

WEEK 10 PROJECT

COVID Analysis :

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

In [3]: df = pd.read_csv("country_vaccinations.csv")

In [4]: df

Out[4]:
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22		0.0	0.0	NaN	NaN	NaN	0.00	0.00
1	Afghanistan	AFG	2021-02-23		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
...
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0	3473523.0	139213.0	69579.0	57.59		31.90
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0	3487962.0	100086.0	83429.0	58.25		32.38
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0	3493763.0	53311.0	90629.0	58.61		32.59
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0	3501493.0	89321.0	100614.0	59.20		32.97
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0	3510256.0	105369.0	103751.0	59.90		33.48

86512 rows × 15 columns

```
In [5]: df.head()

Out[5]:
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22		0.0	0.0	NaN	NaN	NaN	0.0	0.0
1	Afghanistan	AFG	2021-02-23		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
2	Afghanistan	AFG	2021-02-24		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
3	Afghanistan	AFG	2021-02-25		NaN	NaN	NaN	NaN	1367.0	NaN	NaN
4	Afghanistan	AFG	2021-02-26		NaN	NaN	NaN	NaN	1367.0	NaN	NaN

1. Dealing with Null values :

1.1) Checking Null Values :

```
In [15]: df.isnull().sum()

Out[15]:
```

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	

1.2) Remove rows where we're having a Null values:

```
In [16]: df.dropna().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_hundred	people_fully_vaccinated_per_hundred
94	Afghanistan	AFG	2021-05-27	593313.0	479574.0	113739.0	2859.0	6487.0	1.49		1.20
101	Afghanistan	AFG	2021-06-03	630305.0	481800.0	148505.0	4015.0	5285.0	1.58		1.21
339	Afghanistan	AFG	2022-01-27	5081064.0	4517380.0	3868832.0	6868.0	9802.0	12.76		11.34
433	Albania	ALB	2021-02-18	3049.0	2438.0	611.0	1348.0	254.0	0.11		0.08
515	Albania	ALB	2021-05-11	622507.0	440921.0	181586.0	9548.0	12160.0	21.67		15.35

2. Filling Null Values with 0 :

```
In [21]: df["country"] = df["country"].fillna('0')
df["iso_code"] = df["iso_code"].fillna('0')
df["date"] = df["date"].fillna('0')
df["total_vaccinations"] = df["total_vaccinations"].fillna('0')
df["people_vaccinated"] = df["people_vaccinated"].fillna('0')
df["people_fully_vaccinated"] = df["people_fully_vaccinated"].fillna('0')
df["daily_vaccinations_raw"] = df["daily_vaccinations_raw"].fillna('0')
df["daily_vaccinations"] = df["daily_vaccinations"].fillna('0')
df["total_vaccinations_per_hundred"] = df["total_vaccinations_per_hundred"].fillna('0')
df["people_vaccinated_per_hundred"] = df["people_vaccinated_per_hundred"].fillna('0')
df["people_fully_vaccinated_per_hundred"] = df["people_fully_vaccinated_per_hundred"].fillna('0')
df["daily_vaccinations_per_million"] = df["daily_vaccinations_per_million"].fillna('0')
df["vaccines"] = df["vaccines"].fillna('0')
df["source_name"] = df["source_name"].fillna('0')
df["source_website"] = df["source_website"].fillna('0')

In [23]: df

Out[23]:
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	0	0	0	0.0		0.0
1	Afghanistan	AFG	2021-02-23	0	0	0	0	1367.0	0		0
2	Afghanistan	AFG	2021-02-24	0	0	0	0	1367.0	0		0
3	Afghanistan	AFG	2021-02-25	0	0	0	0	1367.0	0		0
4	Afghanistan	AFG	2021-02-26	0	0	0	0	1367.0	0		0
...
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0	3473523.0	139213.0	69579.0	57.59		31.9
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0	3487962.0	100086.0	83429.0	58.25		32.38
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0	3493763.0	53311.0	90629.0	58.61		32.59
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0	3501493.0	89321.0	100614.0	59.2		32.97
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0	3510256.0	105369.0	103751.0	59.9		33.48

86512 rows × 15 columns

```
In [24]: df.head()

Out[24]:
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinated_per_hundred
0	Afghanistan	AFG	2021-02-22	0.0	0.0	0	0	0	0.0		0.0
1	Afghanistan	AFG	2021-02-23	0	0	0	0	1367.0	0		0
2	Afghanistan	AFG	2021-02-24	0	0	0	0	1367.0	0		0
3	Afghanistan	AFG	2021-02-25	0	0	0	0	1367.0	0		0
4	Afghanistan	AFG	2021-02-26	0	0	0	0	1367.0	0		0