

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

## loading the Dataset

```
In [114]: df = pd.read_csv('country_vaccinations.csv')
```

```
In [103]: df.head()
```

```
Out[103]:
```

	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 [104]: df.isnull().sum()
```

```
Out[104]: 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
```

```
In [115]: df = df.fillna(0)
```

## Also add Year Column in the dataset from Date Column

## Approach 1

```
In [116]: year = []
for i in df['date'].values:
    year.append(i.split('-')[0])
    month.append(i.split('-')[1])
    date.append(i.split('-')[2])
df['Year'] = year
```

```
In [117]: df.head()
```

```
Out[117]:
```

	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	

## Add Month Column in the dataset from Date Column (1-12)

```
In [118]: month = []
for i in df['date'].values:
    year.append(i.split('-')[0])
    month.append(i.split('-')[1])
    date.append(i.split('-')[2])
df['Month'] = month
```

In [119]: `df.head()`

Out[119]:

	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	

## Add Day Column in the dataset from Date Column

In [120]:

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

In [121]: `df.head()`

Out[121]:

	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	

## Approach 2 - Combining Date/month/year by changing the datatype of the entire date column

```
In [70]: # changing the datatype of the data
df['date'] = pd.to_datetime(df['date'])
```

```
In [77]: df['Date'] = df['date'].dt.day
df['Month'] = df['date'].dt.month
df['Year'] = df['date'].dt.year
```

```
In [78]: df.head()
```

```
Out[78]:
```

	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	