

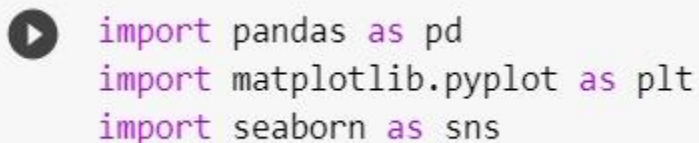
# COVID-19 VACCINES ANALYSIS

## INTRODUCTION:-

In this phase of our Covid-19 vaccines analysis project, we advance by focusing on feature engineering, model training, and evaluation. Feature engineering involves crafting and selecting meaningful attributes from our data to enhance predictive accuracy. Subsequently, we'll train our model, employing suitable machine learning algorithms and parameter optimization. Evaluation then allows us to gauge our model's performance, refining it for improved vaccines forecasts. This progression is essential to empower us with informed investment decisions.

## INITIALIZING THE OBJECT OR CALLING THE LIBRARIES:-

1. "import pandas as pd" allows you to use Pandas for data manipulation, providing data structures like DataFrames and Series to work with tabular data efficiently.
2. "import matplotlib.pyplot as plt" imports Matplotlib, a powerful library for creating various types of data visualizations, with the alias "plt" for convenience.
3. "import seaborn as sns" imports Seaborn, which is built on top of Matplotlib and provides a high-level interface for creating aesthetically pleasing statistical graphics.
4. Together, these libraries are commonly used for data analysis and visualization in Python, enabling you to load, clean, and analyze data and then create informative plots and charts.
5. These libraries are often used in conjunction to explore, analyze, and present data effectively in data science and data visualization.



```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

## PRINTING THE DATAFRAME:-

### 1. Data Retrieval :

The code first specifies the URL of a COVID-19 data file hosted on GitHub. It uses this URL to retrieve the data, which is in CSV format. This data contains information about COVID-19 cases, vaccinations, and related statistics.

### 2. Data Loading:

The Pandas library is used to load the data from the specified URL into a DataFrame, which is stored in the variable `df\_w`. Pandas is a powerful library for data manipulation and analysis and is commonly used for working with tabular data.

### 3. Data Filtering:

After loading the data, the code filters it to include only the records where the 'date' column is less than or equal to '2021-12-08'. This filtering operation restricts the data to a specific time range, in this case, up to and including December 8, 2021. This can be helpful when you're interested in analyzing a particular period of time.

### 4. Information Display:

The code concludes by printing out some basic information about the filtered DataFrame. It displays the number of rows and columns in the DataFrame. This information is useful for understanding the size and structure of the data you're working with.

```
[4] PATH = "https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv"
df_w=pd.read_csv(PATH)
df_w=df_w[df_w.date<='2021-12-08']
```

```
print("The DataFrame has:",
      "\nNumber of rows:",df_w.shape[0],
      "\nand",df_w.shape[1], "number of columns"
)
```

```
The DataFrame has:
Number of rows: 178211
and 67 number of columns
```

## .SUMMARY OF FIRST FIVE DATA:-

Providing a summary of the first five rows of data (often called a "head" of the data) is a common practice in data analysis and data science projects.

### 1. Data Inspection:

Examining the first few rows of the dataset allows you to quickly inspect the data and get a sense of its structure. We can see the names of columns, the data types, and some example values. This initial inspection helps you understand what kind of information is available in the dataset.

### 2. Data Quality:

It helps you assess the quality of the data. If there are any missing values, anomalies, or unexpected values in the first few rows, they can be identified early in the analysis. This is important for data cleaning and preprocessing.

### 3. Column Names:

we can verify that the column names and labels are correct. This is crucial for ensuring that you're working with the right variables.

### 4. Quick Overview:

A summary of the first few rows provides a quick overview of the dataset's contents. We can see the range of values and get an initial impression of the data's distribution.

### 5. Debugging:

If there are issues with data loading or preprocessing, they may become evident when inspecting the first few rows. This can help in debugging the data retrieval and transformation process.

```
df_w.head()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	male_smokers	handwashing_facilities	hospital_
0	AFG	Asia	Afghanistan	2020-01-03	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
1	AFG	Asia	Afghanistan	2020-01-04	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
2	AFG	Asia	Afghanistan	2020-01-05	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
3	AFG	Asia	Afghanistan	2020-01-06	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	
4	AFG	Asia	Afghanistan	2020-01-07	NaN	0.0	NaN	NaN	0.0	NaN	...	NaN	37.746	

5 rows x 67 columns

## SUMMARY OF LAST FIVE DATA:-

Examining the last few rows of the data (often referred to as the "tail" of the data) in a COVID-19 vaccine can be useful for

### 1. Data Completeness:

It helps ensure that the data is complete and up to date. By looking at the last few rows, you can verify that there are no missing data or incomplete records. This is crucial for data integrity.

### 2. Temporal Trends:

If the data is time-series data (as in many COVID-19 datasets), examining the last few rows provides insights into the most recent trends and developments. It allows you to see how the data has evolved over time, which can be important for tracking the progression of the pandemic or vaccination efforts.

### 3. Quality Control:

Checking the last rows can reveal any potential issues or anomalies that might have occurred near the end of the data collection period. This is important for data quality control and anomaly detection

### 4. Data Transformation:

In some cases, data transformations or calculations might have been performed, and these changes may be most evident in the last few rows. By examining the tail of the data, you can check if any post-processing has been applied.

## 5. Decision-Making:

If we are making real-time decisions or updating models based on the most recent data, the last few rows provide the latest information that can inform your decisions.

In the context of a COVID-19 vaccine project, looking at the last few rows helps ensure that your analysis or modeling takes into account the most recent data and events related to the pandemic and vaccination efforts. It's especially important when working with dynamic, evolving datasets that are updated over time.

df\_w.tail()

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	male_smokers	handwashing_facilities	hospi
351151	ZWE	Africa	Zimbabwe	2021-12-04	137441.0	1062.0	515.000	4708.0	1.0	0.571	...	30.7	36.791	
351152	ZWE	Africa	Zimbabwe	2021-12-05	138523.0	1082.0	653.143	4709.0	1.0	0.571	...	30.7	36.791	
351153	ZWE	Africa	Zimbabwe	2021-12-06	139046.0	523.0	722.143	4710.0	1.0	0.714	...	30.7	36.791	
351154	ZWE	Africa	Zimbabwe	2021-12-07	141601.0	2555.0	1053.571	4713.0	3.0	1.000	...	30.7	36.791	
351155	ZWE	Africa	Zimbabwe	2021-12-08	145632.0	4031.0	1568.571	4718.0	5.0	1.571	...	30.7	36.791	

5 rows × 67 columns

## EXTRACTION OF VARIABLES:-

In the context of the COVID-19 vaccine extracting specific variables or columns from the DataFrame (`df_w``) can be beneficial for

### 1. Focus on Relevant Data:

By extracting a subset of columns, we can focus on the specific variables that are most relevant to your analysis or research question. This can help we reduce clutter and simplify your analysis.

### 2. Computational Efficiency:

Working with a smaller set of columns can improve computational efficiency. Unnecessary columns can consume memory and processing power, so extracting only the variables of interest can speed up operations.

### 3. Simplifying Analysis:

Simplifying the dataset by selecting specific variables makes it easier to perform statistical analysis, data visualization, and modeling. We can more easily see patterns and relationships in the data.

### 4. Data Privacy and Security:

If the dataset contains sensitive information, extracting only the necessary variables can help protect privacy and security. Unneeded data can be a risk if not properly handled.

### 5. Clarity and Readability:

Extracting variables helps improve the clarity and readability of your code. It's easier for others to understand and work with a more focused subset of data.

## 6. Custom Data Processing:

We can apply custom data processing and transformations to the selected variables without affecting the entire dataset. This is particularly useful if you need to perform specialized calculations or modifications.

## 7. Modeling and Prediction:

If we are building models or making predictions, it's important to choose the right set of input features. Selecting relevant variables is crucial for model accuracy and interpretability.

In specific case, it appears that you've listed a set of COVID-19-related variables that we want to extract from the Data Frame. These variables might be of particular interest for your analysis or research related to the pandemic, vaccination efforts, or other specific aspects of the data. Extracting them will allow us to work with this subset of data more effectively.



```
df_w.columns.to_list()
```



```
'new_deaths_smoothed',  
'total_cases_per_million',  
'new_cases_per_million',  
'new_cases_smoothed_per_million',  
'total_deaths_per_million',  
'new_deaths_per_million',  
'new_deaths_smoothed_per_million',  
'reproduction_rate',  
'icu_patients',  
'icu_patients_per_million',  
'hosp_patients',  
'hosp_patients_per_million',  
'weekly_icu_admissions',  
'weekly_icu_admissions_per_million',  
'weekly_hosp_admissions',  
'weekly_hosp_admissions_per_million',  
'total_tests',  
'new_tests',  
'total_tests_per_thousand',  
'new_tests_per_thousand',  
'new_tests_smoothed',  
'new_tests_smoothed_per_thousand',  
'positive_rate',  
'tests_per_case',  
'tests_units',  
'total_vaccinations',  
'people_vaccinated',  
'people_fully_vaccinated',  
'total_boosters',
```

## **ANALYSIS:-**

Analyzing a COVID-19 vaccine project using graphs and models is important for

### **1. Visualizing Trends:**

Graphs and charts can visually represent data trends, making it easier to spot patterns and changes over time. Visualizations can reveal vaccination rates, infection rates, and other important trends related to the pandemic.

### **2. Identifying Correlations:**

Statistical models and correlation analyses can help identify relationships between variables. For instance, you can assess whether vaccination rates have an impact on the reduction of infection rates.

### **3. Predictive Modeling:**

Statistical and machine learning models can be used to make predictions and forecasts. These models can help estimate future COVID-19 trends, vaccination needs, or the impact of policy decisions.

### **4. Public Health Decision-Making:**

Data analysis and models can inform public health decision-making. Government agencies, healthcare organizations, and policymakers can use the insights to plan vaccination campaigns, allocate resources, and make informed decisions.

### **5. Monitoring and Surveillance:**

Real-time data analysis and models can aid in monitoring the spread of the virus, detecting outbreaks, and identifying areas with vaccination coverage gaps. This information is crucial for targeted interventions.

### **6. Resource Allocation:**

Models can assist in allocating healthcare resources efficiently, such as hospital beds, ventilators, and medical personnel. This helps ensure that healthcare systems are prepared for surges in COVID-19 cases.

### **7. Vaccination Strategy Optimization:**

Analysis and models can help optimize vaccination strategies, including prioritizing vaccination for high-risk populations, determining the timing of booster shots, and managing vaccine supply chains.

### **8. Communication:**

Visualizations and models provide a means to communicate complex data and insights to the public, healthcare workers, and policymakers. This aids in raising awareness and promoting responsible behavior.

### **9. Research and Development:**

Data analysis and modeling can support ongoing research into COVID-19 vaccines, treatments, and the virus itself. It can guide researchers in refining vaccines and treatments.

### **10. Data-Driven Policy:**

Governments and health organizations can use data-driven analysis to develop policies and guidelines. This is especially important in managing the pandemic effectively.

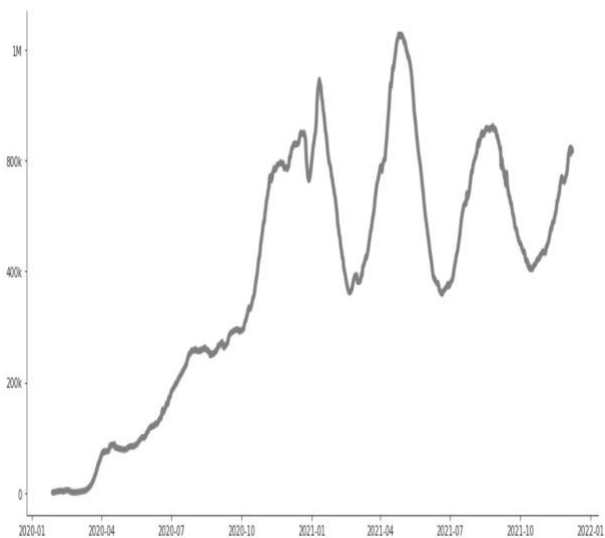
Data analysis, graphs, and models play a critical role in understanding, managing, and responding to the COVID-19 pandemic, particularly in the context of vaccine distribution and the broader public health response. They provide evidence-based insights and support informed decision-making.

```
[ ] # Calling the libraries
import matplotlib.pyplot as plt
from pywaffle import Waffle
import numpy as np
```

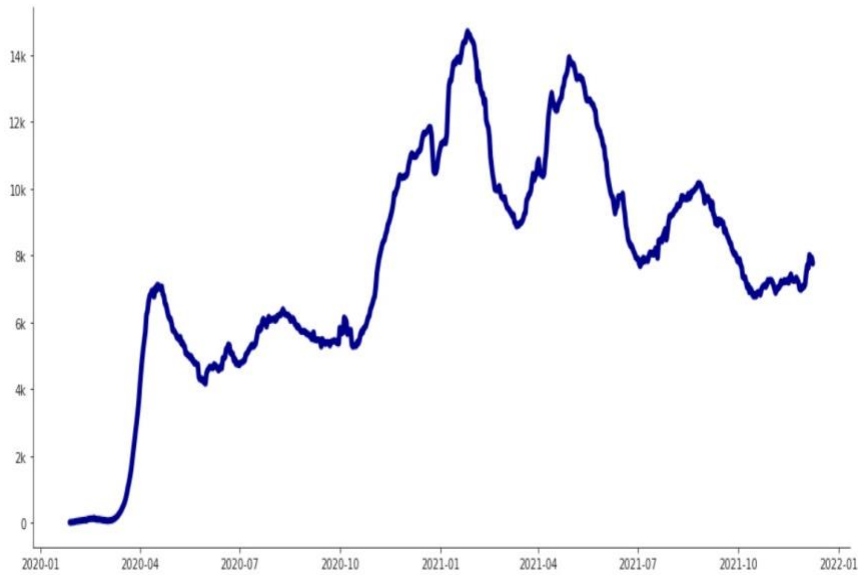
```
fig = plt.figure(
    FigureClass=Waffle,
    rows=5,
    figsize = (16,4),
    values={'Fully vaccinated': 44, 'Only partly vaccinated': 11, 'Did not take any dose': 45},
    title={
        'label': 'Number of people vaccinated against COVID-19\n',
        'loc': 'left',
        'fontdict': {
            'fontsize': 20
        }
    },
    colors=("OliveDrab", "DarkBlue", "DarkRed"),
    icons = ['syringe', 'syringe', 'skull'],
    legend={'loc': 'upper left',
            'bbox_to_anchor': (1, 1),
            'fontsize': 20},
    icon_size=24,
    icon_legend=True
)

fig.set_tight_layout(False)
plt.show();
```

Moving Average Smoothed of the number of New Cases of COVID-19 in the World



Moving Average Smoothed of the number of New Deaths of COVID-19 in the World



## CONCLUSION:

We notice the how fast is the way the virus Covid-19 can spread into the world. The world didn't know how to deal with the problem at the start because of the number of cases and deaths. The countries with more income were able to get resources and track the virus more efficient than others. The vaccines are the way to combat the virus. Only 19% of the population of the World are fully vaccinated Vaccination is the best way to deal with the virus.