Covid analysis

In [3]:
```
import pandas as pd
```

Loading the dataset

In [33]:
```
df = pd.read_csv('country_vaccinations.csv')
df.head()
```

Out[33]:

	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	

1.Dealing with the null values

In [34]:
```
# learning about the dataset
print('Gaining Information from the dataset        : ' , df.info())
print('\n')
print('Describing the dataset                      : ' , df.describe
print('\n')
print('Length of the dataset                       : ' , len(df))
print('\n')
print('Size of the entire dataset                  : ' , df.size)
print('\n')
print('Shape of the dataset                        : ',df.shape )
print('\n')
```

```
<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
Gaining Information from the dataset        :  None


Describing the dataset                        :         total_vaccin
ations  people_vaccinated  people_fully_vaccinated  \
count        4.360700e+04       4.129400e+04             3.880200e+
04
mean         4.592964e+07       1.770508e+07             1.413830e+
07
std          2.246004e+08       7.078731e+07             5.713920e+
07
min          0.000000e+00       0.000000e+00             1.000000e+
00
25%          5.264100e+05       3.494642e+05             2.439622e+
05
50%          3.590096e+06       2.187310e+06             1.722140e+
06
75%          1.701230e+07       9.152520e+06             7.559870e+
06
max          3.263129e+09       1.275541e+09             1.240777e+
09

       daily_vaccinations_raw  daily_vaccinations  \
count            3.536200e+04        8.621300e+04
mean             2.705996e+05        1.313055e+05
```

```
std              1.212427e+06        7.682388e+05
min              0.000000e+00        0.000000e+00
25%              4.668000e+03        9.000000e+02
50%              2.530900e+04        7.343000e+03
75%              1.234925e+05        4.409800e+04
max              2.474100e+07        2.242429e+07

       total_vaccinations_per_hundred  people_vaccinated_per_hundre
d  \
count                    43607.000000                   41294.00000
0
mean                        80.188543                      40.92731
7
std                         67.913577                      29.29075
9
min                          0.000000                       0.00000
0
25%                         16.050000                      11.37000
0
50%                         67.520000                      41.43500
0
75%                        132.735000                      67.91000
0
max                        345.370000                     124.76000
0

       people_fully_vaccinated_per_hundred  daily_vaccinations_per_
million
count                         38802.000000                    8621
3.000000
mean                             35.523243                     325
7.049157
std                              28.376252                     393
4.312440
min                               0.000000
0.000000
25%                               7.020000                      63
6.000000
50%                              31.750000                     205
0.000000
75%                              62.080000                     468
2.000000
max                             122.370000                   11749
7.000000


Length of the dataset                       :  86512


Size of the entire dataset                  :  1297680


Shape of the dataset                        :  (86512, 15)
```

In [35]:
```
# for total null values in the entire dataset
total_null = df.isnull().sum().sum()
```

In [36]:
```
print(df.isnull().sum())
```

```
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
```

2. Filling the null values with 0

In [37]:
```
df = df.fillna(0)
df
```

Out[37]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_
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	
...	...	...	...	...	...	
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0	
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0	
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0	
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0	
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0	

86512 rows × 15 columns

In [38]:
```
# rechecking if the dataset contains null values or not
df.isnull().sum()
```

Out[38]:
```
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
```

In [39]:
```
df = df.to_csv('F:\\desktop folders\\python files (ETG)\\projects\\We
```