

# 1. COVID Analysis

## Import libraries

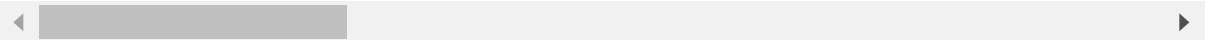
```
In [1]: import pandas as pd  
import numpy as np
```

## Loading dataset

```
In [2]: df = pd.read_csv('Files/country_vaccination_preprocessed.csv')  
df.head()
```

Out[2]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	c
0	Afghanistan	AFG	2021-05-27	593313.0	479574.0	113739.0	
1	Afghanistan	AFG	2021-06-03	630305.0	481800.0	148505.0	
2	Afghanistan	AFG	2022-01-27	5081064.0	4517380.0	3868832.0	
3	Albania	ALB	2021-02-18	3049.0	2438.0	611.0	
4	Albania	ALB	2021-05-11	622507.0	440921.0	181586.0	



In [3]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30847 entries, 0 to 30846
Data columns (total 18 columns):
 #   Column                                  Non-Null Count  Dtype
---  -
 0   country                                30847 non-null  object
 1   iso_code                               30847 non-null  object
 2   date                                   30847 non-null  object
 3   total_vaccinations                    30847 non-null  float64
 4   people_vaccinated                     30847 non-null  float64
 5   people_fully_vaccinated                30847 non-null  float64
 6   daily_vaccinations_raw                 30847 non-null  float64
 7   daily_vaccinations                     30847 non-null  float64
 8   total_vaccinations_per_hundred         30847 non-null  float64
 9   people_vaccinated_per_hundred          30847 non-null  float64
10   people_fully_vaccinated_per_hundred    30847 non-null  float64
11   daily_vaccinations_per_million         30847 non-null  float64
12   vaccines                               30847 non-null  object
13   source_name                            30847 non-null  object
14   source_website                         30847 non-null  object
15   year                                   30847 non-null  int64
16   month                                   30847 non-null  int64
17   day                                    30847 non-null  int64
dtypes: float64(9), int64(3), object(6)
memory usage: 4.2+ MB
```

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

```
In [4]: def month(date):
        lst = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
        day = int( date.split('-')[1] )
        return lst[day - 1]

df['month_name'] = df['date'].apply(month)
```

```
In [5]: df[['country', 'iso_code', 'date', 'year', 'month', 'day', 'month_name']].head()
```

Out[5]:

	country	iso_code	date	year	month	day	month_name
0	Afghanistan	AFG	2021-05-27	2021	5	27	May
1	Afghanistan	AFG	2021-06-03	2021	6	3	Jun
2	Afghanistan	AFG	2022-01-27	2022	1	27	Jan
3	Albania	ALB	2021-02-18	2021	2	18	Feb
4	Albania	ALB	2021-05-11	2021	5	11	May

```
In [6]: df.to_csv('country_vaccination_preprocessed.csv', index=False)
```