

- S** Analizar la relevancia del uso de redes sociales durante la  
pandemia
- M** Obteniendo analisis por número de tweets, hashtags, número  
de fuentes, y localidades con mayor frecuencia
- A** Realizando gráficas de resumen y filtración de datos mediante  
la implementación de herramientas computacionales
- R** Para entender la relación del uso de los tweets  
(hashtags) y la globalidad durante la pandemia
- T** Del 12 de mayo al 13 de mayo del 2022

- Procesar el data set covid19\_tweets2.csv mediante las técnicas de analítica de datos que discutimos durante la semana.

In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
data = pd.read_csv("covid19_tweets2.csv")
data.head()
```

Out[1]:

		user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites
0	Vi@L€†	wednesday	astroworld	addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	187
1	Tom Basile us	Husband, Father, Columnist & Commentator. Auth...	New York, NY		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	72
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...		2019-03-07 01:45:06	197	987	14
4	DIPR-J&K	Jammu and Kashmir	Official Twitter handle of Department of Inf...		2017-02-12 06:45:15	101009	168	1



In [2]:

```
data.describe()
```

Out[2]:

	user_followers	user_friends	user_favourites
<b>count</b>	1.791080e+05	179108.000000	1.791080e+05
<b>mean</b>	1.090555e+05	2121.701566	1.444411e+04
<b>std</b>	8.414670e+05	9162.553072	4.452270e+04
<b>min</b>	0.000000e+00	0.000000	0.000000e+00
<b>25%</b>	1.720000e+02	148.000000	2.060000e+02
<b>50%</b>	9.920000e+02	542.000000	1.791000e+03
<b>75%</b>	5.284000e+03	1725.250000	9.388000e+03
<b>max</b>	4.944256e+07	497363.000000	2.047197e+06

In [3]: `data.count()`

```
Out[3]: user_name      179108
user_location    142337
user_description 168822
user_created     179108
user_followers   179108
user_friends     179108
user_favourites  179108
user_verified    179108
date             179108
text              179108
hashtags         127774
source            179031
is_retweet        179108
dtype: int64
```

In [4]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 179108 entries, 0 to 179107
Data columns (total 13 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   user_name        179108 non-null   object 
 1   user_location     142337 non-null   object 
 2   user_description  168822 non-null   object 
 3   user_created      179108 non-null   object 
 4   user_followers    179108 non-null   int64  
 5   user_friends      179108 non-null   int64  
 6   user_favourites   179108 non-null   int64  
 7   user_verified     179108 non-null   bool   
 8   date              179108 non-null   object 
 9   text               179108 non-null   object 
 10  hashtags          127774 non-null   object 
 11  source             179031 non-null   object 
 12  is_retweet        179108 non-null   bool  
dtypes: bool(2), int64(3), object(8)
memory usage: 15.4+ MB
```

In [5]: `data.drop(['is_retweet','source','user_created'],inplace = True, axis = 1)`  
data

	user_name	user_location	user_description	user_followers	user_friends	user_favourites	user...
0	Viñi Léon	astroworld	wednesday addams as a disney princess keepin i...	624	950	18775	
1	Tom Basile us	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2253	1677	24	
2	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	9275	9525	7254	

		user_name	user_location	user_description	user_followers	user_friends	user_favourites	user_statuses
3	ethel mertz	Stuck in the Middle		#Browns #Indians #ClevelandProud #[...] #Cavs ...	197	987	1488	
4	DIPR-J&K	Jammu and Kashmir		📝 Official Twitter handle of Department of Inf...	101009	168	101	
...	...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.	Ilorin, Nigeria		Animal Scientist   Muslim   Real Madrid/Chelsea	412	1609	1062	
179104	Jason	Ontario		When your cat has more baking soda than Ninja	150	182	7295	
179105	BEEHEMOTH	CA Canada		🛠️ The Architects of Free Trade 🛠️ Really Did ...	1623	2160	98000	
179106	Gary DelPonte	New York City		Global UX UI Visual Designer. StoryTeller, Mus...	1338	1111	0	
179107	TUKY II	Aliwal North, South Africa		TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	97	1697	566	

179108 rows × 10 columns



In [6]:

```
import sys
print(sys.executable)
```

c:\Program Files\Python310\python.exe

In [7]:

```
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer as SIA
from nltk.corpus import subjectivity
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import re
from textblob import TextBlob

%matplotlib inline
```

[nltk\_data] Downloading package vader\_lexicon to  
[nltk\_data] C:\Users\Guillermo\AppData\Roaming\nltk\_data...

```
[nltk_data]  Package vader_lexicon is already up-to-date!
```

In [8]:

```
def text_cleaner(text) :
    text = re.sub(r"\n", " ", text)
    text = re.sub(r"\t", " ", text)
    text = re.sub(r"\r", " ", text)
    text = re.sub(r"(@)|(#)|(RT[\s]+)|(https?:\/\/\S+)|([^\w-])", "", text)
    text = text.strip(" ")
    return text
```

In [9]:

```
def Covid_or_Coronavirus_remover(text) :
    text = re.sub(r"((Covid)|(COVID)|(covid)|(Corona)|(corona)|(CORONA))+", "", text)
    text = re.sub(r"((Covid)|(COVID)|(covid)|(Corona)|(corona)|(CORONA))[A-Za-z0-9]+",
    text = text.strip(" ")
    return text
```

In [10]:

```
def hashtag_collector(text):
    list_of_hashtags = re.findall(r"\#[A-Za-z0-9\-\.\_]+", text, re.DOTALL)
    if(list_of_hashtags != None):
        list_of_hashtags = [word.replace('#', '') for word in list_of_hashtags]
        list_of_hashtags = [word.replace('.', '') for word in list_of_hashtags]
        list_of_hashtags = [word.upper() for word in list_of_hashtags]
    return list_of_hashtags
```

In [11]:

```
def at_collector(text) :
    list_of_ats = re.findall(r"@[A-Za-z0-9\-\.\_]+", text, re.DOTALL)
    if(list_of_ats != None):
        list_of_ats = [name.replace('@', '') for name in list_of_ats]
    return list_of_ats
```

In [12]:

```
def compound_sentiment_score(tweet):
    VaderSent = SIA()
    Overall_sentiment = VaderSent.polarity_scores(tweet)
    return Overall_sentiment
```

In [13]:

```
def getSubjectivity(tweet):
    return TextBlob.subjectivity(tweet)
```

In [14]:

```
def getPolarity (tweet):
    return TextBlob.polarity(tweet)
```

In [15]:

```
def doAnalysis(score) :
    return 'Neutral' if (score == 0) else ('Negative' if (score < 0) else 'Positive')
```

In [16]:

```
def wordClouder(string_column):
    Words = ''.join([words for words in string_column])
    wordcloud = WordCloud(width = 1000, height = 600, random_state = 10, max_font_size
    wordcloud.generate(Words)
    plt.style.use('fivethirtyeight')
```

```
plt.figure(figsize = (20,28))
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```

In [17]:

```
data["hashtags_in_tweets"] = data["text"].apply(hashtag_collector)
data
```

Out[17]:

		user_name	user_location	user_description	user_followers	user_friends	user_favourites	use...
0		⚡i⊕l€†	astroworld	wednesday addams as a disney princess keepin i...	624	950	18775	
1	Tom Basile	us	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2253	1677	24	
2	Time4fisticuffs		Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	9275	9525	7254	
3	ethel mertz		Stuck in the Middle	#Browns #Indians #ClevelandProud #[]_[] #Cavs ...	197	987	1488	
4	DIPR-J&K		Jammu and Kashmir	✍ Official Twitter handle of Department of Inf...	101009	168	101	
...	...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.		Ilorin, Nigeria	Animal Scientist   Muslim   Real Madrid/Chelsea	412	1609	1062	
179104	Jason		Ontario	When your cat has more baking soda than Ninja ...	150	182	7295	
179105	BEEHEMOTH		ca Canada	🛠 The Architects of Free Trade 🛠 Really Did ...	1623	2160	98000	
179106	Gary DelPonte		New York City	Global UX UI Visual Designer. StoryTeller, Mus...	1338	1111	0	
179107	TUKY II		Aliwal North, South Africa	TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	97	1697	566	

179108 rows × 11 columns

In [18]:

```
data["ats_in_tweets"] = data["text"].apply(at_collector)
data
```

Out[18]:

		user_name	user_location	user_description	user_followers	user_friends	user_favourites	user_statuses	user_statuses_favorited	user_statuses_retweeted	user_statuses_replies
0		“Viñet”	astroworld	wednesday addams as a disney princess keepin i...	624	950	18775				
1	Tom Basile	us	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2253	1677	24				
2	Time4fisticuffs		Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	9275	9525	7254				
3	ethel mertz		Stuck in the Middle	#Browns #Indians #ClevelandProud #[ ] #Cavs ...	197	987	1488				
4	DIPR-J&K		Jammu and Kashmir	✍ Official Twitter handle of Department of Inf...	101009	168	101				
...	...	...	...	...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.		Ilorin, Nigeria	Animal Scientist   Muslim   Real Madrid/Chelsea	412	1609	1062				
179104	Jason		Ontario	When your cat has more baking soda than Ninja	150	182	7295				
179105	BEEHEMOTH		CA Canada	🛠 The Architects of Free Trade 🛠 Really Did ...	1623	2160	98000				
179106	Gary DelPonte		New York City	Global UX UI Visual Designer. StoryTeller, Mus...	1338	1111	0				
179107	TUKY II		Aliwal North, South Africa	TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	97	1697	566				

179108 rows × 12 columns

In [19]:

```
data["text"] = data["text"].apply(text_cleaner)
data
```

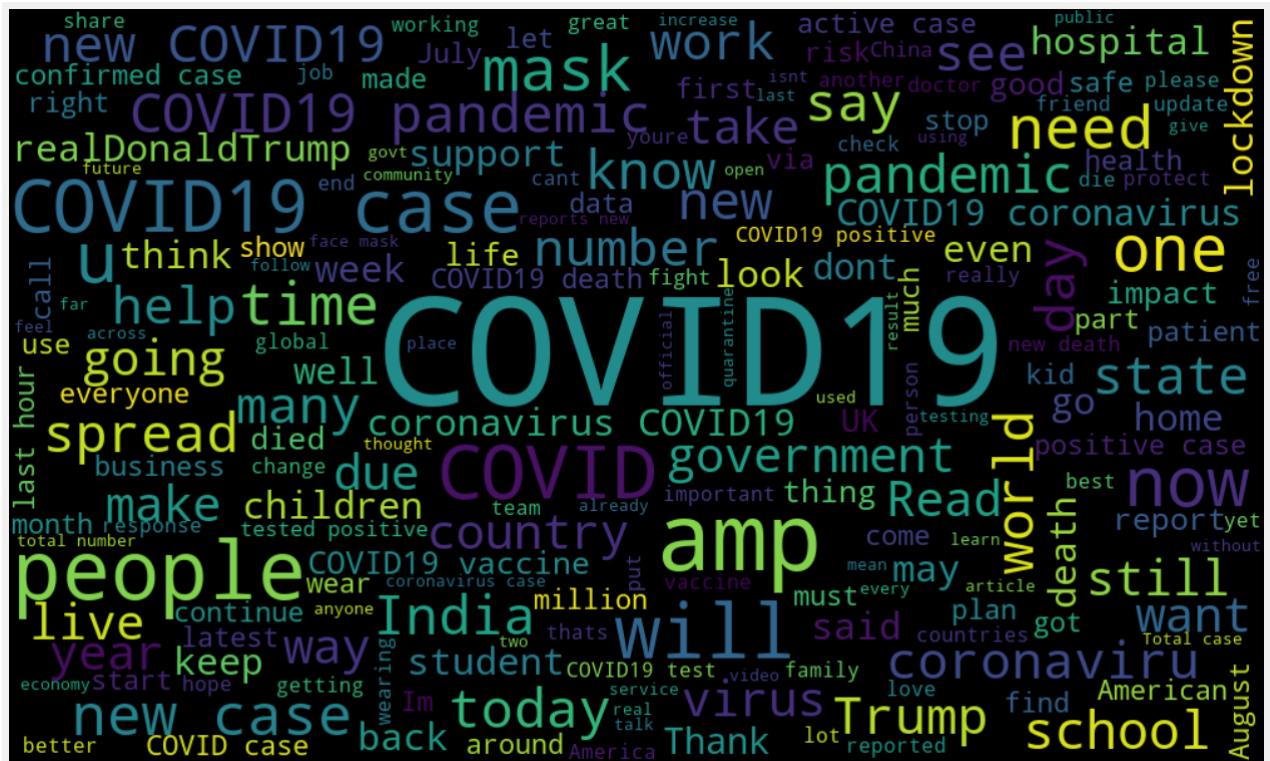
Out[19]:

		user_name	user_location	user_description	user_followers	user_friends	user_favourites	user_statuses	user_verified	user_protected	user_pinned_tweet_id
0		Vi L€†	astroworld	wednesday addams as a disney princess keepin i...	624	950	18775	...	...	...	...
1	Tom Basile	us	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2253	1677	24	...	...	...	...
2	Time4fisticuffs		Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	9275	9525	7254	...	...	...	...
3	ethel mertz		Stuck in the Middle	#Browns #Indians #ClevelandProud #[ ] #Cavs ...	197	987	1488	...	...	...	...
4	DIPR-J&K		Jammu and Kashmir	Official Twitter handle of Department of Inf...	101009	168	101	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.		Ilorin, Nigeria	Animal Scientist   Muslim   Real Madrid/Chelsea	412	1609	1062	...	...	...	...
179104	Jason		Ontario	When your cat has more baking soda than Ninja	150	182	7295	...	...	...	...
179105	BEEHEMOTH		ca Canada	The Architects of Free Trade The Really Did ...	1623	2160	98000	...	...	...	...
179106	Gary DelPonte		New York City	Global UX UI Visual Designer. StoryTeller, Mus...	1338	1111	0	...	...	...	...
179107	TUKY II		Aliwal North, South Africa	TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	97	1697	566	...	...	...	...

179108 rows × 12 columns



```
In [20]: wordCloud(data['text'])
```

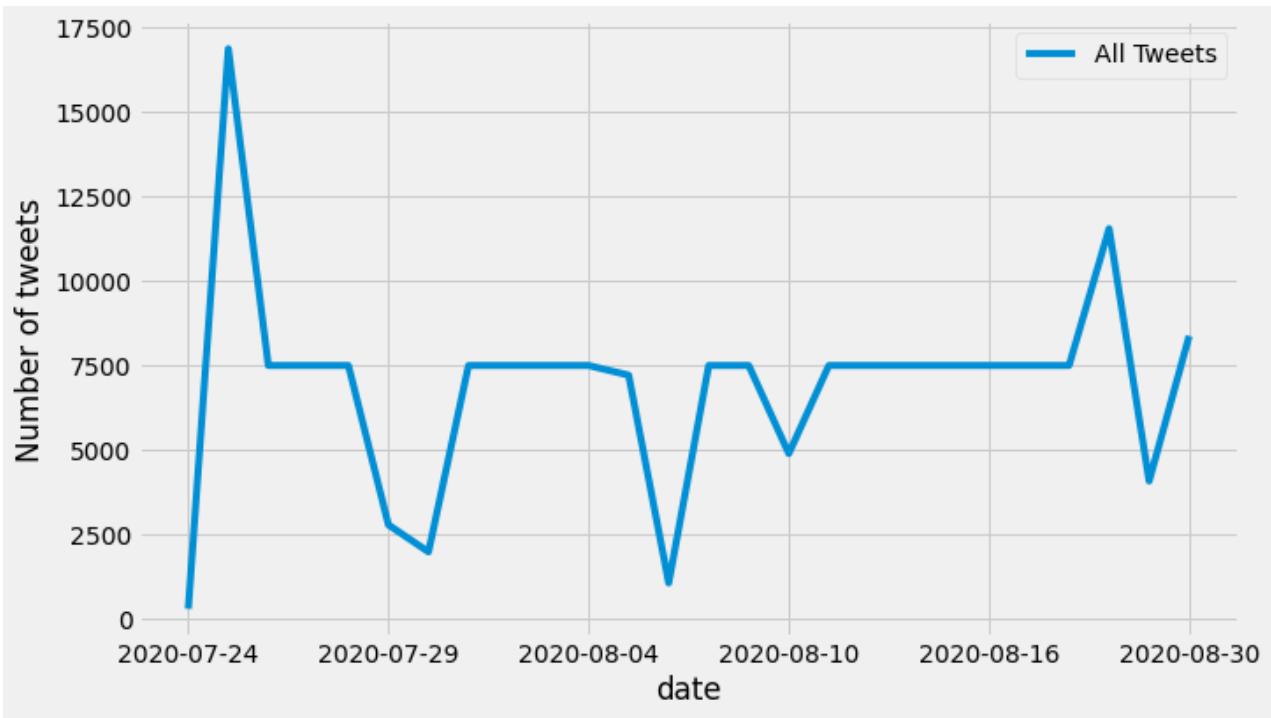


```
In [21]: data["hashtags"].unique()
```

```
Out[21]: array([['COVID19'], ['CoronaVirusUpdates', 'COVID19'], ...,  
    ['COVID19', 'coronavirus', 'TrumpVirus'],  
    ['COVID19', 'CloseTheSchools', 'KeepTheSchoolsClosed', 'KeepTeachersAlive', 'K1  
2'],  
    ['nurses', 'COVID19', 'coronavirus', 'schools']], dtype=object)
```

```
In [22]: data["date"] = data["date"].astype(str).str[0:10]
```

```
In [23]: y = data[["date","text"]].groupby("date").count()
y.rename({"text":"All Tweets"},inplace = True, axis= 1)
ax = y.plot(figsize = (10,6), ylabel = "Number of tweets")
```



In [24]:

```
from nltk import FreqDist
our_hashtags_in_tweets = data["hashtags_in_tweets"]
our_hashtags_in_tweets = sum(our_hashtags_in_tweets, [])
our_hashtags_in_tweets
```

Out[24]:

```
['COVID19',
 'COVID19',
 'CORONAVIRUSUPDATES',
 'COVID19',
 'CORONAVIRUS',
 'COVID19',
 'COVID19',
 'RECRUITING',
 'COVID19',
 'COVIDPOSITIVE',
 'HURRICANEHANNA',
 'COVID19',
 'LOGO',
 'GRAPHICDESIGNER',
 'LOGODESIGN',
 'LOGODESINGER',
 'ICON',
 'MINIMALIST',
 'ABSTRACT',
 'COVID19',
 'COVID19',
 'COVID19',
 'COVID19',
 'KEEPSOCIALDISTANCE',
 'COVID19',
 'TAMILNADU',
 'CHENNAI',
 'COVID19',
 'HOMEWORK',
 'LIGHT',
 'WASH',
 'COVID19',
```

'COVID19',  
'KOLAR',  
'BLOOD',  
'COVID19',  
'BOSNIAANDHERZEGOVINA',  
'COVID19',  
'TNCORONAUPDATE',  
'TN',  
'COVID19',  
'FEMA',  
'PUERTORICO',  
'COVID19',  
'COVID19',  
'VISHAL',  
'COVID19',  
'AFRICA',  
'COVID19',  
'COVID19',  
'COVID19',  
'CLIMATE',  
'BORISJOHNSON',  
'SCOBRA',  
'FIVE',  
'COVID19',  
'PLAYYOURPART',  
'COVID19',  
'COVID19',  
'COVID19',  
'DELHI',  
'COVID19',  
'TAMILNADU',  
'COVID19',  
'RUTOSWINDLINGGEMA',  
'COVID19',  
'TAMILNADU',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'INDIAFIGHTSCORONA',  
'COVID19',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'OBESITY',  
'CARRYMINATI',  
'MY200VSBUTCH200',  
'COVID19',  
'FERGUSONS',  
'DURBANJULY',  
'ALREADYVIDEO',  
'COVID19',  
'WISEUP',  
'ETSY',  
'EPICONETSY',  
'MASKSFORSALE',  
'COVID19',  
'COVID19',  
'MUMBAIPINTERS',

'INDUSTRY',  
'SATURDAYVIBES',  
'COVID19',  
'COVID19',  
'COVID19',  
'POLITICAL',  
'COVID19',  
'CLEAN',  
'SALESCOACH',  
'BUSINESSCOACH',  
'SALESFORCE',  
'COVID19',  
'NSTWORLD',  
'NATIONALDAYRALLY',  
'NDR',  
'AMITSHAH',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'DURBANJULY',  
'COVID19',  
'ALREADYVIDEO',  
'COVID19',  
'HURRICANEHANNA',  
'TAMILNADU',  
'COVID19',  
'PLAYYOURPART',  
'COVID19',  
'INTHISTOGETHER',  
'COVID19',  
'BELTANDROAD',  
'COVID19',  
'VISHAL',  
'HANNA',  
'COVID19',  
'COVID19',  
'TRUMPVIRUS',  
'MASKITORCASKET',  
'FOXNEWS',  
'COVID19',  
'WALTDISNEYWORLD',  
'COVID19',  
'TRAVEL',  
'CORONAVIRUS',  
'COVID19',  
'HISPANIC',  
'LATINX',  
'COVID19',  
'NIGERIA',  
'DRKAWANA',  
'CORONAVIRUS',  
'COVID19',  
'STAYATHOME',  
'COVID19',  
'BENPC',  
'LOCKDOWNPHOTOCHALLENGE',  
'COVID19',  
'ONLYFANSBABE',  
'IHAVEAJOKE',

'ONLYFANSGIRL',  
'ALREADYVIDEO',  
'COVID19',  
'COVID19',  
'AMSTERDAM',  
'COVID19',  
'TRAVEL',  
'COVID19',  
'COVID19',  
'COVID',  
'COVIDIOTS',  
'CDC',  
'COVID19',  
'COVID19',  
'LEICESTERSHIRE',  
'COVID19',  
'MY200VSBUTCH200',  
'ROYALEHIGHTRADES',  
'ALREADYVIDEO',  
'MTVHOTTTEST',  
'COVID19',  
'VISHAL',  
'COVID19',  
'FACTS',  
'SATURDAYMOTIVATION',  
'TRUTH',  
'COVID19',  
'UTTARPRADESH',  
'COVID-19',  
'VOTEBLUE2020',  
'VOTEFORBIDEN',  
'VOTEBLUETOENDTHISNIGHTMARE',  
'VIETNAM',  
'COVID19',  
'CORONAVIRUS',  
'NCOV',  
'COVID19',  
'FATIGUE',  
'TWOWAYRADIO',  
'COVID19',  
'1',  
'MASKUP',  
'MASK',  
'MASKS',  
'MASKMOANERS',  
'COVID19',  
'COVID',  
'REPUBLICANSFORBIDENNOW',  
'COVID19',  
'COVID19',  
'COVID19',  
'THEBIGSPEND',  
'COVID19',  
'TRAVELTUESDAY',  
'COVID19',  
'COVID19',  
'ANDHRAPRADESH',  
'APFIGHTSCORONA',  
'STAYHOME',  
'COVID19',

'AR',  
'BBNAJIA',  
'COVID19',  
'TACHADANCECHALLANGE',  
'SAYNOTONIPOSTFEE',  
'KERALA',  
'COVID',  
'CORONAVIRUS',  
'CORONAVIRUSUPDATE',  
'COVID19',  
'CORONAVIRUSPANDEMIC',  
'BIGNEWS',  
'RAJASTHAN',  
'COVID19',  
'WEARAMASK',  
'COVID',  
'COVID19',  
'UAE',  
'COVID19',  
'COVID19',  
'QATAR',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'VIRTUAL',  
'CHRO',  
'PANELDISCUSSION',  
'WORKPLACES',  
'COVID19',  
'COVID19',  
'ENGLAND',  
'AIRLINES',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'COVID19-',  
'FAKENEWS',  
'DEMOCRATS',  
'AIRLINES',  
'COVID19',  
'COVID19',  
'MAHARASHTRA',  
'COVID19',  
'JAIPUR',  
'PLASMABANK',  
'COVID19',  
'WIN',  
'MENTALHEALTH',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'FOXNEWS',  
'COVID19',  
'COVID19',  
'TECHSTRATEGY',  
'DANANG',

'COVID19',  
'ECMO',  
'SOUTHAFRICA',  
'TOURISM',  
'CARTEBLANCHE',  
'COVID19',  
'CRICKET',  
'RUGBY',  
'COLLEAGUESFORLIFE',  
'DELOITTEALUMNI',  
'OPENINGDAY',  
'GRANDSLAM',  
'WALKOFF',  
'BRAZIL',  
'BILSONARO',  
'COVID19',  
'COVID19',  
'COVID19',  
'DELHI',  
'COVID19',  
'FACECOVERINGS',  
'MASKPROBLEMS',  
'FACEMASK',  
'HACK',  
'MASKHACK',  
'MASKMOANERS',  
'COVID19',  
'COVID19',  
'TIMES',  
'GOOD',  
'DEPRESSION',  
'COVID19',  
'VISHAL',  
'BENPC',  
'LOCKDOWNPHOTOCHALLENGE',  
'COVID19',  
'SATURDAYMORNING',  
'FLOWERS',  
'COVID19',  
'COVID19',  
'ANDHRAPRADESH',  
'CORONAVIRUS',  
'COVID19',  
'PROTESTS',  
'COVID19',  
'COVID19',  
'FIGHTAGAINSTCORONA',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'MANUFACTURING',  
'COVIDUPDATES',  
'APEC',  
'NEWS',  
'TODAY',  
'POLITICS',  
'TRUMP',  
'MONEY',  
'WISDOM',

'LEADERSHIP',  
'APEC',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'TESTINGWHILEBLACK',  
'COVID19',  
'COVID19',  
'DONATE',  
'MASKS',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'SATURDAYMORNING',  
'SATURDAYVIBES',  
'CORONAVIRUSSPAIN',  
'COVID19',  
'SPAIN',  
'CATALONIA',  
'BARCELONA',  
'GOODADVICE',  
'COVID19',  
'DELHI',  
'COVID19',  
'SOCIALDISTANCING',  
'BBQ',  
'COVID19',  
'COVID19',  
'WAKEUP',  
'COVID19',  
'FAKE',  
'PLANDEMIC',  
'EXOSOMES',  
'VIRUSTHEORY',  
'5G',  
'VIRUS',  
'SAYNOTOBILLGATES',  
'COVID19',  
'INTHISTOGETHER',  
'MEDBIKINI',  
'COVID19',  
'HOMESCHOOLING',  
'COVID19',  
'CORONAVIRUS',  
'NC',  
'PANDEMIC',  
'COVID',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'ALREADYVIDEO',  
'RIHANNAXTIWA',  
'OBASEKIFLAGSOFTODAY',  
'COVID19',  
'TACHADANCECHALLANGE',  
'FRAUDBYQUODJOE',  
'BBNAIJIA2020',  
'COVID19',

'COVID19',  
'COVID19',  
'STIMULUS',  
'FACECOVERINGMANDATE',  
'COVID19',  
'COVID19',  
'UKRAINE',  
'COVID19',  
'GOODJOB COP',  
'SATURDAYMOTIVATION',  
'COVID19',  
'BBNAIJIA2020',  
'RIHANNAXTIWA',  
'ALREADYVIDEO',  
'COVID19',  
'COVID19',  
'COVID19',  
'WEARAMASK',  
'REALMENWEARMASKS',  
'TOUGHGUYSWEARMASKS',  
'COVID19',  
'COVID19',  
'TRAYVONMARTIN',  
'COVID19',  
'COVID19',  
'COVID19',  
'TESTING',  
'CYPRUS',  
'COVID19',  
'COVIDUPDATES',  
'COVID19',  
'APFIGHTSCORONA',  
'COVID19PANDEMIC',  
'BBCAQ',  
'DISTRESS',  
'WEARINGAMASK',  
'WEARAMASK',  
'VIRALLOAD',  
'COVID19',  
'KERRY',  
'STAYCATION2020',  
'CLOGHANE',  
'COVID19',  
'BIGPHARMA',  
'WHO',  
'5G',  
'COVID19',  
'TOPNEWS',  
'RAJASTHAN',  
'EXGOP-ER',  
'PANDEMIC',  
'COVID19',  
'REOPENINGSCHOOLS',  
'COVID19',  
'NEWYORK',  
'COVID19',  
'FRANCE',  
'COVID19',  
'CATHOLIC',  
'CHRISTIAN',

'TRUTH',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'CORONAVIRUS',  
'CANTSTOPTHETRUTH',  
'COVID19',  
'COVID19',  
'CROATIA',  
'COVID19',  
'COVID19',  
'COVID19',  
'STOCKMARKET',  
'ANDHRAPRADESH',  
'COVID19',  
'WEARAMASK',  
'REALMENWEARMASKS',  
'TOUGHGUYSWEARMASKS',  
'MASKSSAVELIVES',  
'ONTARIO',  
'GOP',  
'COVID19',  
'QUARANTINEPORTRAITS',  
'QUARANTINE',  
'DAILYPORTRAIT',  
'COVID19',  
'FINEART',  
'US',  
'COVID19',  
'CORONAVIRUS',  
'CORONAVIRUSUPDATE',  
'COVID19',  
'CORONAVIRUSPANDEMIC',  
'COVID19',  
'COVID19',  
'COVID19',  
'TELECOMMUTING',  
'REMOTETEACHING',  
'REMOTELEARNING',  
'BARBADOS',  
'BEACHOFFICE',  
'ANNAPOLIS',  
'COVID19',  
'COVID19',  
'BICOLNEWS',  
'COVID19',  
'CLIMATECHANGE',  
'HURRICANEHANNA',  
'COVID19',  
'COVID19',  
'SOUTHCAROLINA',  
'HEALTHCARE',  
'COVID19',  
'CORONAVIRUS',  
'ANDREWMLANGENI',  
'COVID19',  
'COVID19',  
'COVID19',  
'DELHI',  
'COVID19',



'KASHMIR',  
'SRINAGAR',  
'COVID19',  
'COVID19',  
'KAGA',  
'MAGA',  
'COVID19',  
'COVID19',  
'ANDHRAPRADESH',  
'COVID19',  
'AUSOPEN',  
'AUSTRALIANOPEN2021',  
'CORONAVIRUS',  
'COVID19',  
'OBSESE',  
'CORONAVIRUS',  
'ANIMSXAVI',  
'COVID19',  
'ISTHISSERIOUS',  
'GAMBIA',  
'COVID19',  
'COVID19',  
'COVID19',  
'SUMMER2020',  
'COVID19',  
'WEATHER',  
'GARDENPARTY',  
'COVID19',  
'KATIEHOPKINS',  
'LUFC',  
'COVID19',  
'VOTEBLUETOENDTHISNIGHTMARE',  
'COVID19',  
'BESAFE',  
'PERAKTBG',  
'ABSTRACTMAX',  
'PUMA',  
'PUMAONE',  
'COVID19',  
'JANASENA',  
'POWERSTAR',  
'VAKEELSAAB',  
'TRUMPDEATHTOLL147K',  
'COVID19',  
'TRUMPVIRUSCATASTROPHE',  
'COVID19',  
'COVID19',  
'UK',  
'LEFTYTOSSERS',  
'BORIS',  
'COVID19',  
'COVID19',  
'AHMEDABADRAIN',  
'COVID19',  
'REOPENSAFELY',  
'REOPENINGSCHOOLS',  
'HEALTH',  
'COVID19',  
'COVID19',  
'COVID19',

'CORONAVIRUS',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID',  
'CORONAVIRUS',  
'BLACKLIVESMATTER',  
'SATCHAT',  
'PTECH',  
'WEAREPTECH',  
'GOSPELHOUSE',  
'GODGOTTHIS',  
'COVID19',  
'COVID19',  
'HOW',  
'HEIR',  
'HAVE',  
'LITTLE',  
'KNOWLEDGE',  
'THINKS',  
'ANIMALS',  
'UNIVERSITY',  
'COVID19',  
'BERKLEYFORUM',  
'COVID19',  
'HOPE',  
'USA',  
'TRUMP',  
'USPS',  
'AMAZON',  
'POSTMASTER',  
'ANDHRAPRADESH',  
'COVID19',  
'CORONAVIRUS',  
'ANDORRA',  
'COVID19',  
'COVID19',  
'LONDON',  
'COVIDUPDATES',  
'DRUMPF',  
'TEXASISHEELING',  
'PIC',  
'ASSAM',  
'ODISHA',  
'COVID19',  
'COVID19',  
'WORSTPRESIDENTINHISTORY',  
'TRUMPDEATHTOLL147K',  
'COVID19',  
'COVID19',  
'THEYHAVENOIDEA',  
'MASK',  
'NOMASKS',  
'MP',  
'SHIVRAJSINGHCHOUHAN',  
'COVID19',  
'BLUEWAVE2020',  
'COVID19',  
'COVID19',

'ANDHRAPRADESH',  
'COVID19',  
'LIGHTSKINNED',  
'COVID19',  
'VISHAL',  
'COVID19',  
'JANASENA',  
'POWERSTAR',  
'VAKEELSAAB',  
'WWE',  
'WRESTLEMANIA',  
'COVID19',  
'CORONAVIRUS',  
'PRODUCTIVE',  
'COVID19',  
'COVID19',  
'KUPWARA',  
'CORONAVIRUSUPDATE',  
'FIGHTAGAINSTCOVID19',  
'PANDAMIC',  
'ADR',  
'EXAMS',  
'COVID19',  
'COVID19',  
'COVID19',  
'AIRLINES',  
'CORONAVIRUS',  
'COVID19',  
'MASK',  
'TRUMP',  
'CNN',  
'AUDUBON',  
'SATURDAYMORNING',  
'SATURDAYVIBES',  
'COVID19',  
'SATURDAYTHOUGHTS',  
'SATURDAYMOTIVATION',  
'FREE',  
'ORANGEBABOON',  
'COVID19',  
'VACCINE',  
'COVID19',  
'COVID19',  
'ECONOMICRECOVERY',  
'COVID19',  
'AMERICAN',  
'COVID19',  
'DORATHIANS',  
'COIMBATORE',  
'LOCKDOWN',  
'TRUMPDEATHTOLL148K',  
'COVID19',  
'TRUMPPANDEMICFAILURE',  
'VIRUS-KILLING',  
'CORONAVIRUS',  
'COVID19',  
'HURRICANE',  
'COVID19',  
'FREEPALESTINE',  
'COVID19',

'PALESTINIANSLIVESMATTER',  
'ENDAPARTHEIDNOW',  
'CORONAVIRUS',  
'CORONAVIRUSUPDATE',  
'COVID19',  
'CORONAVIRUSPANDEMIC',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'PROTESTS',  
'COVID19',  
'BRAZIL',  
'BILSONARO',  
'COVID19',  
'DELHI',  
'COVID19',  
'COVID19',  
'STAYHOME',  
'WASHYOURHANDS',  
'PHYSICALDISTANCING',  
'COVID19',  
'COVIDVIC',  
'COVID19',  
'COVID19',  
'JIB',  
'COVID19',  
'MLB',  
'DH',  
'NATIONALLEAGUE',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'MONEYMATTERS',  
'INSANITY',  
'COVID19',  
'COVID',  
'PURPOSE',  
'COVID19',  
'PURI',  
'COVID19',  
'ODISHA',  
'COVID19',  
'COVID19',  
'STAYATHOMEANDSTAYSAFE',  
'STAYHOMEZIM',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'INDIA',  
'MAHARASHTRA',  
'DELHI',  
'GOI',  
'MUMBAI',  
'LOCAL',  
'COVID19',  
'SUPPORTSMALLBUSINESSES',

'COVID19',  
'COVID19',  
'FIRE',  
'PPE',  
'COVID19',  
'KNOWTHEFACTS',  
'NSTNATION',  
'FARMER',  
'DAUGHTERS',  
'COVID19',  
'SUICIDE',  
'CANCER',  
'COVID19',  
'VISHAL',  
'MAILINVOTING',  
'FRAUD',  
'COVID19',  
'TRUMP2020',  
'JOKE',  
'WETHEPEOPLE',  
'RUSSIA',  
'RUSSIA',  
'RUSSIA',  
'SCHIFF',  
'RUSSIANCOLLUSION',  
'INTERFERENCE',  
'COVID19',  
'POTUS',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID',  
'GEORGIA',  
'COVID19',  
'CORONAVIRUSUPDATE',  
'CORONAVIRUSPANDEMIC',  
'CORONAVIRUS',  
'COVID19',  
'FLORIDA',  
'COVID19',  
'OBESITY',  
'CARRYMINATI',  
'MY200VSBUTCH200',  
'COVID19',  
'COVID19',  
'TEACHERS',  
'COVID19',  
'GLOBALPANDEMIC',  
'COVID19',  
'ORACLECLOUD',  
'COVID19',  
'TEXASCOVID19',  
'COVID19',  
'COVID19',  
'KIDS',  
'CORONAVIRUS',  
'VISHAL',  
'COVID19UK',  
'TORYBREXIT',

'HEALTHWORKERS',  
'COVID19',  
'THEATREINAFRICA',  
'COVID19',  
'PAKISTAN',  
'COVID19',  
'IRELAND',  
'2',  
'MASKUP',  
'MASK',  
'COVID19',  
'CORONAVIRUS',  
'LONDON',  
'ARTWORK',  
'ENTAGLEMENT',  
'RIHANNAXTIWA',  
'OBASEKIFLAGSOFTODAY',  
'COVID19',  
'BBNAIJIA2020',  
'BBNAJIA',  
'LAYCON',  
'DORA',  
'COVID19',  
'COVID19',  
'GOODQUESTION',  
'COVID19',  
'KBF',  
'SCAMDEMIC',  
'NOMASKS',  
'COVID19',  
'BORISVOTEONCONFIDENCE',  
'BORISHASFAILEDUK',  
'SECONDPEAK',  
'SECONDWAVE',  
'CORONAVIRUSUK',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'COVID19',  
'CALLTHEMFIRST',  
'SELECTRESTORATION',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'AUSTRALIA',  
'COVID19',  
'LOCKDOWNMELBOURNE',  
'QUARANTINELIFE',  
'NEWYORK',  
'CORONAVIRUS',  
'GUERNSEY',  
'CORONA',  
'COVID19',  
'BOUNTYONOURMILITARY',  
'TRE45SON',  
'COVID19',  
'COVID19',  
'MADHYAPRADESHCM',

'SHIVRAJSINGHCHOUHAN',  
'COVID19',  
'COVID19',  
'MORONS',  
'ETHNIC',  
'MAGA',  
'COVID19',  
'COVID19',  
'PA18',  
'PAPOL',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'COVID19USA',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'PLANDEMIA',  
'SCAM',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'DIGITALTRANSFORMATION',  
'DIGITAL',  
'COVID19',  
'COVID19',  
'COVID19',  
'BANGLADESH',  
'COVID19',  
'SUNBELT',  
'CORONAVIRUS',  
'COVID19',  
'COVID19',  
'COVID19',  
'COVID19',  
'JAMMUANDKASHMIR',  
'REFERENDUM2020',  
'COVID19',  
'VACCINES',  
'COVID19',  
'WEAREINTHISTOGETHER',  
'TRUCKING',  
'COVID19',  
'SETU\_APP',  
'SHIVRAJSINGHCHOUHAN',  
'COVID19',  
'TRUMP',  
'ISRAEEL',  
'AMERICAN',  
'BLACKTRANSLIVESMATTER',  
'IRAN',  
'NETANYAHU',  
'COVID19',  
'COVID19',  
'CORONAVIRUS',  
'COVID19',  
'FRONTLINEWORKERS',

```
'RUNNING',
'SINGING',
'ANXIETY',
'COVID19',
'PANDEMIC',
'COVID19',
'AMERICA',
'COVID19',
'MARXISTBLM',
...]
```

In [25]:

```
our_freq = FreqDist(our_hashtags_in_tweets)
dictionary_hashtags = pd.DataFrame({ "Hashtag" : list(our_freq.keys()), "Count" : list(our_freq.values()) })
dictionary_hashtags
```

Out[25]:

	Hashtag	Count
0	COVID19	100055
1	CORONAVIRUSUPDATES	521
2	CORONAVIRUS	10184
3	RECRUITING	3
4	COVIDPOSITIVE	4
...	...	...
37903	AGECONMATTERS	1
37904	LAWLESSTRUMP	1
37905	COMMUNITYHEALTH	1
37906	KEEPTHESSCHOOLSCLOSED	1
37907	KEEPTEACHERSALIVE	1

37908 rows × 2 columns

In [26]:

```
nlargest_dictionary_hashtags = dictionary_hashtags.nlargest(columns='Count', n = 15)
nlargest_dictionary_hashtags
```

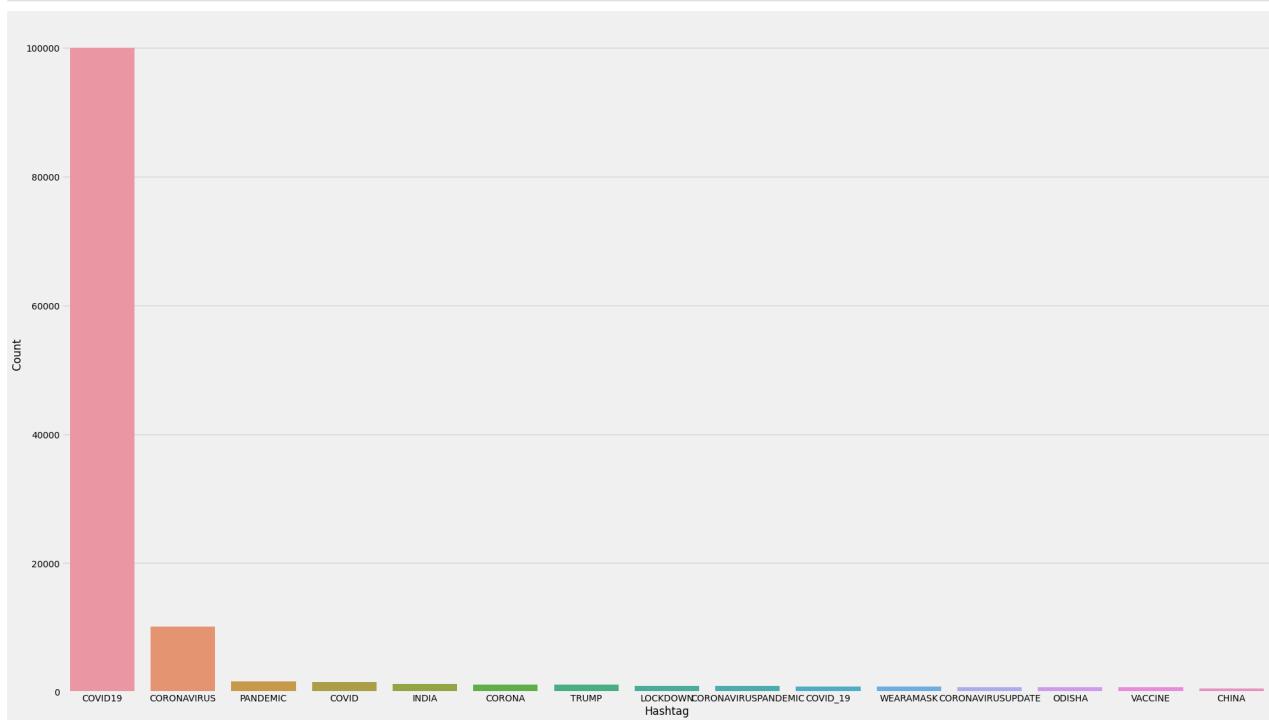
Out[26]:

	Hashtag	Count
0	COVID19	100055
2	CORONAVIRUS	10184
189	PANDEMIC	1623
77	COVID	1568
245	INDIA	1191
417	CORONA	1159
165	TRUMP	1099
343	LOCKDOWN	961

	Hashtag	Count
112	CORONAVIRUSPANDEMIC	882
255	COVID_19	827
115	WEARAMASK	801
111	CORONAVIRUSUPDATE	721
318	ODISHA	708
261	VACCINE	702
622	CHINA	541

In [44]:

```
plt.figure(figsize=(30,18))
sns.barplot(data=nlargest_dictionary_hashtags, x="Hashtag", y="Count")
plt.show()
```



In [28]:

```
nlargest_dictionary_hashtags = dictionary_hashtags.nlargest(columns='Count', n = 15)
nlargest_dictionary_hashtags
```

Out[28]:

	Hashtag	Count
0	COVID19	100055
2	CORONAVIRUS	10184
189	PANDEMIC	1623
77	COVID	1568
245	INDIA	1191
417	CORONA	1159

	Hashtag	Count
165	TRUMP	1099
343	LOCKDOWN	961
112	CORONAVIRUSPANDEMIC	882
255	COVID_19	827
115	WEARAMASK	801
111	CORONAVIRUSUPDATE	721
318	ODISHA	708
261	VACCINE	702
622	CHINA	541

In [31]:

```

import pandas as pd; pd.set_option("display.max_rows", None, "display.max_columns", Non
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

def readData(file):
    data = pd.read_csv(file)
    return data

def tweet_preparation():
    tweets = readData("covid19_tweets2.csv")
    tweets[['Date', 'Hour']] = tweets['date'].str.split(' ', expand=True)
    tweets.set_index('Date')
    tweets.index = pd.to_datetime(tweets.index)
    tweets[['Count']] = 1
    grouped_tweets = tweets.groupby(['Date']).count()

    return grouped_tweets.Count.to_frame(name='tweets')

def cases_preparation():
    df = readData("time_series_covid19_confirmed_global.csv")
    data = df.loc[:, '7/24/20':'8/30/20']

    data = data.T.sum(axis = 1)
    data = data.to_frame(name = 'cases')
    data.index = pd.to_datetime(data.index)
    return data

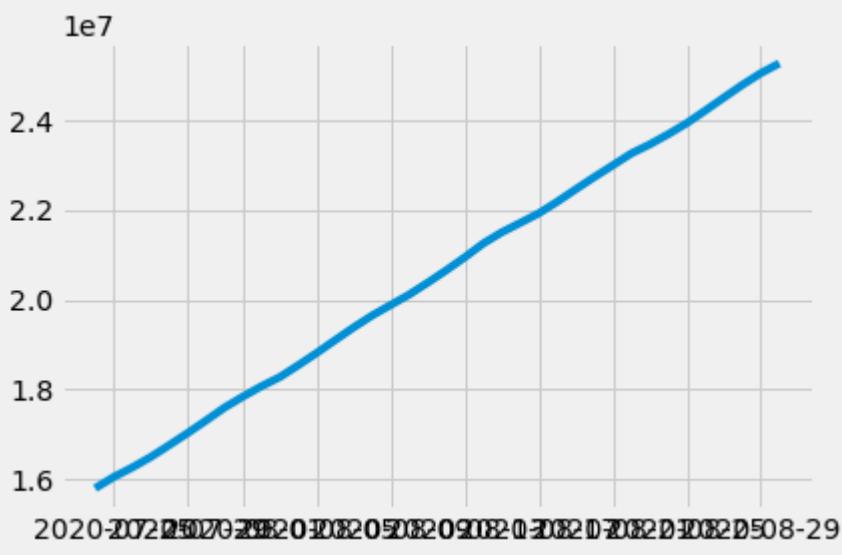
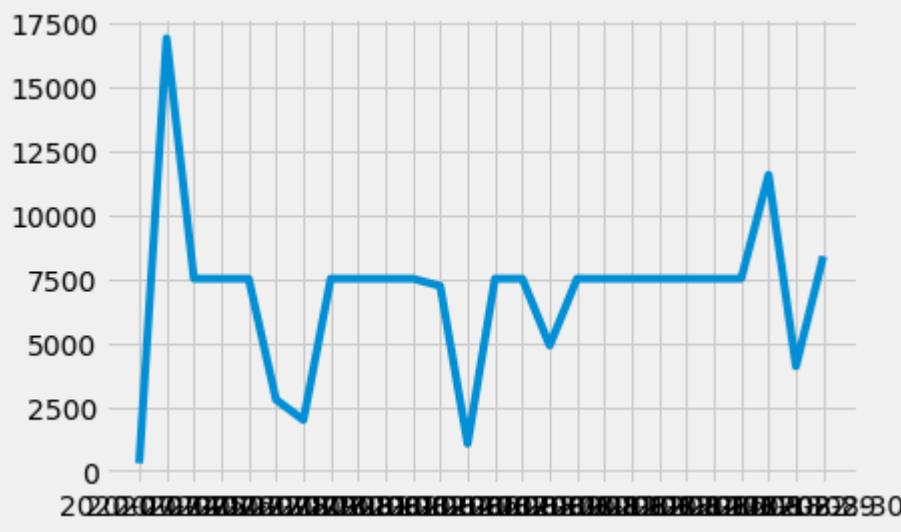
data_tweets = tweet_preparation()
plt.plot(data_tweets)
plt.show()

data_cases = cases_preparation()
plt.plot(data_cases)
plt.show()

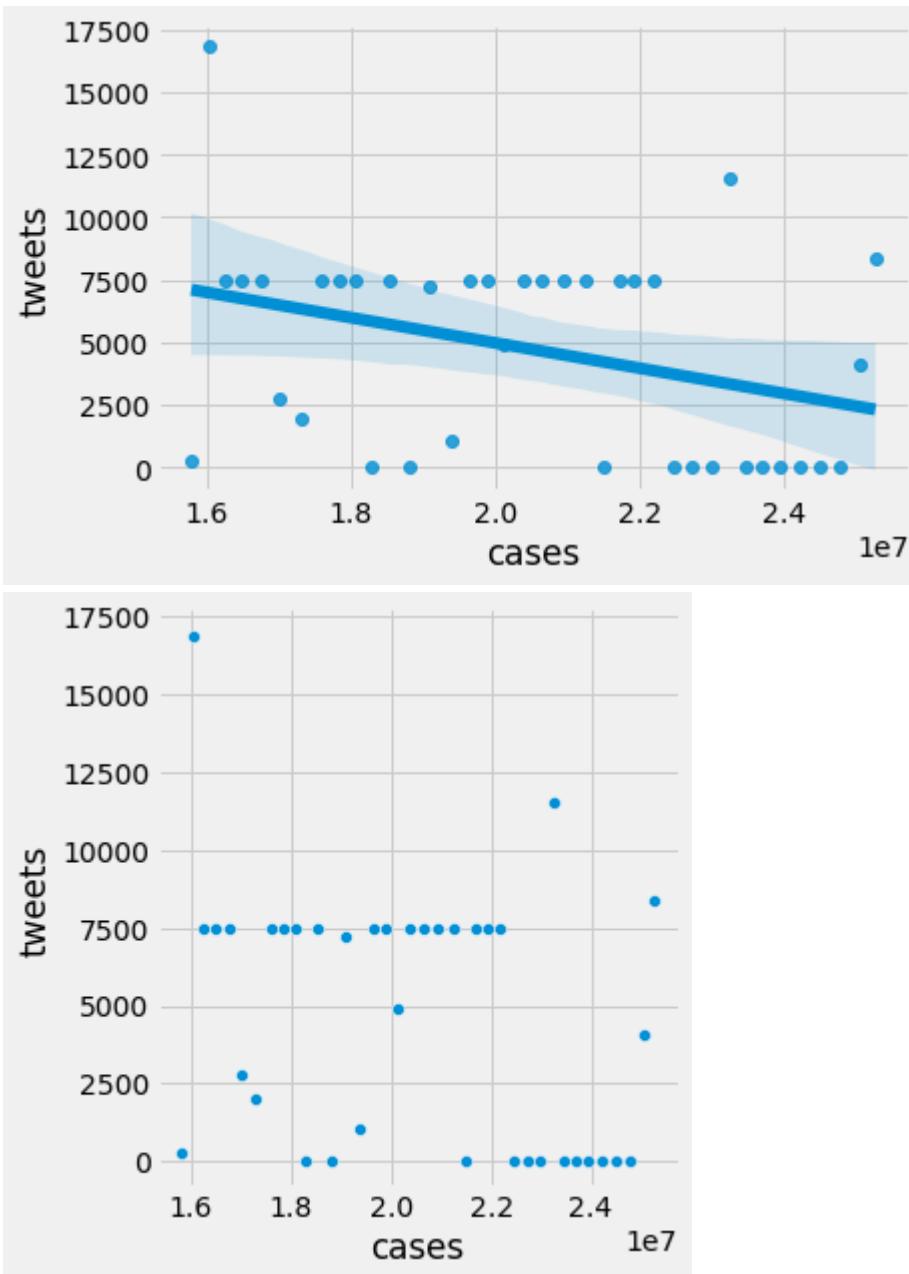
data_cases.index = data_cases.index.strftime('%Y-%m-%d')

```

```
df_corr = pd.concat([data_tweets, data_cases], axis=1)
df_corr.fillna(0, inplace=True)
df_corr
sns.regplot(x=df_corr["cases"], y=df_corr["tweets"])
sns.relplot(x=df_corr["cases"], y=df_corr["tweets"])
```



Out[31]: <seaborn.axisgrid.FacetGrid at 0x18533ada770>



In [32]:

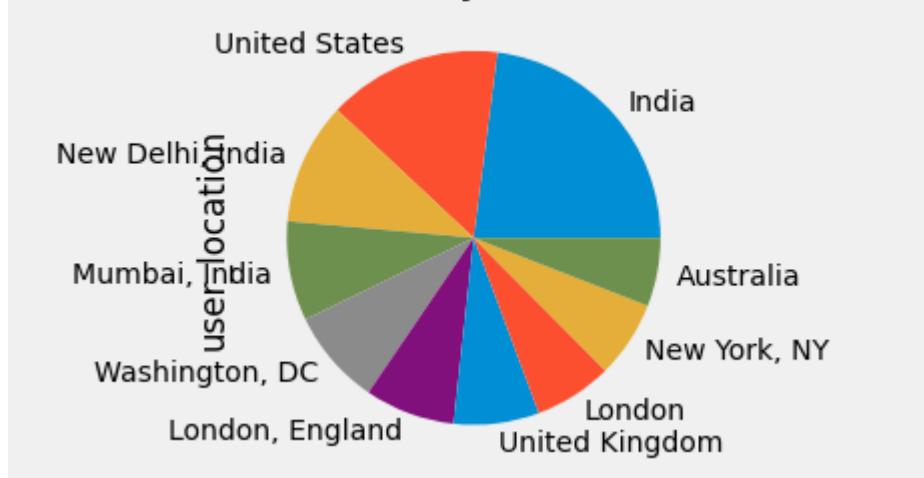
```
df_data = pd.DataFrame(data)
locations = df_data['user_location'].value_counts()
locations[0:10]

# Grafica
fig1, ax1 = plt.subplots()
locations[0:10].plot(kind='pie', title="localidades con la mayor cantidad de tweets")
```

Out[32]:

```
<AxesSubplot:title={'center':'localidades con la mayor cantidad de tweets'}, ylabel='use
r_location'>
```

## localidades con la mayor cantidad de tweets



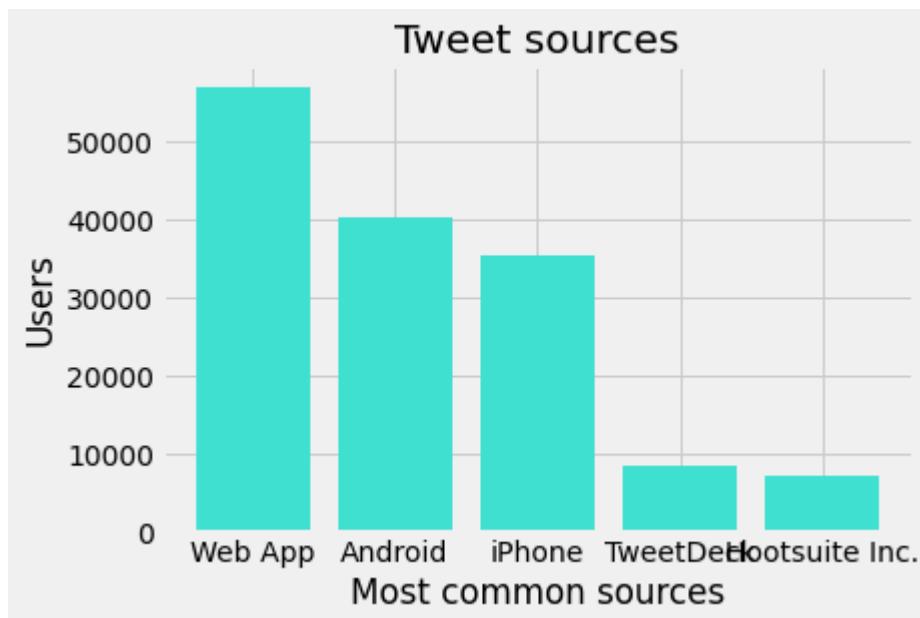
```
In [35]: data = pd.read_csv('covid19_tweets2.csv')
```

```
In [36]: sources = data.pivot_table(index=['source'], aggfunc='size')
sources.sort_values()

nameLabels_x = ["Web App", "Android", "iPhone", "TweetDeck", "Hootsuite Inc."]

numberSources_y = []
numberSources_y.append(sources["Twitter Web App"])
numberSources_y.append(sources["Twitter for Android"])
numberSources_y.append(sources["Twitter for iPhone"])
numberSources_y.append(sources["TweetDeck"])
numberSources_y.append(sources["Hootsuite Inc."])

plt.bar(nameLabels_x, numberSources_y, color='turquoise')
plt.title("Tweet sources")
plt.xlabel("Most common sources")
plt.ylabel("Users")
plt.show()
```



```
In [37]: username = data['user_name'].value_counts().reset_index()
username = pd.DataFrame(username)
username.columns = ['Twitter UserName', 'Number of Tweets']
username_10 = username.head(10)
username_10
```

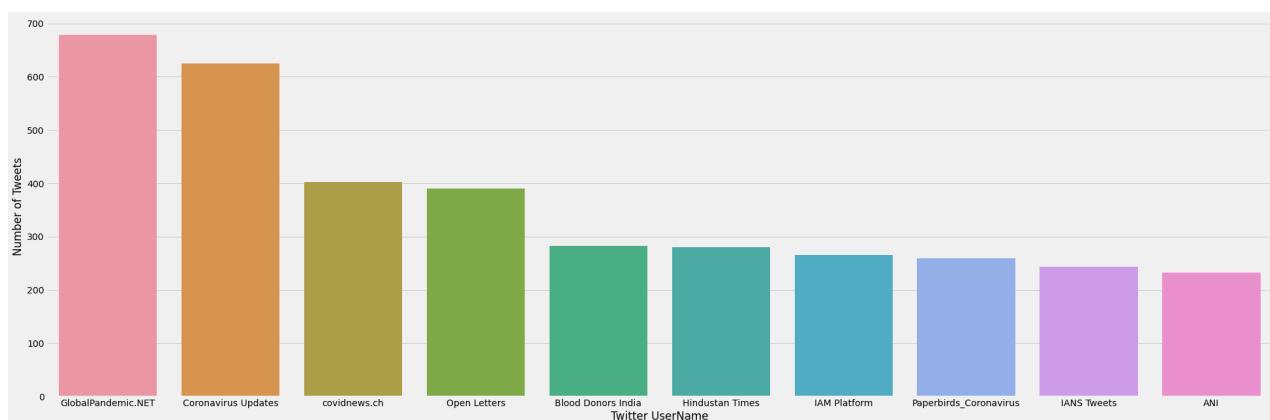
Out[37]:

	Twitter UserName	Number of Tweets
0	GlobalPandemic.NET	679
1	Coronavirus Updates	625
2	covidnews.ch	402
3	Open Letters	390
4	Blood Donors India	282
5	Hindustan Times	280
6	IAM Platform	266
7	Paperbirds_Coronavirus	259
8	IANS Tweets	244
9	ANI	233

In [43]:

```
plt.figure(figsize = (30,10))
sns.barplot(x = 'Twitter UserName', y = 'Number of Tweets', data = username_10)
```

Out[43]:



In [39]:

```
followers = data[['user_name', 'user_followers']].copy()
followers = followers.drop_duplicates(subset = ['user_name'])
followers = followers.sort_values(by = ['user_followers'], ascending = False)
followers_10 = followers.head(10)
followers_10
```

Out[39]:

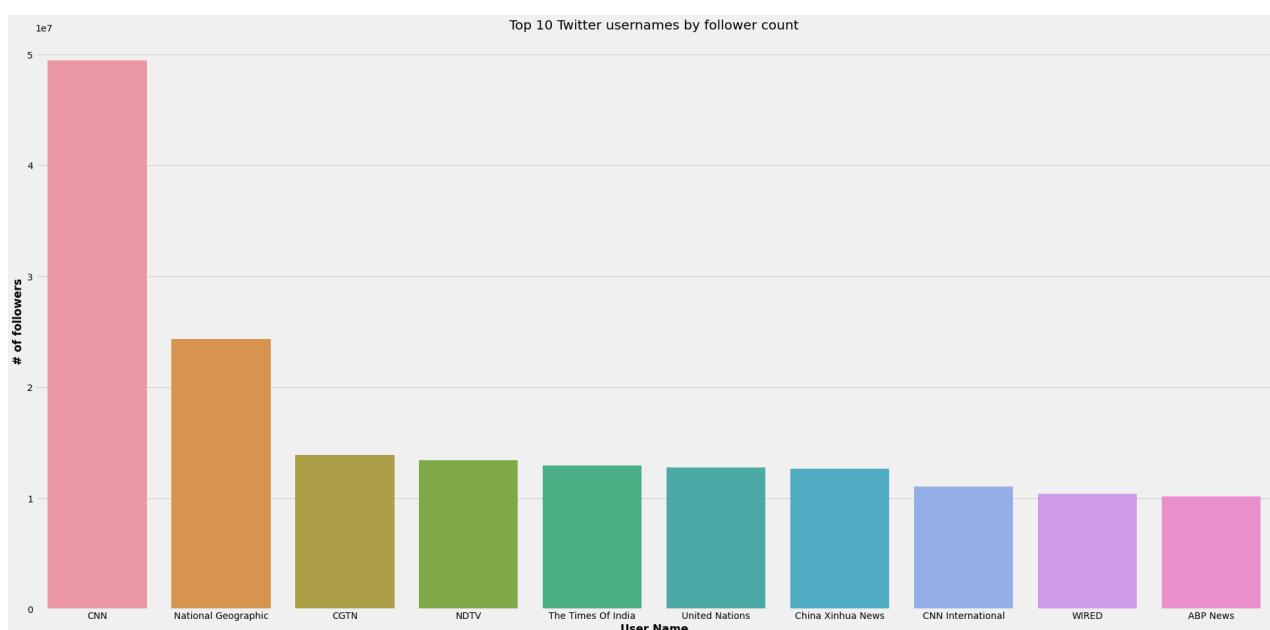
	user_name	user_followers
104946	CNN	49442559
174909	National Geographic	24359165
235	CGTN	13892795

	user_name	user_followers
<b>1161</b>	NDTV	13417047
<b>785</b>	The Times Of India	12978547
<b>3872</b>	United Nations	12748145
<b>3251</b>	China Xinhua News	12681122
<b>105719</b>	CNN International	11081219
<b>129519</b>	WIRED	10380131
<b>13557</b>	ABP News	10172489

In [40]:

```
plt.figure(figsize = (30,15))
sns.barplot(x = 'user_name', y = 'user_followers', data = followers_10)
plt.xlabel('User Name', fontweight = 'bold')
plt.ylabel('# of followers', fontweight = 'bold')
plt.title('Top 10 Twitter usernames by follower count')
```

Out[40]:



In [41]:

```
tweets = data[['text', 'user_favourites']].copy()
tweets = tweets.drop_duplicates(subset = ['text'])
tweets = tweets.sort_values(by = ['user_favourites'], ascending = False)
tweets_10 = tweets.head(10)
tweets_10
```

Out[41]:

	text	user_favourites
<b>14504</b>	So Trump wants everyone else's child to get #C...	2047197
<b>173409</b>	How do you pioneer a growing fatality rate and...	1320166
<b>172861</b>	Interesting, since some say asymptomatic carri...	1320166
<b>172825</b>	So many conflicting reports, so much we don't ...	1320166

		text	user_favourites
<b>148145</b>	Herd immunity at 20% is still too high and wou...	1317944	
<b>148540</b>	Sounds like the need for a booster shot, once ...	1317944	
<b>177848</b>	Nuovo rimbalzo del #covid19 in #Turchia Corona...	1161820	
<b>168379</b>	Sotto i 300 casi #covid19 #SudCorea S. Korea c...	1161816	
<b>156746</b>	Il rapporto tra casi accertati e mortalità è e...	1159748	
<b>141678</b>	Mortalità molto elevata rispetto ai casi #covi...	1158416	

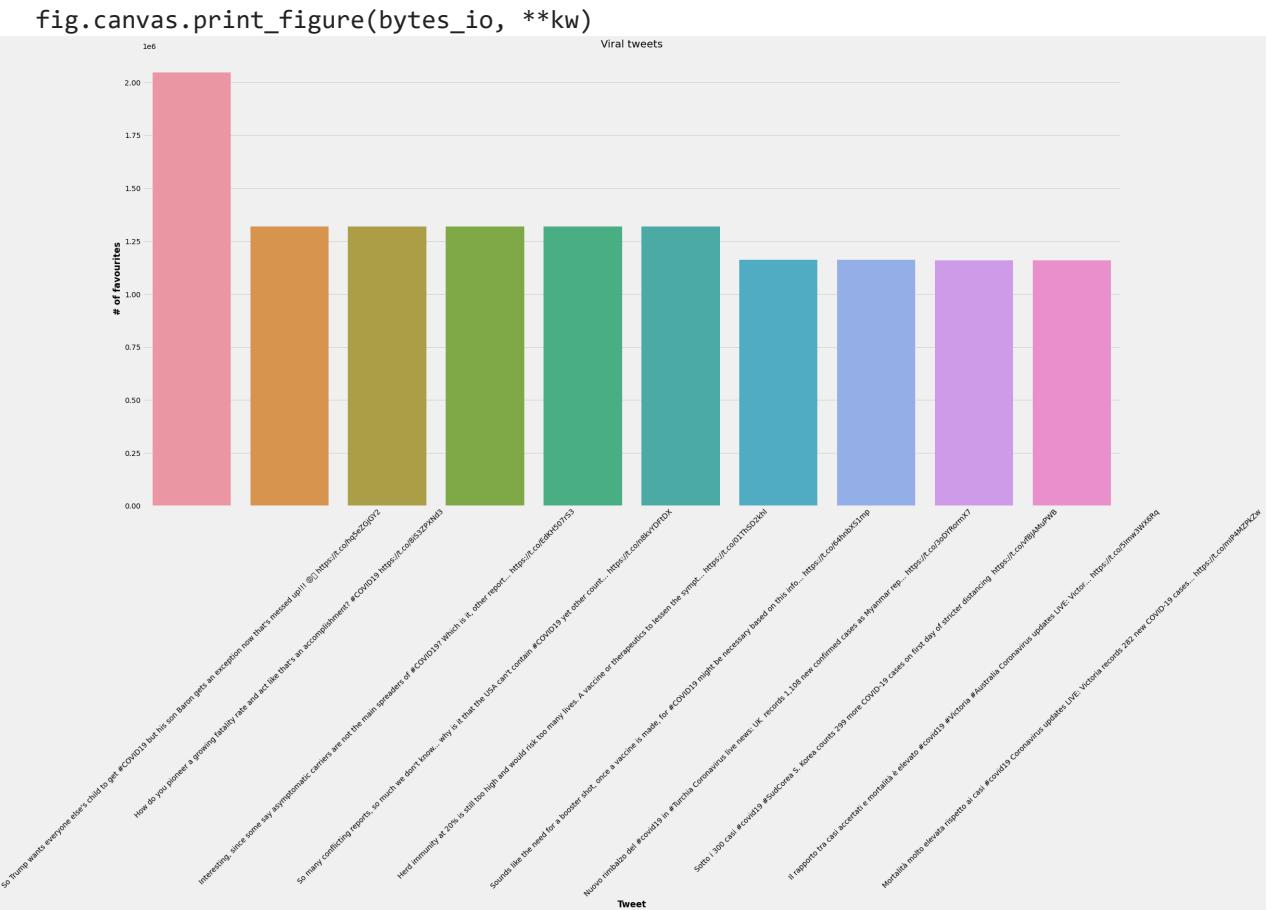
In [42]:

```
plt.figure(figsize = (30, 15))
sns.barplot(x = 'text', y = 'user_favourites', data = tweets_10)

plt.xlabel('Tweet', fontweight = 'bold')
plt.ylabel('# of favourites', fontweight = 'bold')
plt.xticks(rotation = 45)
plt.title('Viral tweets')
```

Out[42]:

C:\Users\Guillermo\AppData\Roaming\Python\Python310\site-packages\IPython\core\pylabtool s.py:151: UserWarning: Glyph 128548 (\N{FACE WITH LOOK OF TRIUMPH}) missing from current font.



In [26]:

```

import nltk
from wordcloud import WordCloud
import pandas as pd
import seaborn as sns
import numpy as np
import re
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans

%matplotlib inline
from mpl_toolkits.mplot3d import Axes3D
plt.rcParams['figure.figsize'] = (16, 9)
plt.style.use('ggplot')

```

In [19]:

```

dataframe = pd.read_csv("covid19_tweets2.csv")
dataframe.head()

```

Out[19]:

		user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites
0	Violi€†	astroworld	wednesday addams as a disney princess keepin i...		2017-05-26 05:46:42	624	950	187
1	Tom Basile us	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	72
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...		2019-03-07 01:45:06	197	987	14
4	DIPR-J&K	Jammu and Kashmir	Official Twitter handle of Department of Inf...		2017-02-12 06:45:15	101009	168	1



In [20]:

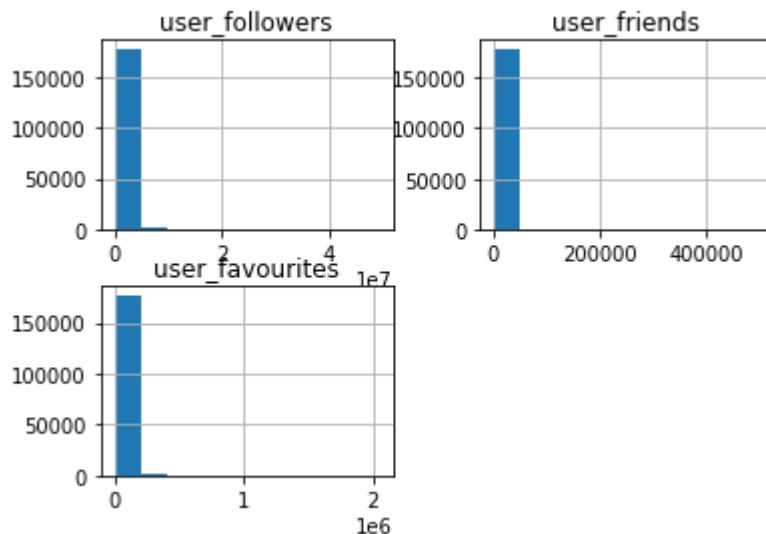
```

dataframe.drop(['user_verified'],1).hist()
plt.show()

```

C:\Users\Guillermo\AppData\Local\Temp\ipykernel\_13848\1697134265.py:1: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only.

```
dataframe.drop(['user_verified'],1).hist()
```

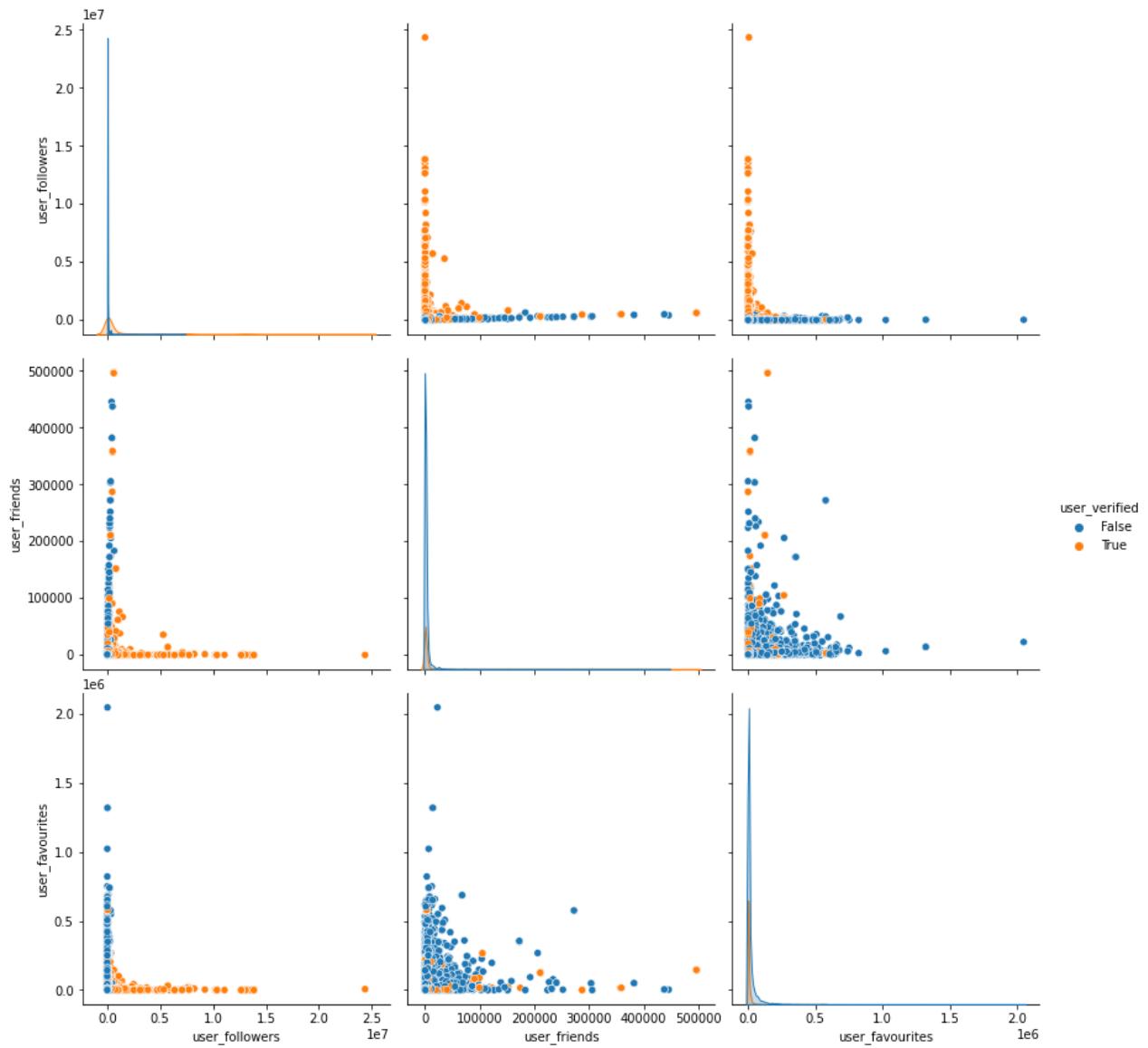


```
In [21]: sns.pairplot(dataframe.dropna(), hue='user_verified', size=4, vars=["user_followers", "use
```

C:\Users\Guillermo\AppData\Roaming\Python\Python310\site-packages\seaborn\axisgrid.py:20  
76: UserWarning: The `size` parameter has been renamed to `height`; please update your c  
ode.

```
    warnings.warn(msg, UserWarning)
```

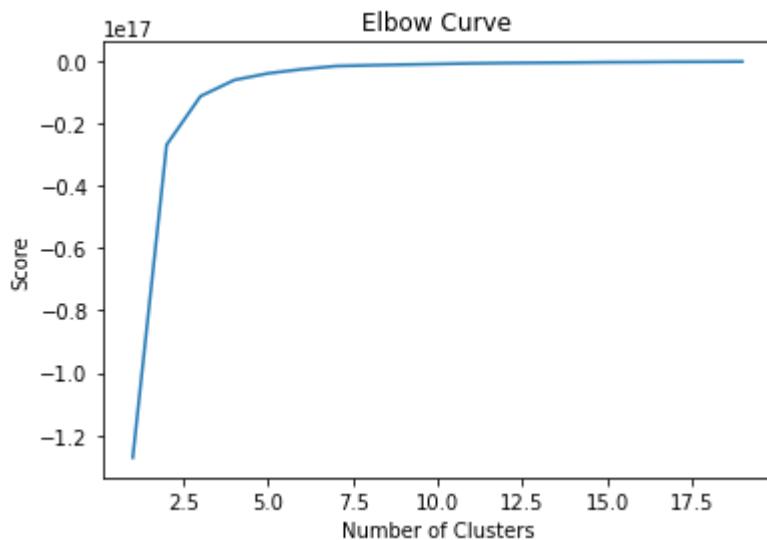
```
Out[21]: <seaborn.axisgrid.PairGrid at 0x17040ca7be0>
```



```
In [22]: X = np.array(dataframe[["user_followers", "user_friends", "user_favourites"]])
y = np.array(dataframe['user_verified'])
X.shape
```

```
Out[22]: (179108, 3)
```

```
In [23]: Nc = range(1, 20)
kmeans = [KMeans(n_clusters=i) for i in Nc]
kmeans
score = [kmeans[i].fit(X).score(X) for i in range(len(kmeans))]
score
plt.plot(Nc, score)
plt.xlabel('Number of Clusters')
plt.ylabel('Score')
plt.title('Elbow Curve')
plt.show()
```



```
In [24]: kmeans = KMeans(n_clusters=3).fit(X)
centroids = kmeans.cluster_centers_
print(centroids)
```

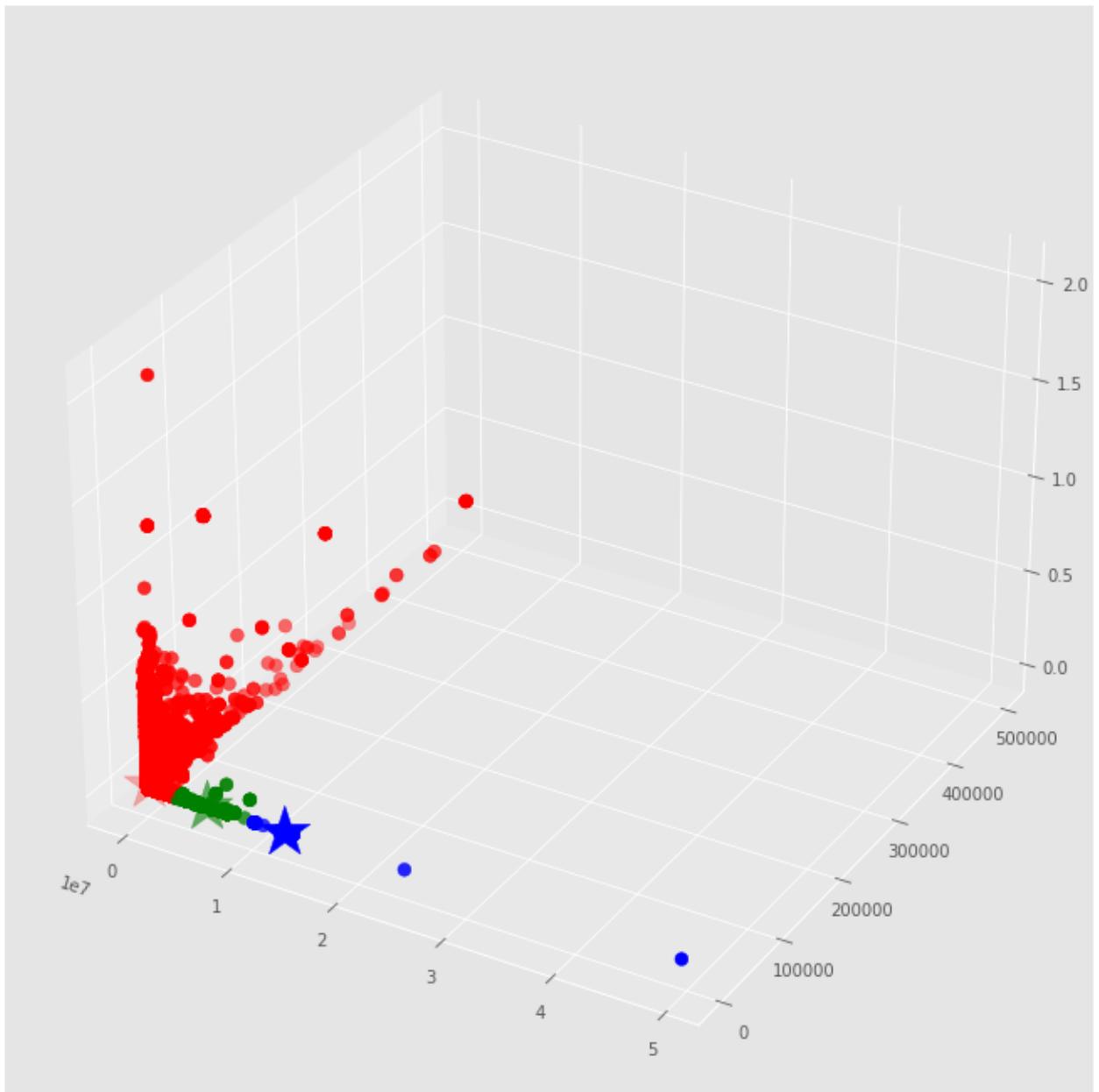
```
[[3.20424443e+04 2.13836740e+03 1.45774127e+04]
 [5.71070570e+06 6.64066007e+02 2.59558086e+03]
 [1.31393268e+07 2.40845960e+02 1.23171717e+02]]
```

```
In [27]: # Predicting the clusters
labels = kmeans.predict(X)
# Getting the cluster centers
C = kmeans.cluster_centers_
colores=['red','green','blue']
asignar=[]
for row in labels:
    asignar.append(colores[row])

fig = plt.figure()
ax = Axes3D(fig)
ax.scatter(X[:, 0], X[:, 1], X[:, 2], c=asignar,s=60)
ax.scatter(C[:, 0], C[:, 1], C[:, 2], marker='*', c=colores, s=1000)
```

C:\Users\Guillermo\AppData\Local\Temp\ipykernel\_13848\2144393596.py:11: MatplotlibDeprecationWarning: Axes3D(fig) adding itself to the figure is deprecated since 3.4. Pass the keyword argument auto\_add\_to\_figure=False and use fig.add\_axes(ax) to suppress this warning. The default value of auto\_add\_to\_figure will change to False in mpl3.5 and True values will no longer work in 3.6. This is consistent with other Axes classes.

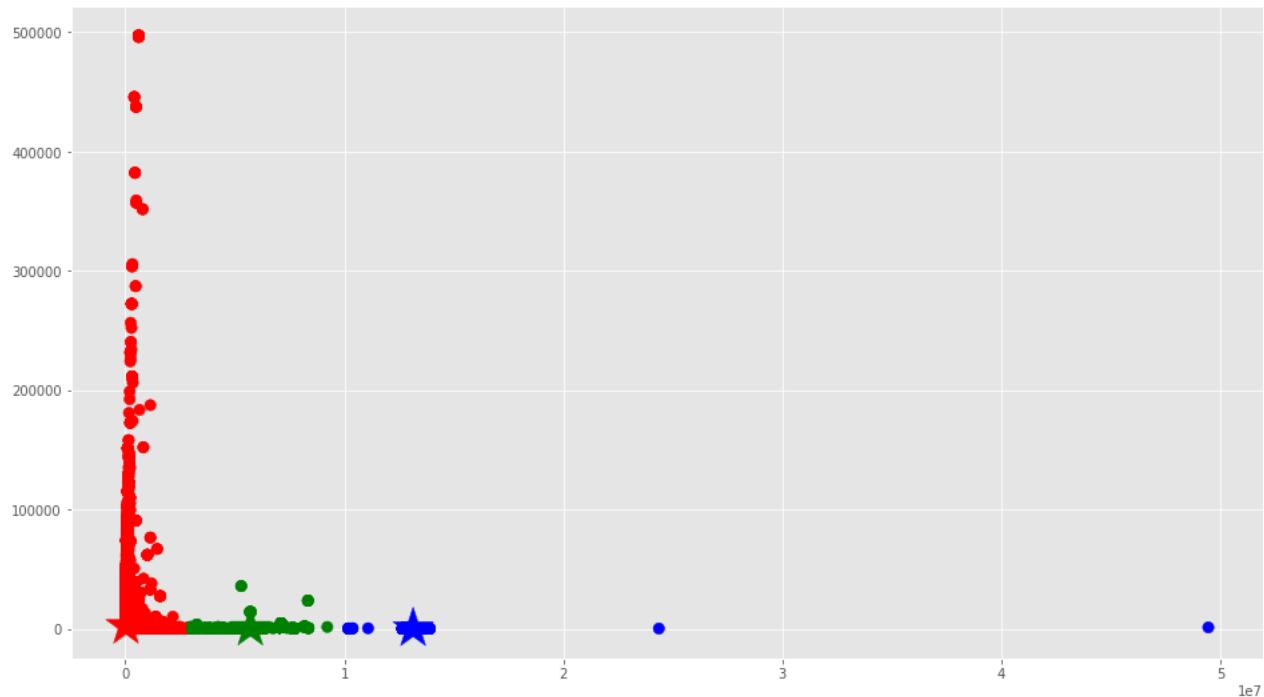
```
Out[27]: <mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x17055c32ad0>
```



In [28]:

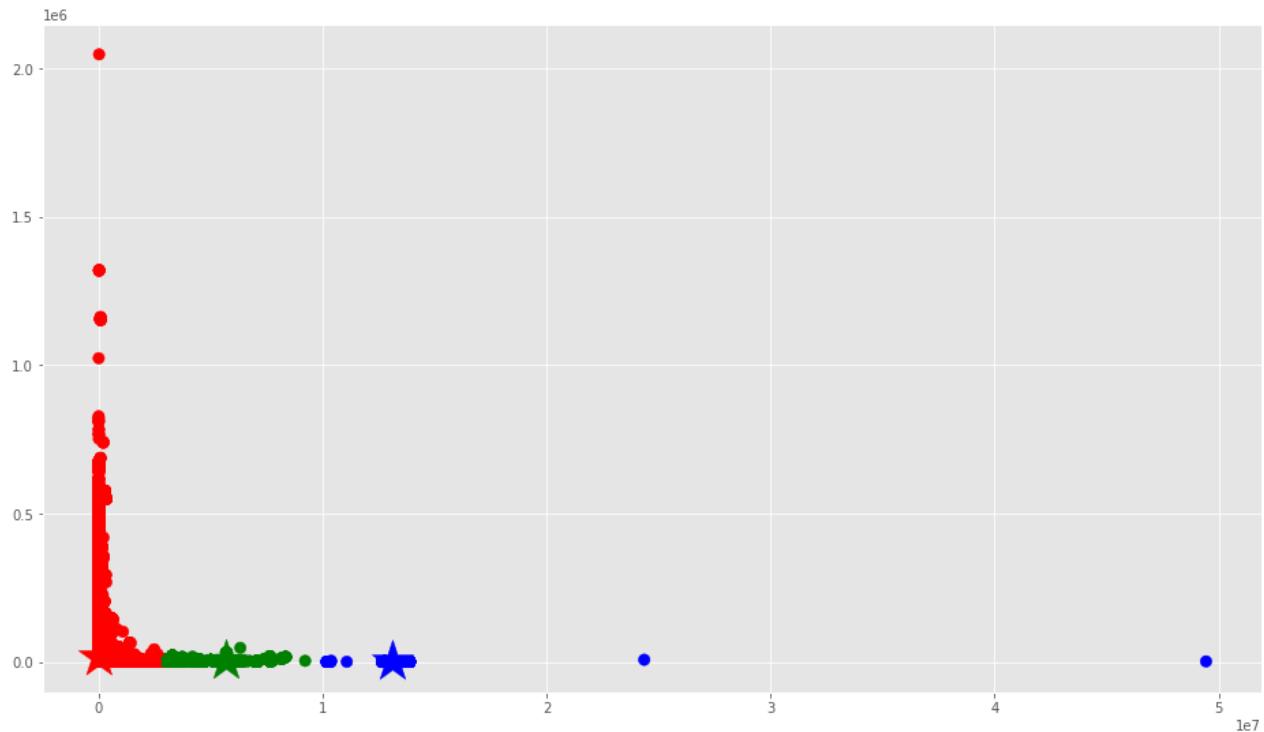
```
f1 = dataframe['user_followers'].values
f2 = dataframe['user_friends'].values

plt.scatter(f1, f2, c=asignar, s=70)
plt.scatter(C[:, 0], C[:, 1], marker='*', c=colores, s=1000)
plt.show()
```



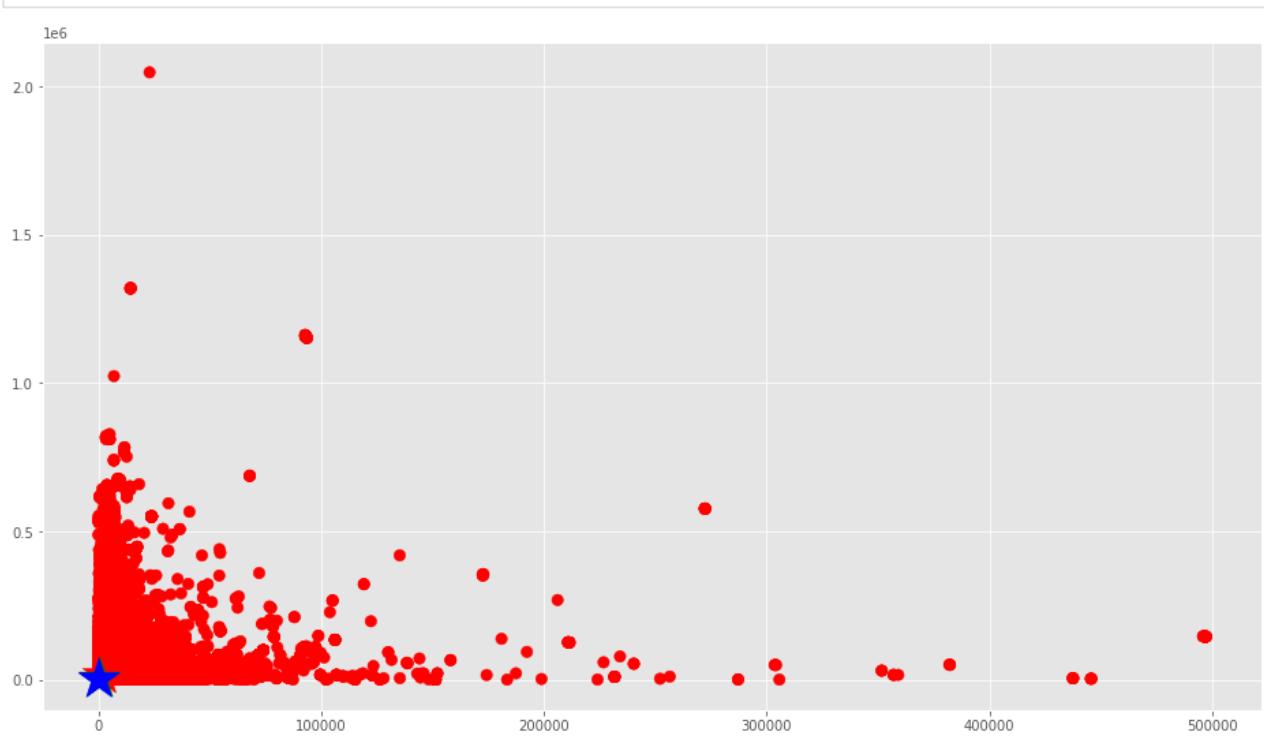
```
In [29]: f1 = dataframe['user_followers'].values
f2 = dataframe['user_favourites'].values

plt.scatter(f1, f2, c=asignar, s=70)
plt.scatter(C[:, 0], C[:, 2], marker='*', c=colores, s=1000)
plt.show()
```



```
In [30]: f1 = dataframe['user_friends'].values
f2 = dataframe['user_favourites'].values

plt.scatter(f1, f2, c=asignar, s=70)
plt.scatter(C[:, 1], C[:, 2], marker='*', c=colores, s=1000)
plt.show()
```



```
In [2]: from wordcloud import WordCloud, STOPWORDS
def word_cloud_func(data):
    word_cloud = WordCloud(
        background_color = 'black',
        stopwords = set(STOPWORDS),
        max_words = 50,
        max_font_size = 40,
    ).generate(str(data))
    fig = plt.figure(
        figsize = (20,20),
        facecolor = 'k',
        edgecolor = 'k')
    plt.axis('off')

    fig.subplots_adjust(top = 2.3)
    plt.imshow(word_cloud)
    plt.show()
```

```
In [3]: def text_cleaner(text):
    text = re.sub(r"\n", " ", str(text))
    text = re.sub(r"\t", " ", str(text))
    text = re.sub(r"\r", " ", str(text))
    text = re.sub(r"(@)|(#)|(RT[\s]+)|(https?:\/\/\S+)|([^\w\d-])", "", text)
    text = text.strip(" ")
    return text
```

```
In [4]: data = pd.read_csv('covid19_tweets2.csv')
vec = TfidfVectorizer(stop_words = "english")
vec.fit(data['text'].values)
features = vec.transform(data['text'].values)
```

## K-Means Clustering (2 Clusters)

```
In [5]: kmeans = KMeans(n_clusters = 2, random_state = 0)
kmeans.fit(features)
```

```
Out[5]: KMeans(n_clusters=2, random_state=0)
```

```
In [6]: temp = kmeans.predict(features)
data['Cluster'] = temp
data
```

		user_name	user_location	user_description	user_created	user_followers	user_friends	user_f...
0	Violi€†	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950		
1	Tom Basile us	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		
...	...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.	Ilorin, Nigeria	Animal Scientist   Muslim   Real Madrid/Chelsea	2013-12-30 18:59:19	412	1609		
179104	Jason	Ontario	When your cat has more baking soda than Ninja	2011-12-21 04:41:30	150	182		
179105	BEEHEMOTH	ca Canada	The Architects of Free Trade Really Did ...	2016-07-13 17:21:59	1623	2160		
179106	Gary DelPonte	New York City	Global UX UI Visual Designer. StoryTeller, Mus...	2009-10-27 17:43:13	1338	1111		

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_f...
179107	TUKY II	Aliwal North, South Africa	TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	2018-04-14 17:30:07	97	1697	

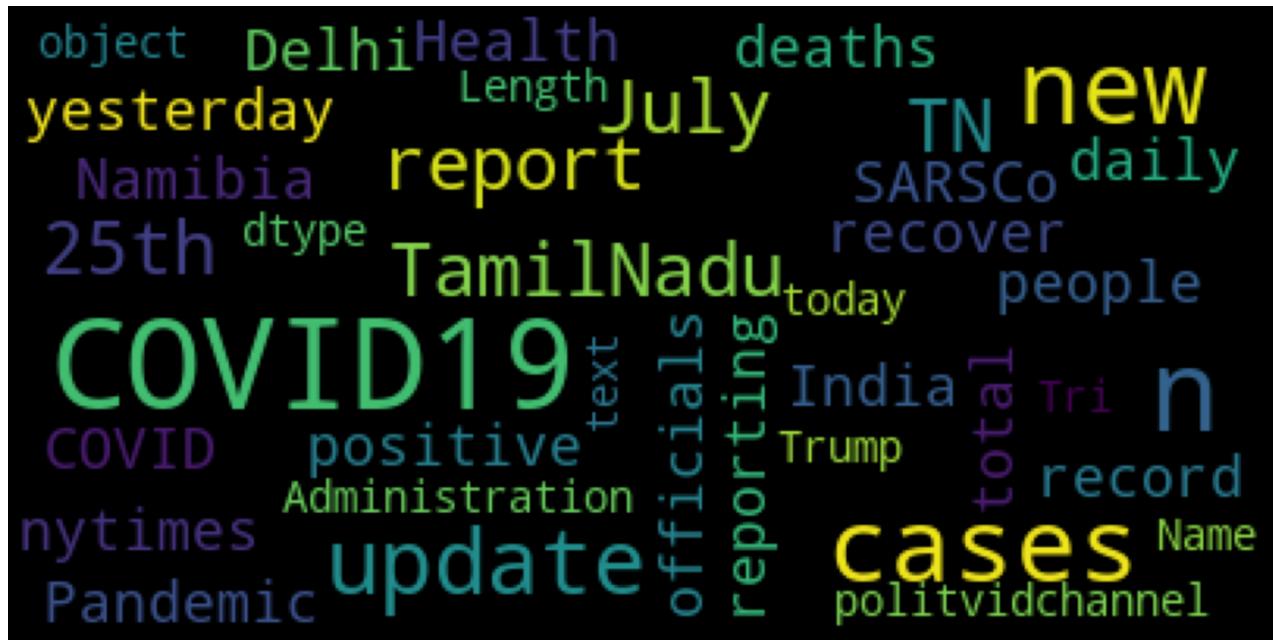
179108 rows × 14 columns

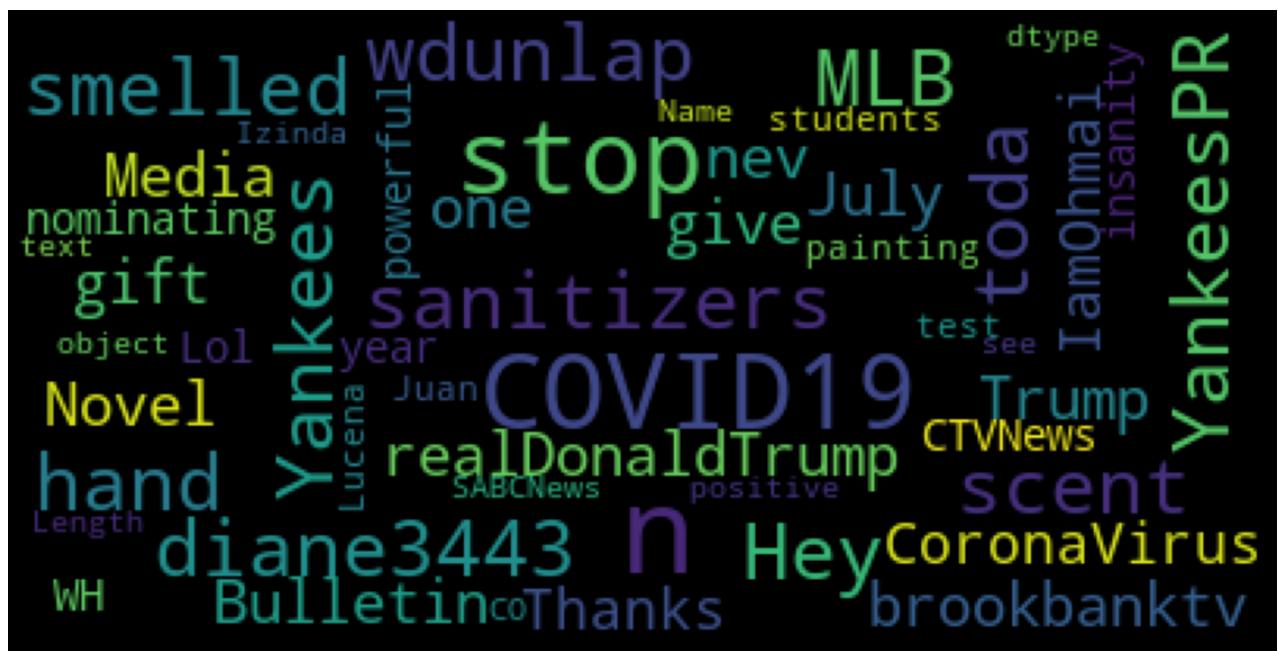
In [7]: `len(data[data['Cluster'] == 0])`

Out[7]: 15187

In [8]: `len(data[data['Cluster'] == 1])`

Out[8]: 163921

In [9]: `word_cloud_func(data[data['Cluster'] == 0]['text'])`In [10]: `word_cloud_func(data[data['Cluster'] == 1]['text'])`



## K-Means Clustering (4 clusters)

```
In [11]: kmeans1 = KMeans(n_clusters = 4, random_state = 0)
kmeans1.fit(features)
```

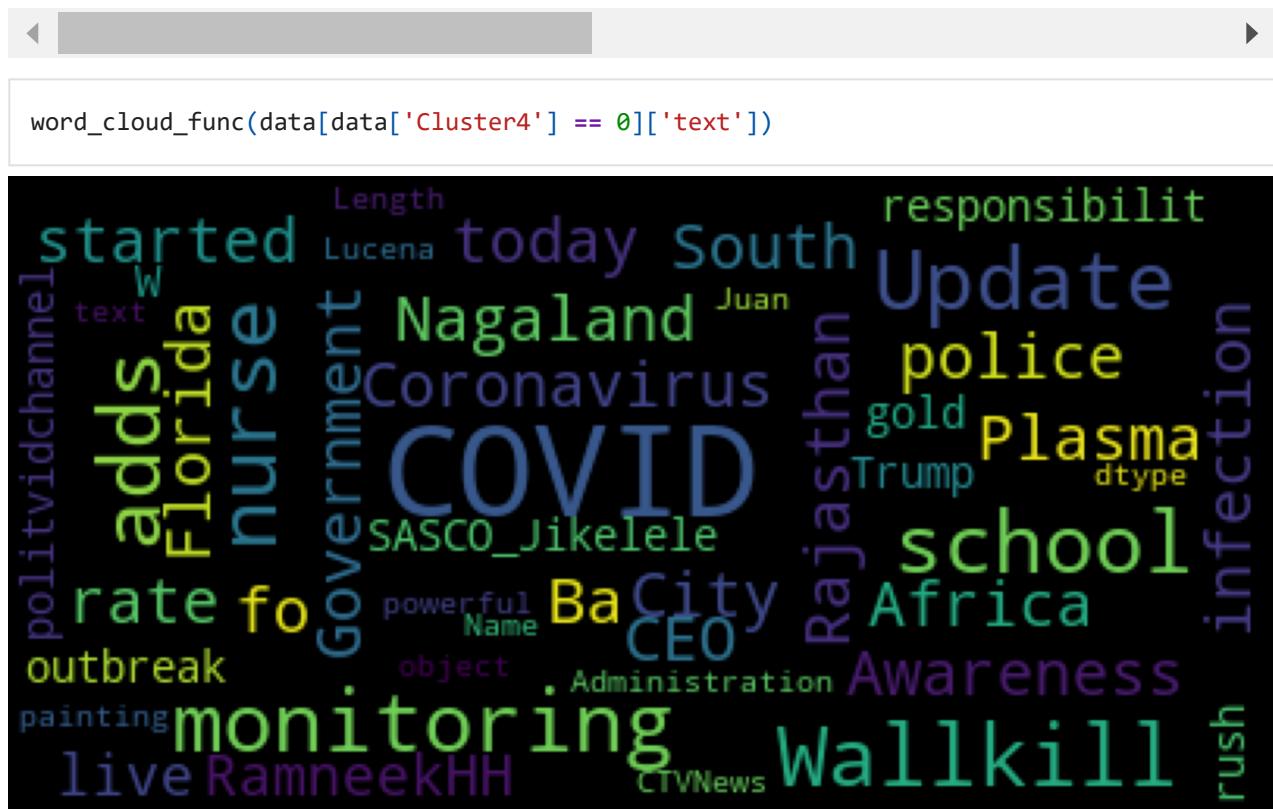
```
Out[11]: KMeans(n_clusters=4, random_state=0)
```

```
In [12]: temp1 = kmeans1.predict(features)
data['Cluster4'] = temp1
data
```

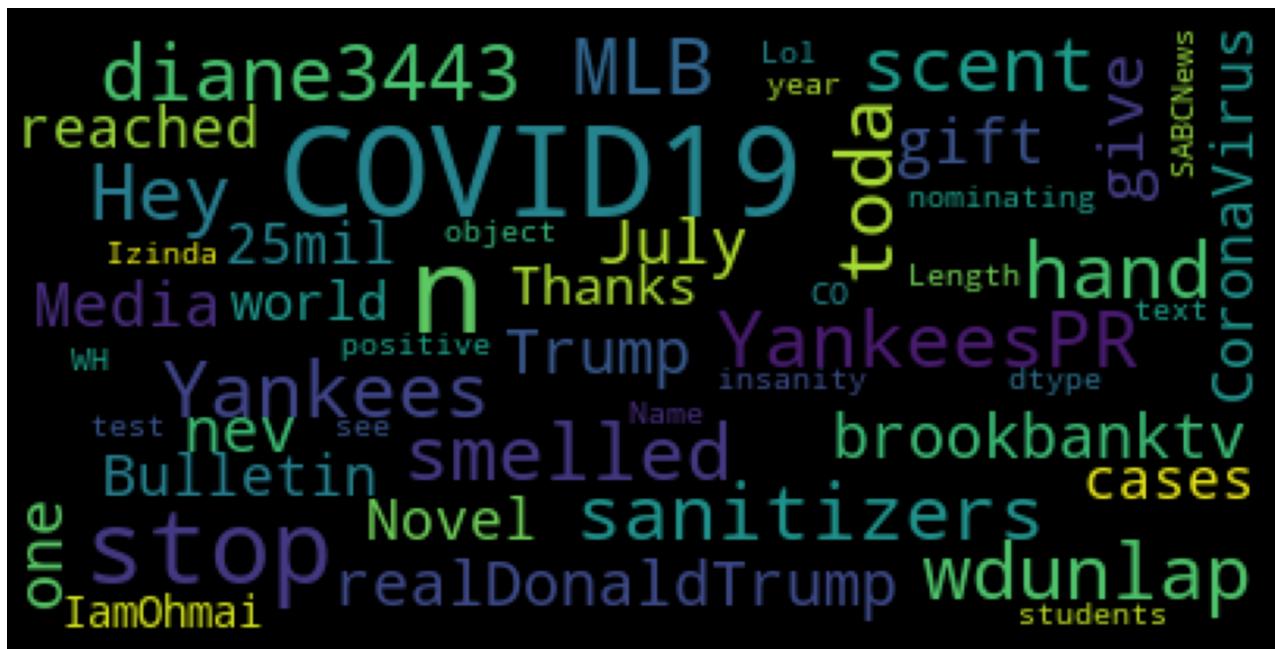
	user_name	user_location	user_description	user_created	user_followers	user_friends	user_f
0	'ViøL€†	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	
1	Tom Basile us	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	
2	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	

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_f
...	...	...	...	...	...	...	...
179103	AJIMATI AbdulRahman O.	Ilorin, Nigeria	Animal Scientist   Muslim   Real Madrid/Chelsea	2013-12-30 18:59:19	412	1609	
179104	Jason	Ontario	When your cat has more baking soda than Ninja	2011-12-21 04:41:30	150	182	
179105	BEEHEMOTH 	ca Canada	The Architects of Free Trade  Really Did ...	2016-07-13 17:21:59	1623	2160	
179106	Gary DelPonte	New York City	Global UX UI Visual Designer. StoryTeller, Mus...	2009-10-27 17:43:13	1338	1111	
179107	TUKY II	Aliwal North, South Africa	TOKELO SEKHOPA   TUKY II   LAST BORN   EISH TU...	2018-04-14 17:30:07	97	1697	

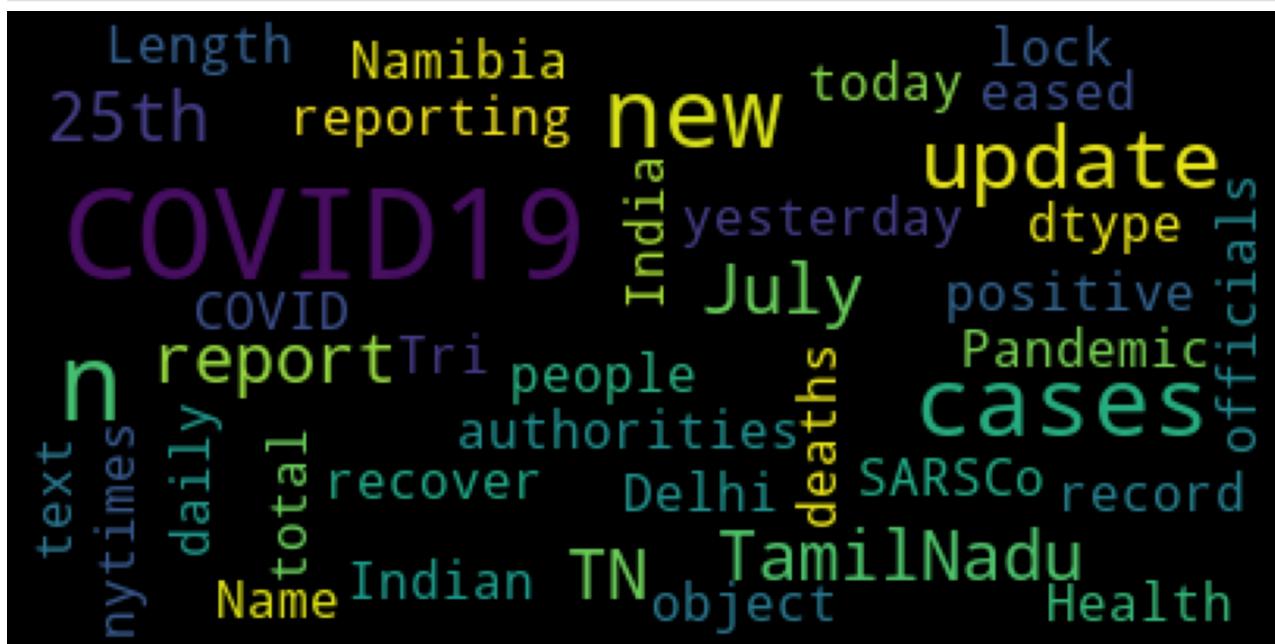
179108 rows × 15 columns



In [14]: word\_cloud\_func(data[data['Cluster4'] == 1]['text'])



```
In [15]: word_cloud_func(data[data['Cluster4'] == 2]['text'])
```



```
In [16]: word_cloud_func(data[data['Cluster4'] == 3]['text'])
```

