

WEEKLY PROJECT 11

COVID Analysis-Data Preprocessing

Importing libraries and dataset

In [11]:

```
import numpy as np
import pandas as pd
import datetime
```

In [12]:

```
df=pd.read_csv('country_vaccinations.csv')
df.head()
```

Out[12]:

| | country | iso_code | date | total_vaccinations | people_vaccinated | people_fully_vaccinated | daily_vaccinations_raw | daily_vaccinations | total_vaccinations_per_hundred | people_vaccinated_per_hundred |
|---|-------------|----------|------------|--------------------|-------------------|-------------------------|------------------------|--------------------|--------------------------------|-------------------------------|
| 0 | Afghanistan | AFG | 2021-02-22 | | 0.0 | 0.0 | NaN | NaN | NaN | 0.0 |
| 1 | Afghanistan | AFG | 2021-02-23 | | NaN | NaN | NaN | NaN | 1367.0 | NaN |
| 2 | Afghanistan | AFG | 2021-02-24 | | NaN | NaN | NaN | NaN | 1367.0 | NaN |
| 3 | Afghanistan | AFG | 2021-02-25 | | NaN | NaN | NaN | NaN | 1367.0 | NaN |

Date Preprocessing

Checking null values

| | |
|----------|---|
| In [13]: | <pre>df.isnull().sum()</pre> |
| Out[13]: | <pre>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</pre> |
| In [14]: | <pre>df=df.dropna()</pre> |

1)Adding year column

In [15]:

```
year=[]
for i in df['date']:
    year.append(i.split('-')[0])
df['year']=year
df.head()
```

Out[15]:

| | country | iso_code | date | total_vaccinations | people_vaccinated | people_fully_vaccinated | daily_vaccinations_raw | daily_vaccinations | total_vaccinations_per_hundred | people_vaccinated_per_hundred |
|-----|-------------|----------|------------|--------------------|-------------------|-------------------------|------------------------|--------------------|--------------------------------|-------------------------------|
| 94 | Afghanistan | AFG | 2021-05-27 | 593313.0 | 479574.0 | 113739.0 | 2859.0 | 6487.0 | 1.49 | 1 |
| 101 | Afghanistan | AFG | 2021-06-03 | 630305.0 | 481800.0 | 148505.0 | 4015.0 | 5285.0 | 1.58 | 1 |
| 339 | Afghanistan | AFG | 2022-01-27 | 5081064.0 | 4517380.0 | 3868832.0 | 6868.0 | 9802.0 | 12.76 | 11 |
| 433 | Albania | ALB | 2021-02-18 | 3049.0 | 2438.0 | 611.0 | 1348.0 | 254.0 | 0.11 | 0 |
| 515 | Albania | ALB | 2021-05-11 | 622507.0 | 440921.0 | 181586.0 | 9548.0 | 12160.0 | 21.67 | 15 |

2)Adding month column

</

3)Adding date column

</

4)Adding day column

</

| | |
|---------|--|
| In []: | |
|---------|--|