

## Importing required library

```
In [2]: import pandas as pd
```

## Loading the dataset

```
In [13]: df = pd.read_csv('country_vaccinations.csv')
```

```
In [14]: df.head()
```

```
Out[14]:
```

	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	

```
In [15]: # removing the columns
```

```
In [16]: 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[16]:
```

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

## Dealing with the null values

```
In [17]: df.isnull().sum()
```

```
Out[17]: country          0
date          0
total_vaccinations    42905
people_vaccinated    45218
dtype: int64
```

```
In [18]: df = df.fillna(0)
```

```
In [19]: df.head()
```

```
Out[19]:
```

	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

## Add Month Column in the dataset from Date Column (Jan-Dec)

```
In [20]: dct = {1: 'January',
               2: 'February',
               3: 'March',
               4: 'April',
               5: 'May',
               6: 'June',
               7: 'July',
               8: 'August',
               9: 'Septemeber',
               10: 'October',
               11: 'November',
               12: 'December'}
```

```
In [23]: month = []
for i in df['date'].values:
    month.append(i.split('-')[1])

new_month = []
for i in month:
    for j in dct.keys():
        if (int(i) == j):
            new_month.append(dct[j])

df['Month'] = new_month
df.head()
```

```
Out[23]:
```

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