WEEKLY PROJECT 16 SANJAY ANAND V

COVID Analysis- Visualisation

1)Importing libraries and datasets

In [20]:
```
import numpy as np
import pandas as pd
import datetime
import matplotlib.pyplot as plt
import seaborn as sns
```

In [21]:
```
df=pd.read_csv('country_vaccinations.csv')
df.head()
```

Out[21]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	0.0	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	NaN	NaN

Adding year column

In [22]:
```
year=[]
for i in df['date']:
    year.append(i.split('-')[0])
df['year']=year
df.head()
```

Out[22]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	0.0	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	NaN	NaN

Adding month column

In [23]:
```
month_name=[]
for i in df['date']:
    date=int(i.split('-')[2])
    month=int(i.split('-')[1])
    year=int(i.split('-')[0])
    month_name.append(datetime.date(year,month,date).strftime('%b'))
df['month_name']=month_name
df.head()
```

Out[23]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	0.0	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	NaN	NaN

Adding date column

In [24]:
```
date_=[]
for i in df['date']:
    date_.append(i.split('-')[2])
df['date_']=date_
df.head()
```

Out[24]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	0.0	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	NaN	NaN

2)Dealing with null values

In [25]:
```
df.isnull().sum()
```

Out[25]:
```
country                              0
iso_code                             0
date                                 0
total_vaccinations                42905
people_vaccinated                 45218
people_fully_vaccinated           47710
daily_vaccinations_raw            51150
daily_vaccinations                  299
total_vaccinations_per_hundred    42905
people_vaccinated_per_hundred     45218
people_fully_vaccinated_per_hundred    47710
daily_vaccinations_per_million      299
vaccines                             0
source_name                          0
source_website                       0
year                                 0
month_name                           0
date_                                0
dtype: int64
```

In [26]:
```
df=df.dropna()
```

In [27]:
```
df.isnull().sum()
```

Out[27]:
```
country                           0
iso_code                          0
date                              0
total_vaccinations                0
people_vaccinated                 0
people_fully_vaccinated           0
daily_vaccinations_raw            0
daily_vaccinations                0
total_vaccinations_per_hundred    0
people_vaccinated_per_hundred     0
people_fully_vaccinated_per_hundred    0
daily_vaccinations_per_million    0
vaccines                          0
source_name                       0
source_website                    0
year                              0
month_name                        0
date_                             0
dtype: int64
```

3)Number of total vaccinations in India in Year 2020,2021,2022

In [28]:
```
total_vaccinations_2020 = df[(df['country']=='India') & (df['year']=='2020')]['total_vaccinations'].sum()
total_vaccinations_2021 = df[(df['country']=='India') & (df['year']=='2021')]['total_vaccinations'].sum()
total_vaccinations_2022 = df[(df['country']=='India') & (df['year']=='2022')]['total_vaccinations'].sum()

print("Total vaccinations in India in 2020: ", total_vaccinations_2020)
print("Total vaccinations in India in 2021: ", total_vaccinations_2021)
print("Total vaccinations in India in 2022: ", total_vaccinations_2022)
```

```
Total vaccinations in India in 2020:  0.0
Total vaccinations in India in 2021:  170720070946.0
Total vaccinations in India in 2022:  149321759019.0
```

4)Number of total vaccinations in Year 2020 of India and USA

In [29]:
```
total_vaccinations_india_2020 = df[(df['country']=='India') & (df['year']=='2020')]['total_vaccinations'].sum()
total_vaccinations_usa_2020 = df[(df['country']=='United States') & (df['year']=='2020')]['total_vaccinations'].sum()

print("Total vaccinations in India in 2020: ", total_vaccinations_india_2020)
print("Total vaccinations in USA in 2020: ", total_vaccinations_usa_2020)
```

```
Total vaccinations in India in 2020:  0.0
Total vaccinations in USA in 2020:  41064128.0
```

5)Number of total vaccinations in Year 2021 of India and China

In [30]:
```
total_vaccinations_india_2021 = df[(df['country']=='India') & (df['year']=='2021')]['total_vaccinations'].sum()
total_vaccinations_china_2021 = df[(df['country']=='China') & (df['year']=='2021')]['total_vaccinations'].sum()

print("Total vaccinations in India in 2021: ", total_vaccinations_india_2021)
print("Total vaccinations in China in 2021: ", total_vaccinations_china_2021)
```

```
Total vaccinations in India in 2021:  170720070946.0
Total vaccinations in China in 2021:  11509664000.0
```

6)Number of vaccinations in each month in India in year 2021

In [32]:
```
n=len(df['country'])
l=['Jan','Feb','Mar','Apr','May','June','July','Aug','Sep','Oct','Nov','Dec']
r=[]
for i in l:
    c = 0
    for j in range(n):
        try:
            if df['month_name'][j] == i and df['year'][j] == '2021':
                c += df['total_vaccinations'][j].sum()
        except KeyError:
            pass
    r.append(c)
```

In [33]:
```
df1=pd.DataFrame(list(zip(l,r)),columns=['MONTH','TOTAL_VACCINATIONS'])
print("The number of vaccinations in each month in India in 2021 are:")
df1.head(20)
```

```
The number of vaccinations in each month in India in 2021 are:
```

Out[33]:

	MONTH	TOTAL_VACCINATIONS
0	Jan	2.092826e+08
1	Feb	9.100231e+08
2	Mar	2.397304e+09
3	Apr	4.756036e+09
4	May	8.256727e+09
5	June	0.000000e+00
6	July	0.000000e+00
7	Aug	2.424152e+10
8	Sep	2.888268e+10
9	Oct	2.898878e+10
10	Nov	3.308481e+10
11	Dec	3.468745e+10

7)The month which has the most number of total vaccinations in India in 2021

In [34]:
```
m=len(df1['MONTH'])
g=max(r)
for h in range(m):
    if(df1['TOTAL_VACCINATIONS'][h]==g):
        print("The month which has the most number of total vaccinations is ",df1['MONTH'][h])
```

```
The month which has the most number of total vaccinations is  Dec
```

In [ ]:
```
```