In [1]:
```
import pandas as pd
```

In [85]:
```
df = pd.read_csv('country_vaccinations.csv')
df.head()
```

Out[85]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vacci
0	Afghanistan	AFG	2021-02-22	0.0	0.0	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	

describing the dataframe and getting the valueable information from it

In [87]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   country                               86512 non-null  object 
 1   iso_code                              86512 non-null  object 
 2   date                                  86512 non-null  object 
 3   total_vaccinations                    43607 non-null  float64
 4   people_vaccinated                     41294 non-null  float64
 5   people_fully_vaccinated               38802 non-null  float64
 6   daily_vaccinations_raw                35362 non-null  float64
 7   daily_vaccinations                    86213 non-null  float64
 8   total_vaccinations_per_hundred        43607 non-null  float64
 9   people_vaccinated_per_hundred         41294 non-null  float64
 10  people_fully_vaccinated_per_hundred   38802 non-null  float64
 11  daily_vaccinations_per_million        86213 non-null  float64
 12  vaccines                              86512 non-null  object 
 13  source_name                           86512 non-null  object 
 14  source_website                        86512 non-null  object 
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
```

In [88]:
```
df.describe()
```

Out[88]:

	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw
count	4.360700e+04	4.129400e+04	3.880200e+04	3.536200e+04
mean	4.592964e+07	1.770508e+07	1.413830e+07	2.705996e+05
std	2.246004e+08	7.078731e+07	5.713920e+07	1.212427e+06
min	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00
25%	5.264100e+05	3.494642e+05	2.439622e+05	4.668000e+03
50%	3.590096e+06	2.187310e+06	1.722140e+06	2.530900e+04
75%	1.701230e+07	9.152520e+06	7.559870e+06	1.234925e+05
max	3.263129e+09	1.275541e+09	1.240777e+09	2.474100e+07

checking the null values

In [90]:
```
df.isnull().sum()
```

Out[90]:
```
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
dtype: int64
```

eliminating the null values by replacing with 0

In [92]:
```
df = df.fillna(0)
df.head()
```

Out[92]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vacci
0	Afghanistan	AFG	2021-02-22	0.0	0.0	
1	Afghanistan	AFG	2021-02-23	0.0	0.0	
2	Afghanistan	AFG	2021-02-24	0.0	0.0	
3	Afghanistan	AFG	2021-02-25	0.0	0.0	
4	Afghanistan	AFG	2021-02-26	0.0	0.0	

In [93]:
```
df.isnull().sum()
```

Out[93]:
```
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
dtype: int64
```

data preprocessing

eliminating the non required column from the dataset

In [95]:
```
df.head()
```

Out[95]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vacci
0	Afghanistan	AFG	2021-02-22	0.0	0.0	
1	Afghanistan	AFG	2021-02-23	0.0	0.0	
2	Afghanistan	AFG	2021-02-24	0.0	0.0	
3	Afghanistan	AFG	2021-02-25	0.0	0.0	
4	Afghanistan	AFG	2021-02-26	0.0	0.0	

In [96]:
```
del df['source_website']
del df['source_name']
del df['vaccines']
del df['daily_vaccinations_per_million']
del df['people_fully_vaccinated_per_hundred']
del df['people_vaccinated_per_hundred']
del df['total_vaccinations_per_hundred']
del df['daily_vaccinations']
del df['people_fully_vaccinated']
del df['iso_code']
del df['daily_vaccinations_raw']
df.head()
```

Out[96]:

	country	date	total_vaccinations	people_vaccinated
0	Afghanistan	2021-02-22	0.0	0.0
1	Afghanistan	2021-02-23	0.0	0.0
2	Afghanistan	2021-02-24	0.0	0.0
3	Afghanistan	2021-02-25	0.0	0.0
4	Afghanistan	2021-02-26	0.0	0.0

preproccessing the date column

In [98]:
```
year =[]
month =[]
Date =[]

for i in df.values:
    year.append(int(i[1].split('-')[0]))
    month.append(int(i[1].split('-')[1]))
    Date.append(int(i[1].split('-')[2]))

df['Date']  = Date
df['month'] = month
df['year']  = year
```

In [99]:
```
#  preproccessing month column
month ={
    1:'Jan',2:'Feb',3:'Mar',4:'Apr',5:'May',6:'Jun',
    7:'Jul',8:'Aug',9:'Sep',10:'Oct',11:'Nov',12:'Dec'
}
month_ = []
for i in df['month']:
    month_.append(month[i])

df['month'] = month_
df.head()

month_ =list(set(month_))
```

In [100]:
```
df.head()
```

Out[100]:

	country	date	total_vaccinations	people_vaccinated	Date	month	year
0	Afghanistan	2021-02-22	0.0	0.0	22	Feb	2021
1	Afghanistan	2021-02-23	0.0	0.0	23	Feb	2021
2	Afghanistan	2021-02-24	0.0	0.0	24	Feb	2021
3	Afghanistan	2021-02-25	0.0	0.0	25	Feb	2021
4	Afghanistan	2021-02-26	0.0	0.0	26	Feb	2021

pre- processing the total_vaccinations column

In [101]:
```
tot_vacc = []
for i in df['total_vaccinations']:
    tot_vacc.append(int(i))

df['total_vaccinations'] = tot_vacc

df.head()
```

Out[101]:

	country	date	total_vaccinations	people_vaccinated	Date	month	year
0	Afghanistan	2021-02-22	0	0.0	22	Feb	2021
1	Afghanistan	2021-02-23	0	0.0	23	Feb	2021
2	Afghanistan	2021-02-24	0	0.0	24	Feb	2021
3	Afghanistan	2021-02-25	0	0.0	25	Feb	2021
4	Afghanistan	2021-02-26	0	0.0	26	Feb	2021

pre- processing the people_vaccinated column

In [102]:
```
ppl_vacc = []
for i in df['people_vaccinated']:
    ppl_vacc.append(int(i))

df['people_vaccinated'] = ppl_vacc

df.head()
```

Out[102]:

	country	date	total_vaccinations	people_vaccinated	Date	month	year
0	Afghanistan	2021-02-22	0	0	22	Feb	2021
1	Afghanistan	2021-02-23	0	0	23	Feb	2021
2	Afghanistan	2021-02-24	0	0	24	Feb	2021
3	Afghanistan	2021-02-25	0	0	25	Feb	2021
4	Afghanistan	2021-02-26	0	0	26	Feb	2021

Q1.Find Monthly number of total vaccination from Jan to Dec in 2021

In [108]:
```
data = []

for month in month_:
    c = 0
    for i in df.values:
        if (i[6] == 2021) and (i[5] == month):
            c += i[2]
    data.append([month,c])

monthly_vaccinated = pd.DataFrame(data,columns =['Month','total_vacci
monthly_vaccinated
```

Out[108]:

	Month	total_vaccinated
0	Jan	1368363247
1	Feb	4511692243
2	Nov	208464393532
3	Sep	165202298502
4	Dec	237472470374
5	May	46939663406
6	Oct	190039710952
7	Apr	26638153919
8	Aug	141091163032
9	Jul	108476702477
10	Jun	75009721988
11	Mar	12370498506