda3/li
b/python3.8/site-packages (from nltk>=3.1->textblob) (2021.4.4)
Requirement already satisfied: joblib in /home/muskangupta/anaconda3/l
ib/python3.8/site-packages (from nltk>=3.1->textblob) (1.0.1)
Requirement already satisfied: tqdm in /home/muskangupta/anaconda3/li
b/python3.8/site-packages (from nltk>=3.1->textblob) (4.59.0)
Installing collected packages: textblob
Successfully installed textblob-0.15.3
Note: you may need to restart the kernel to use updated packages.
```

In [46]:

```python
from textblob import TextBlob
```

In [47]:

```python
def get_sentiment(text):
    blob = TextBlob(text)
    sentiment_polarity = blob.sentiment.polarity
    sentiment_subjectivity = blob.sentiment.subjectivity
    if sentiment_polarity > 0:
        sentiment_label = 'Positive'
    elif sentiment_polarity < 0:
        sentiment_label = 'Negative'
    else:
        sentiment_label = 'Neutral'
    result = {'polarity':sentiment_polarity,
              'subjectivity':sentiment_subjectivity,
              'sentiment':sentiment_label}
    return result
```

In [48]:

```python
# Text
ex1 = df['clean_tweet'].iloc[0]
```

In [51]:

```python
get_sentiment(ex1)
```

Out[51]:

```
{'polarity': -0.25, 'subjectivity': 0.25, 'sentiment': 'Negative'}
```

```python
df['sentiment_results'] = df['clean_tweet'].apply(get_sentiment)
```

```python
df['sentiment_results']
```

```
0         {'polarity': -0.25, 'subjectivity': 0.25, 'sen...
1         {'polarity': 0.5, 'subjectivity': 0.5, 'sentim...
2         {'polarity': 0.0, 'subjectivity': 0.0, 'sentim...
3         {'polarity': 0.0, 'subjectivity': 0.3571428571...
4         {'polarity': 0.0, 'subjectivity': 0.0, 'sentim...
                            ...
179103    {'polarity': 0.2, 'subjectivity': 0.2, 'sentim...
179104    {'polarity': 0.8, 'subjectivity': 0.7, 'sentim...
179105    {'polarity': 0.3, 'subjectivity': 1.0, 'sentim...
179106    {'polarity': 0.26325757575757575, 'subjectivit...
179107    {'polarity': 0.0, 'subjectivity': 0.0, 'sentim...
Name: sentiment_results, Length: 179108, dtype: object
```

```python
df['sentiment_results'].iloc[0]
```

```
{'polarity': -0.25, 'subjectivity': 0.25, 'sentiment': 'Negative'}
```

```python
pd.json_normalize(df['sentiment_results'].iloc[0])
```

	polarity	subjectivity	sentiment
0	-0.25	0.25	Negative

```
In [59]:

df = df.join(pd.json_normalize(df['sentiment_results']))
```

```
---------------------------------------------------------------------------
-----
ValueError                                Traceback (most recent call
 last)
<ipython-input-59-cfe3258be696> in <module>
----> 1 df = df.join(pd.json_normalize(df['sentiment_results']))

~/anaconda3/lib/python3.8/site-packages/pandas/core/frame.py in join(s
elf, other, on, how, lsuffix, rsuffix, sort)
   8108            5  K5  A5  NaN
   8109            """
-> 8110        return self._join_compat(
   8111            other, on=on, how=how, lsuffix=lsuffix, rsuffix=rs
uffix, sort=sort
   8112        )

~/anaconda3/lib/python3.8/site-packages/pandas/core/frame.py in _join_
compat(self, other, on, how, lsuffix, rsuffix, sort)
   8133                    sort=sort,
   8134                )
-> 8135        return merge(
   8136            self,
   8137            other,

~/anaconda3/lib/python3.8/site-packages/pandas/core/reshape/merge.py i
n merge(left, right, how, on, left_on, right_on, left_index, right_ind
ex, sort, suffixes, copy, indicator, validate)
     87            validate=validate,
     88        )
---> 89     return op.get_result()
     90
     91

~/anaconda3/lib/python3.8/site-packages/pandas/core/reshape/merge.py i
n get_result(self)
    684            join_index, left_indexer, right_indexer = self._get_jo
in_info()
    685
--> 686        llabels, rlabels = _items_overlap_with_suffix(
    687            self.left._info_axis, self.right._info_axis, self.
suffixes
    688        )

~/anaconda3/lib/python3.8/site-packages/pandas/core/reshape/merge.py i
n _items_overlap_with_suffix(left, right, suffixes)
   2176
   2177     if not lsuffix and not rsuffix:
-> 2178        raise ValueError(f"columns overlap but no suffix speci
fied: {to_rename}")
   2179
   2180     def renamer(x, suffix):

ValueError: columns overlap but no suffix specified: Index(['polarit
y', 'subjectivity', 'sentiment'], dtype='object')
```

```
df.head()
```

Out[60]:

	user_name	user_location	user_description	user_created	user_followers	user_friends	us
0	ꙴꙴꙴꙴꙴ	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	
1	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	
2	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[]_[] #Cavs ...	2019-03-07 01:45:06	197	987	
4	DIPR-J&K	Jammu and Kashmir	🖊Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	

In [61]:

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

Out[61]:

```
Neutral     75490
Positive    74154
Negative    29464
Name: sentiment, dtype: int64
```

```
df['sentiment'].value_counts().plot(kind='bar')
```

<AxesSubplot:>
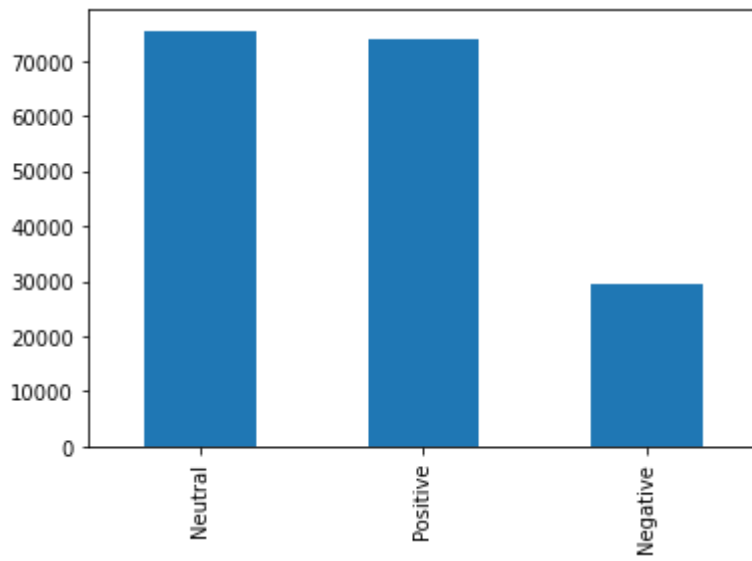
In [63]:

```python
# Plot with seaborn
sns.countplot(df['sentiment'])
```

Out[63]:

```
<AxesSubplot:xlabel='sentiment', ylabel='count'>
```
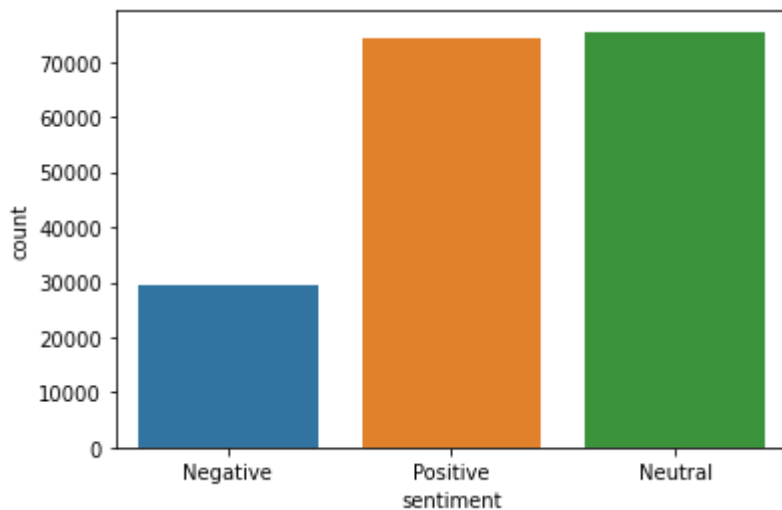

In [70]:

```
### Keyword Extraction
+ For Positive and Negative Sentiment
+  General
```

```
  File "<ipython-input-70-fb86073ddf45>", line 2
    + For Positive and Negative Sentiment
      ^
SyntaxError: invalid syntax
```

In [71]:

```python
positive_tweet = df[df['sentiment'] == 'Positive']['clean_tweet']
```

In [72]:

```python
neutral_tweet = df[df['sentiment'] == 'Neutral']['clean_tweet']
negative_tweet = df[df['sentiment'] == 'Negative']['clean_tweet']
```

In [73]:

```
positive_tweet
```

Out[73]:

```
1          Hey and  wouldnt it have made more sense to ha...
6          How Will Change Work in General (and recruitin...
8                     Praying for good health and recovery of
12         👏 — no one will be safe from until everyone i...
13         Lets all protect ourselves from Its real and t...
                                ...
179096     We were really bummed we couldn't cop one of t...
179103     Thanks for nominating me for the challenge I n...
179104                     2020 The year of insanity Lol
179105      A powerful painting by Juan Lucena Its a trib...
179106     More than 1200 students test positive for at m...
Name: clean_tweet, Length: 74154, dtype: object
```

In [74]:

```
# Remove Stopwords and Convert to Tokens
positive_tweet_list = positive_tweet.apply(nfx.remove_stopwords).tolist()
```

In [75]:

```
negative_tweet_list = negative_tweet.apply(nfx.remove_stopwords).tolist()
neutral_tweet_list = neutral_tweet.apply(nfx.remove_stopwords).tolist()
```

In [76]:

```
positive_tweet_list
```

Out[76]:

```
['Hey wouldnt sense players pay respects A…',
 'Change Work General (and recruiting specifically) via/',
 'Praying good health recovery',
 '👏 — safe safe commit ensure…',
 'Lets protect real numbers climbing fast Continent Lets n…',
 'Second wave Flandersback',
 'COVID Update: infection rate Florida following natural curve exper
ts predicted initial cu…',
 'Good Patriots Volunteer Election Judge Polls open without…',
 'comprehensive review amp Analysis: key ways WASH help reduce trans
mission…',
 'crosses 2 lakh mark 150055 ppp recovered far positive today 6988/t
otal…',
 'Actor father MrGKReddy tested positive 15/20 days s…',
 'safe place visit guests said hotel meticulous applying hand sanita
tion als…',
 'CEO live WSU talking "The effects Covid19 students"…',
 'Actionables healthy recovery',
```

TOKENIZATION

```python
for line in positive_tweet_list:
#     print(line)
    for token in line.split():
        print(token)
```

```
Hey
wouldnt
sense
players
pay
respects
A…
Change
Work
General
(and
recruiting
specifically)
via/
Praying
good
health
recovery
👏
```

```python
pos_tokens = [token for line in positive_tweet_list  for token in line.split()]
```

```python
neg_tokens = [token for line in negative_tweet_list  for token in line.split()]
neut_tokens = [token for line in neutral_tweet_list  for token in line.split()]
```

In [83]:

```
pos_tokens
```

Out[83]:

```
['Hey',
 'wouldnt',
 'sense',
 'players',
 'pay',
 'respects',
 'A…',
 'Change',
 'Work',
 'General',
 '(and',
 'recruiting',
 'specifically)',
 'via/',
 'Praying',
 'good',
 'health',
 'recovery',
```

In [84]:

```
# Get Most Commonest Keywords
from collections import Counter
```

In [85]:

```
def get_tokens(docx,num=30):
    word_tokens = Counter(docx)
    most_common = word_tokens.most_common(num)
    result = dict(most_common)
    return result
```

```
get_tokens(pos_tokens)
```

```
{'cases': 10310,
 'new': 9947,
 'amp': 5298,
 'positive': 3888,
 'people': 3655,
 'New': 3325,
 'deaths': 3292,
 'COVID19': 2745,
 'pandemic': 1929,
 'total': 1783,
 'reported': 1773,
 'safe': 1677,
 'today': 1661,
 'confirmed': 1522,
 'latest': 1468,
 'like': 1455,
 'time': 1455,
 '|': 1454,
 'good': 1449,
 'tested': 1398,
 'need': 1393,
 '24': 1380,
 'coronavirus': 1362,
 'day': 1331,
 'number': 1270,
 'right': 1266,
 'reports': 1264,
 'India': 1230,
 'Total': 1227,
 '2020': 1222}
```

```
most_common_pos_words = get_tokens(pos_tokens)
most_common_neg_words = get_tokens(neg_tokens)
most_common_neut_words = get_tokens(neut_tokens)
```

```
# Plot with seaborn
neg_df = pd.DataFrame(most_common_neg_words.items(),columns=['words','scores'])
```

```
neg_df
```

	words	scores
0	cases	2245
1	amp	2033
2	people	1801
3	pandemic	895
4	spread	870
5	daily	760
6	like	754
7	COVID19	748
8	slow	734
9	new	723
10	risk	713
11	deaths	679
12	symptoms	621
13	day	604
14	Help	573
15	time	568
16	long	550
17	y…	547
18	health	534
19	identify	533
20	sooner	529
21	know	527
22	selfreporting	522
23	virus	516
24	dead	510
25	hard	489
26	mask	478
27	going	475
28	bad	474
29	Trump	459

```python
plt.figure(figsize=(20,10))
sns.barplot(x='words',y='scores',data=neg_df)
plt.xticks(rotation=45)
plt.show()
```
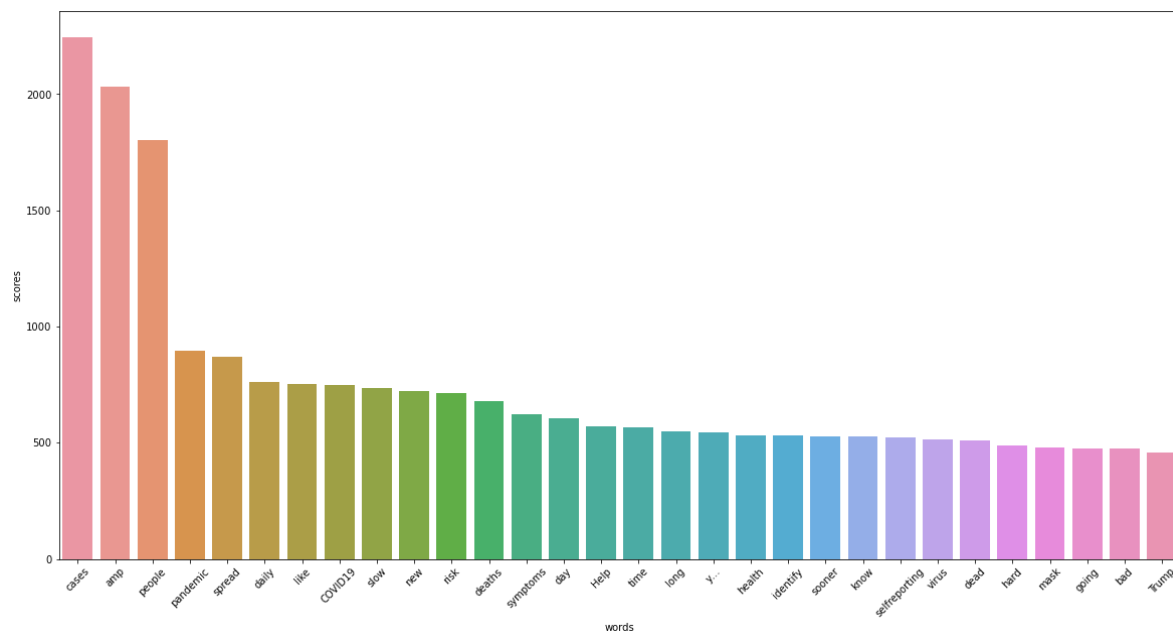
```python
# Plot with seaborn
pos_df = pd.DataFrame(most_common_pos_words.items(),columns=['words','scores'])
plt.figure(figsize=(20,10))
sns.barplot(x='words',y='scores',data=pos_df)
plt.xticks(rotation=45)
plt.show()
```
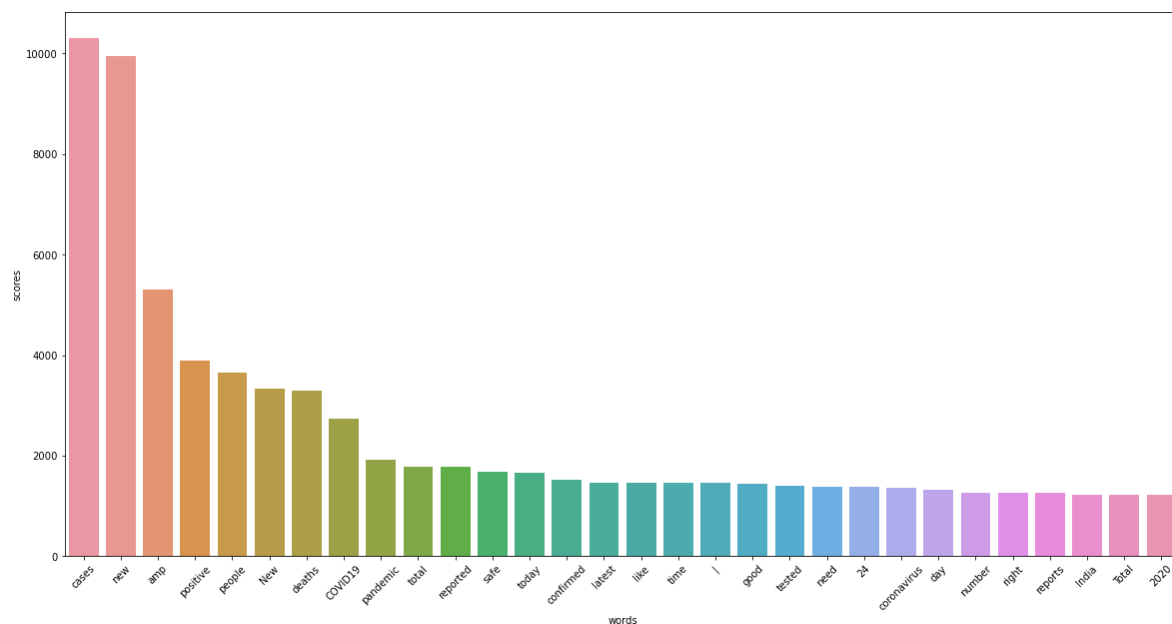
```python
# Plot with seaborn
neut_df = pd.DataFrame(most_common_neut_words.items(),columns=['words','scores'])
plt.figure(figsize=(20,10))
sns.barplot(x='words',y='scores',data=neut_df)
plt.xticks(rotation=45)
plt.show()
```
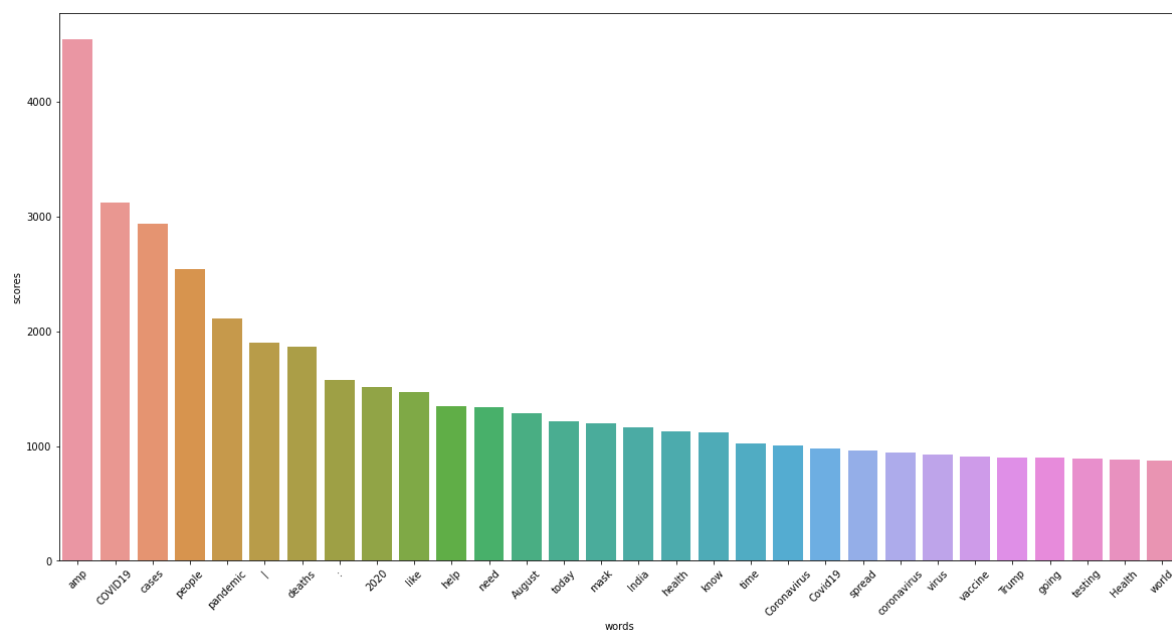
In [99]:

```
pip install wordcloud
```

WARNING: Keyring is skipped due to an exception: Failed to create the
 collection: Remote peer disconnected.
Collecting wordcloud
  Downloading wordcloud-1.8.1-cp38-cp38-manylinux1_x86_64.whl (371 kB)
     |████████████████████████████████| 371 kB 886 kB/s eta 0:00:01
Requirement already satisfied: matplotlib in /home/muskangupta/anacond
a3/lib/python3.8/site-packages (from wordcloud) (3.3.4)
Requirement already satisfied: pillow in /home/muskangupta/anaconda3/l
ib/python3.8/site-packages (from wordcloud) (8.2.0)
Requirement already satisfied: numpy>=1.6.1 in /home/muskangupta/anaco
nda3/lib/python3.8/site-packages (from wordcloud) (1.20.1)
Requirement already satisfied: python-dateutil>=2.1 in /home/muskangup
ta/anaconda3/lib/python3.8/site-packages (from matplotlib->wordcloud)
 (2.8.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.
3 in /home/muskangupta/anaconda3/lib/python3.8/site-packages (from mat
plotlib->wordcloud) (2.4.7)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/muskangupta/
anaconda3/lib/python3.8/site-packages (from matplotlib->wordcloud) (1.
3.1)
Requirement already satisfied: cycler>=0.10 in /home/muskangupta/anaco
nda3/lib/python3.8/site-packages (from matplotlib->wordcloud) (0.10.0)
Requirement already satisfied: six in /home/muskangupta/anaconda3/lib/
python3.8/site-packages (from cycler>=0.10->matplotlib->wordcloud) (1.
15.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.8.1
Note: you may need to restart the kernel to use updated packages.

In [100]:

```
### Word Cloud
from wordcloud import WordCloud
```
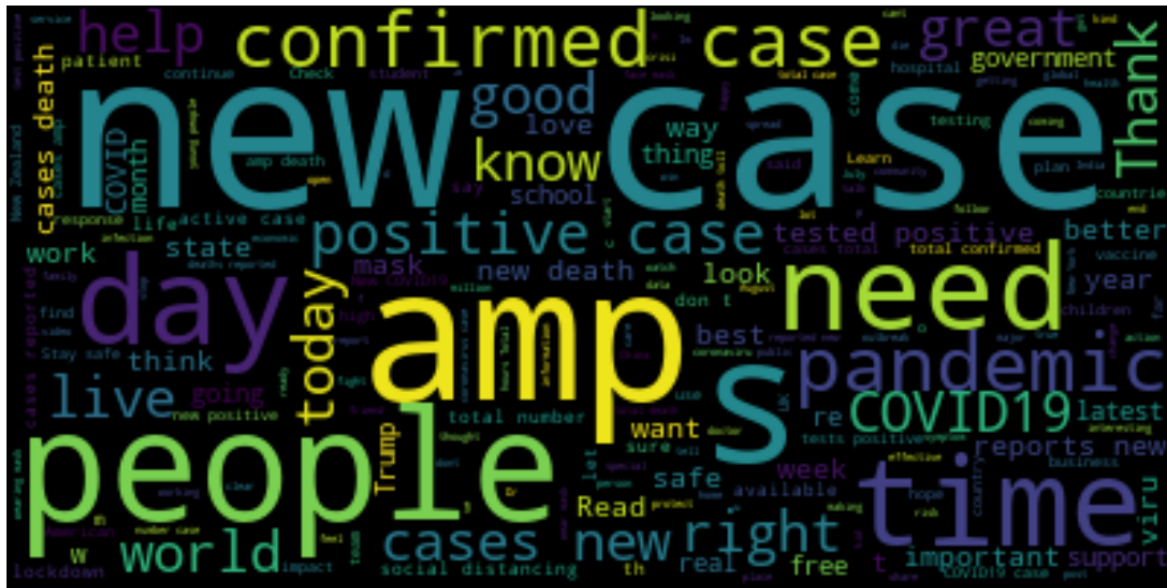
In [105]:

```
def plot_wordcloud(docx):
    plt.figure(figsize=(20,10))
    mywordcloud = WordCloud().generate(docx)
    plt.imshow(mywordcloud,interpolation='bilinear')
    plt.axis('off')
    plt.show()
```
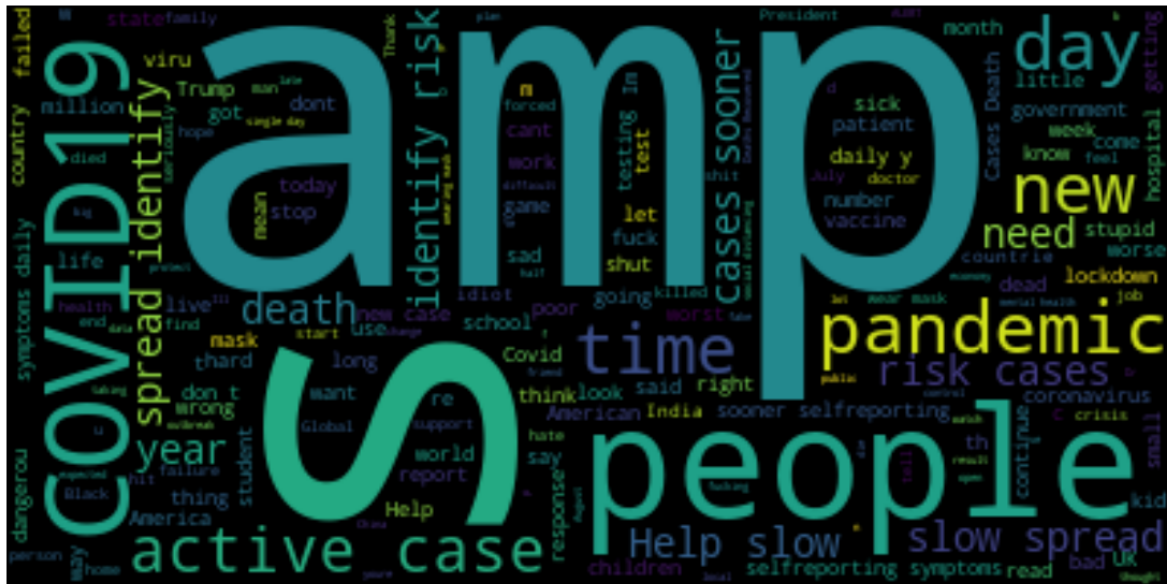
In [106]:

```
pos_docx = ' '.join(pos_tokens)
neg_docx = ' '.join(neg_tokens)
neu_docx = ' '.join(neut_tokens)
```
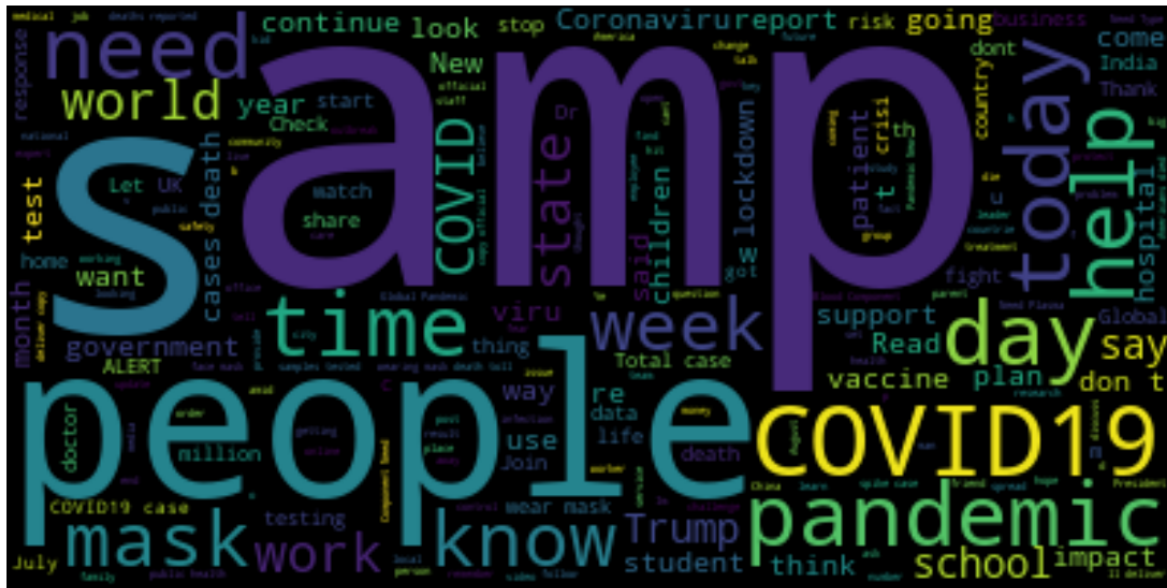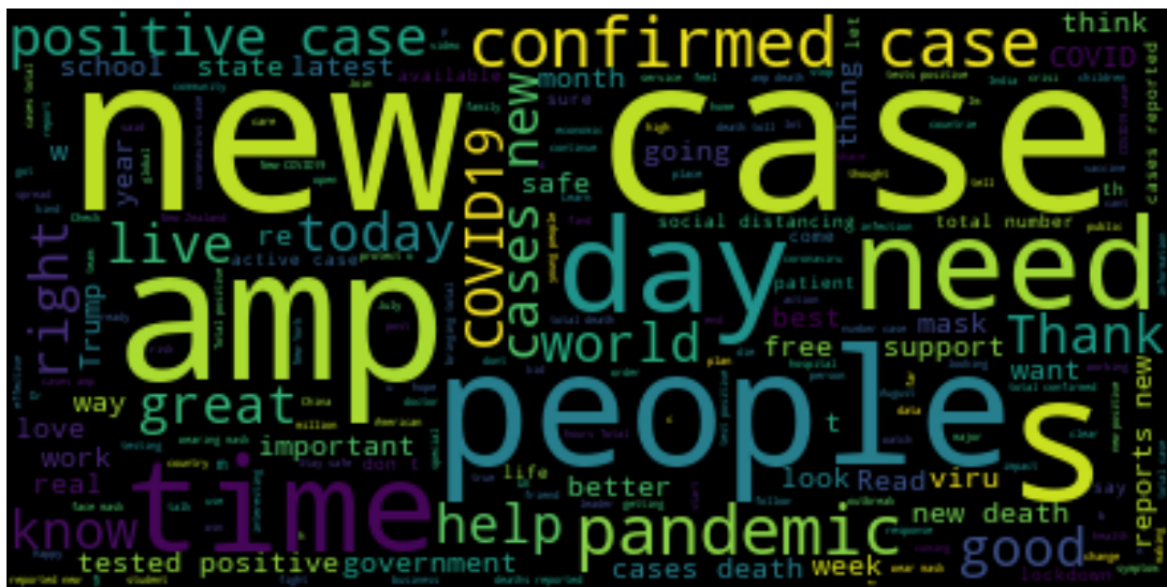
```
plot_wordcloud(pos_docx)
```

```
plot_wordcloud(neg_docx)
```

```
plot_wordcloud(neu_docx)
```

```
plot_wordcloud(pos_docx)
```


In [ ]:

```
### Thanks for your time, MUSKAN GUPTA, M.sc DS
```