WEEKLY PROJECT 11

COVID Analysis-Data Preprocessing

Importing libraries and dataset

```
In [11]: import numpy as np
           import pandas as pd
           \textbf{import} \ \text{datetime}
          df=pd.read_csv('country_vaccinations.csv')
In [12]:
           df.head()
                                    date total_vaccinations people_vaccinated people_fully_vaccinated daily_vaccinations_raw daily_vaccinations total_vaccinations_per_hundred people_vaccinated_per_hundred
Out[12]:
                 country iso_code
                                   2021-
           0 Afghanistan
                              AFG
                                                       0.0
                                                                          0.0
                                                                                                NaN
                                                                                                                      NaN
                                                                                                                                        NaN
                                                                                                                                                                        0.0
                                                                                                                                                                                                      0.0
                                   02-22
                                   2021-
02-23
           1 Afghanistan
                              AFG
                                                      NaN
                                                                         NaN
                                                                                                NaN
                                                                                                                      NaN
                                                                                                                                      1367.0
                                                                                                                                                                       NaN
                                                                                                                                                                                                     Nal
           2 Afghanistan
                                                      NaN
                                                                         NaN
                                                                                                NaN
                                                                                                                      NaN
                                                                                                                                      1367.0
                                                                                                                                                                       NaN
                                                                                                                                                                                                     Nal
                                   2021-
02-25
           3 Afghanistan
                                                      NaN
                                                                         NaN
                                                                                                NaN
                                                                                                                      NaN
                                                                                                                                      1367.0
                                                                                                                                                                       NaN
                                                                                                                                                                                                     Nal
                                                                                                                                       1367.0
           4 Afghanistan
                                                      NaN
                                                                         NaN
                                                                                                NaN
                                                                                                                      NaN
                                                                                                                                                                       NaN
                                                                                                                                                                                                     Nal
```

Date Preprocessing

Checking null values

```
df.isnull().sum()
                                                    0
         country
Out[13]:
                                                    0
         iso_code
         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
                                                    0
         source_name
                                                    0
         source_website
         dtype: int64
        df=df.dropna()
In [14]:
```

1)Adding year column

Out[15]:	C	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hund
	94 Afgh	nanistan	AFG	2021- 05-27	593313.0	479574.0	113739.0	2859.0	6487.0	1.49	1
	101 Afgh	nanistan	AFG	2021- 06-03	630305.0	481800.0	148505.0	4015.0	5285.0	1.58	1
	339 Afgh	nanistan	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	C
	515	Albania	ALB	2021- 05-11	622507.0	440921.0	181586.0	9548.0	12160.0	21.67	15

2)Adding month column

```
In [16]: month=[]
for i in df['date']:
    month.append(i.split('-')[1])
    df['month']=month
    df.head()

Out[16]: country iso_code date total_vaccinations people_vaccinated people_fully_vaccinated daily_vaccinations_raw daily_vaccinations_total_vaccinations_per_hundred people_vaccinated_per_hund
```

TO]:		country	iso_code	uate	total_vaccinations	people_vaccinateu	people_fully_vaccinated	uany_vaccmanons_raw	dany_vaccinations	total_vaccinations_per_nunured	people_vaccinateu_per_nunu
	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	C
	515	Albania	ALB	2021- 05-11	622507.0	440921.0	181586.0	9548.0	12160.0	21.67	15

3)Adding date column

	country	iso_code	aate	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_nundred	people_vaccinated_per_nund
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	C
515	Albania	ALB	2021- 05-11	622507.0	440921.0	181586.0	9548.0	12160.0	21.67	15
	101 339 433	94 Afghanistan101 Afghanistan339 Afghanistan433 Albania	 94 Afghanistan AFG 101 Afghanistan AFG 339 Afghanistan AFG 433 Albania ALB 	94 Afghanistan AFG 2021-05-27 101 Afghanistan AFG 2021-06-03 339 Afghanistan AFG 2022-01-27 433 Albania ALB 2021-02-18	94 Afghanistan AFG 2021- 05-27 593313.0 101 Afghanistan AFG 2021- 06-03 630305.0 339 Afghanistan AFG 2022- 01-27 5081064.0 433 Albania ALB 2021- 02-18 3049.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 433 Albania ALB 2021- 02-18 3049.0 2438.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 2859.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 4015.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 6868.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0	94 Afghanistan AFG 05-27 05-27 593313.0 479574.0 113739.0 2859.0 6487.0 101 Afghanistan AFG 06-03 06-03 630305.0 481800.0 148505.0 4015.0 5285.0 339 Afghanistan AFG 01-27 01-27 5081064.0 4517380.0 3868832.0 6868.0 9802.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0 254.0	101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 4015.0 5285.0 1.58 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 6868.0 9802.0 12.76 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0 254.0 0.11

4)Adding day column

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hund
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	С
515	Albania	ALB	2021- 05-11	622507.0	440921.0	181586.0	9548.0	12160.0	21.67	15
	101 339 433	94 Afghanistan101 Afghanistan339 Afghanistan433 Albania	 94 Afghanistan AFG 101 Afghanistan AFG 339 Afghanistan AFG 433 Albania ALB 	94 Afghanistan AFG 2021-05-27 101 Afghanistan AFG 2021-06-03 339 Afghanistan AFG 2022-01-27 433 Albania ALB 2021-02-18	94 Afghanistan AFG 2021- 05-27 593313.0 101 Afghanistan AFG 2021- 06-03 630305.0 339 Afghanistan AFG 2022- 01-27 5081064.0 433 Albania ALB 2021- 02-18 3049.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 433 Albania ALB 2021- 02-18 3049.0 2438.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 2859.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 4015.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 6868.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 2859.0 6487.0 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 4015.0 5285.0 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 6868.0 9802.0 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0 254.0	94 Afghanistan AFG 2021- 05-27 593313.0 479574.0 113739.0 2859.0 6487.0 1.49 101 Afghanistan AFG 2021- 06-03 630305.0 481800.0 148505.0 4015.0 5285.0 1.58 339 Afghanistan AFG 2022- 01-27 5081064.0 4517380.0 3868832.0 6868.0 9802.0 12.76 433 Albania ALB 2021- 02-18 3049.0 2438.0 611.0 1348.0 254.0 0.11