



# Project Title: Covid-19 Vaccine Analysis

[This Photo](#) by Unknown author is licensed under [CC BY-SA](#).



The screenshot displays a Google Sheets spreadsheet titled "country vaccination". The spreadsheet contains data for various countries, including the United States, France, Germany, Italy, Spain, and the United Kingdom. The data is organized into columns for country, date, total cases, new cases, deaths, and recovery. The data is sorted by date, showing a timeline from early 2020 to mid-2021. The interface includes standard Google Sheets menus and toolbars.

country	date	total	new	deaths	recovery
United States	2020-03-01	0	0	0	0
France	2020-03-01	1367	1367	0	0
Germany	2020-03-01	1367	1367	0	0
Italy	2020-03-01	1367	1367	0	0
Spain	2020-03-01	1367	1367	0	0
United Kingdom	2020-03-01	1367	1367	0	0
United States	2020-03-01	8000	8000	0.02	0.02
France	2020-03-01	1367	1367	0	0
Germany	2020-03-01	1367	1367	0	0
Italy	2020-03-01	1367	1367	0	0
Spain	2020-03-01	1367	1367	0	0
United Kingdom	2020-03-01	1367	1367	0	0
United States	2020-03-01	2018	2018	0	0
France	2020-03-01	2018	2018	0	0
Germany	2020-03-01	2018	2018	0	0
Italy	2020-03-01	2018	2018	0	0
Spain	2020-03-01	2018	2018	0	0
United Kingdom	2020-03-01	2018	2018	0	0
United States	2020-03-01	2862	2862	0.14	0.14
France	2020-03-01	2862	2862	0	0
Germany	2020-03-01	2862	2862	0	0
Italy	2020-03-01	2862	2862	0	0
Spain	2020-03-01	2862	2862	0	0
United Kingdom	2020-03-01	2862	2862	0	0
United States	2020-03-01	54000	54000	0.02	0.02
France	2020-03-01	54000	54000	0	0
Germany	2020-03-01	54000	54000	0	0
Italy	2020-03-01	54000	54000	0	0
Spain	2020-03-01	54000	54000	0	0
United Kingdom	2020-03-01	54000	54000	0	0

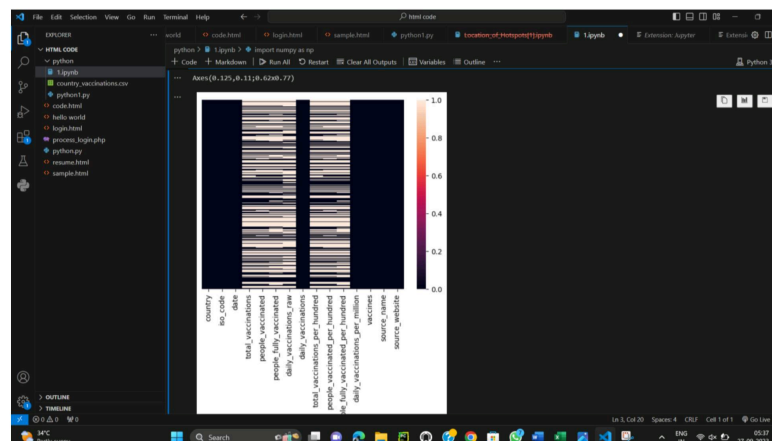
## Data Preprocessing:

- In Case of this part we will be finding the Total number of missing values in the given dataset and Handling it by plotting a Heatmap

```
python > python.py
1 import numpy as np
2 import pandas as pd
3 df=pd.read_csv("VHE_CODE\python\country_vaccinations.csv")
4 print(df.isnull().sum())
5
```

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
80007 ZiaBahan 7x6 ... Ministry of Health https://www.arcgis.com/home/webmap/viewer.html...
80008 ZiaBahan 7x6 ... Ministry of Health https://www.arcgis.com/home/webmap/viewer.html...
80009 ZiaBahan 7x6 ... Ministry of Health https://www.arcgis.com/home/webmap/viewer.html...
80010 ZiaBahan 7x6 ... Ministry of Health https://www.arcgis.com/home/webmap/viewer.html...
80011 ZiaBahan 7x6 ... Ministry of Health https://www.arcgis.com/home/webmap/viewer.html...

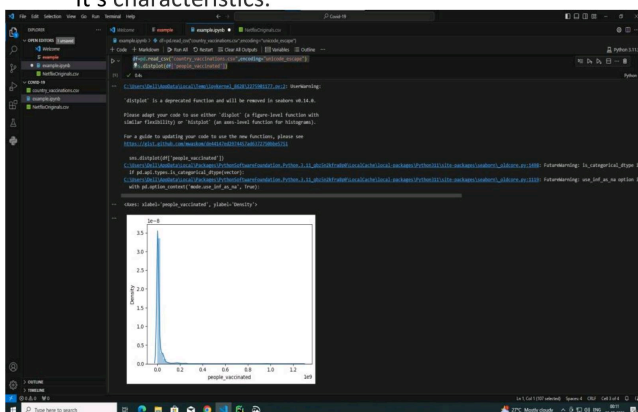
[80012 rows x 15 columns]
PS D:\total code & C:\Users\santh\AppData\Local\Programs\Python\Python310\python.exe "d:\total code\python\python.py"
PS D:\total code & C:\Users\santh\AppData\Local\Programs\Python\Python310\python.exe "d:\total code\python\python.py"
country
iso_code
total_vaccinations 42965
people_vaccinated 42318
people_fully_vaccinated 47718
daily_vaccinations_raw 51156
daily_vaccinations 799
total_vaccinations_per_hundred 42965
people_vaccinated_per_hundred 42318
people_fully_vaccinated_per_hundred 47718
daily_vaccinations_per_million 799
vaccines
source_name
source_website
dhyer: Info4
PS D:\total code &
```



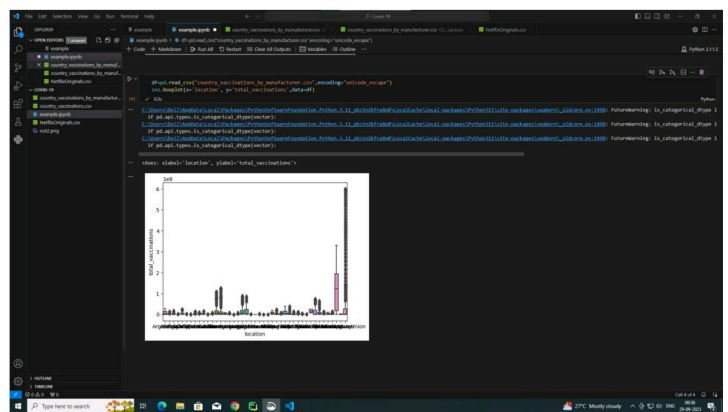
## Exploratory Data Analysis:

In case of Exploratory Data Analysis(EDA) we have gone through the concept of identifying the outliers and plotting them in various type of Graphs.

1. As a very first step of this case we have plotted a distribution chart for the downloaded dataset by understanding it's characteristics.



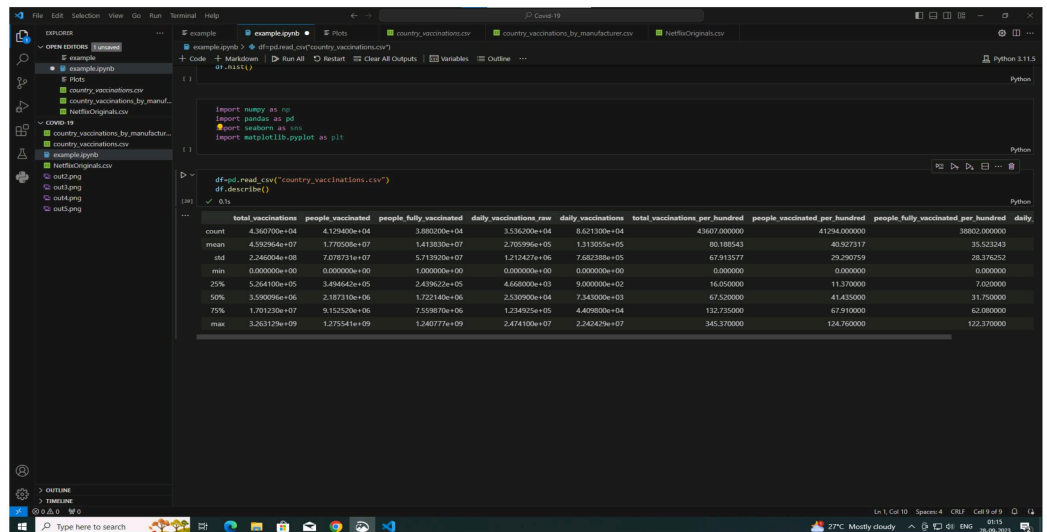
2. Secondly we have plotted the boxplot for given dataset by Understanding about Outliers .



## Statistical Analysis:

- It simply describes the basic Statistics for all continuous variables And Nan values are automatically skipped in these statistics. It Indicates the Count of a variable, Mean, Standard deviation, Minimum and Maximum value

Statistical distribution contains the dataset variables that has either Integers or Floating Point type.

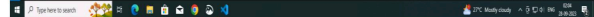
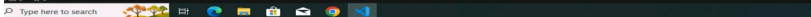


The screenshot shows a Jupyter Notebook interface with a Python script that reads a CSV file and displays its statistical summary. The script is as follows:

```
import numpy as np
import pandas as pd
df = pd.read_csv("country_vaccinations.csv")
df.describe()
```

The output of the `df.describe()` function is a DataFrame containing statistical data for 11 variables. The variables are: `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`, and `daily`. The statistics provided for each variable include count, mean, std, min, 25%, 50%, 75%, and max.

	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	daily
count	4.360700e+04	4.129400e+04	3.880200e+04	3.536200e+04	8.621300e+04	43607.000000	41294.000000	38802.000000	38802.000000
mean	4.592964e+07	1.770506e+07	1.413830e+07	2.705996e+05	1.213055e+05	80.188543	40.927317	35.523343	35.523343
std	2.346504e+08	7.207873e+07	5.713930e+07	1.215427e+06	7.682380e+05	87.918577	29.290759	28.376252	28.376252
min	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	5.264100e+05	3.494642e+05	2.439622e+05	4.668000e+03	9.000000e+02	16.050000	11.370000	7.000000	7.000000
50%	3.590096e+06	2.187310e+06	1.722140e+06	2.530900e+04	7.343000e+03	67.520000	41.435000	31.750000	31.750000
75%	1.701230e+07	9.152520e+06	7.559870e+06	1.234925e+05	4.409800e+04	132.735000	67.910000	62.080000	62.080000
max	3.261129e+09	1.275541e+09	1.240777e+09	2.474100e+07	2.242429e+07	345.370000	124.760000	122.370000	122.370000



## Insights and Recommendations:

---

Variants: Monitoring the impact of COVID-19 variants on vaccine effectiveness is essential. New variants may require booster shots or updated vaccines to maintain protection.

Vaccine Distribution: Ensuring equitable distribution of vaccines globally is critical to achieving widespread immunity. Disparities in vaccine access can prolong the pandemic.

Booster Shots: Research indicated that booster shots might be necessary to maintain immunity, especially for certain populations and in response to new variants.

Long-term Effects: Continuously monitor and research the long-term effects of COVID-19 vaccines, including any potential rare side effects.

Global Collaboration: Collaboration between countries, organizations, and pharmaceutical companies is essential for effective vaccine research, production, and distribution.