**COVID VACCINES ANALYSIS**

**TEAM MEMBERS**

**310821243012: DHANUUSH A**

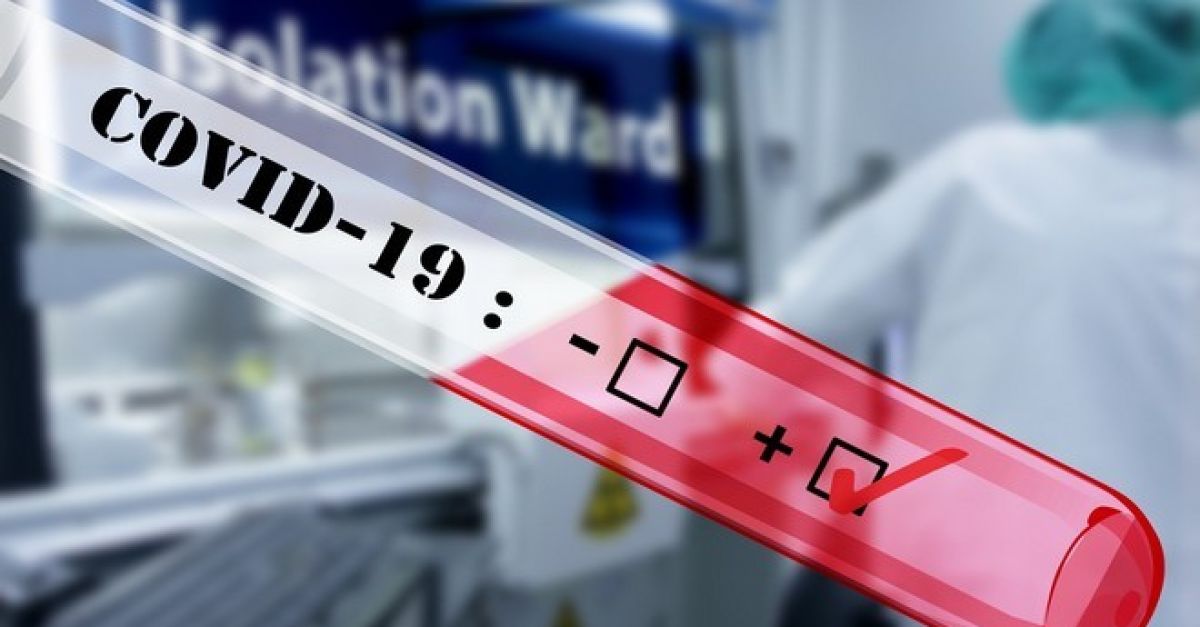
**310821243048: SAI NITHISH KUMAR S**

**310821243034: MURALI KRISHNAN M**

**310821243033: LOKESH V**

**PHASE-II:** Document submission

**PROJECT:** *DATA ANALYSIS ON COVID VACCINATION DATA*



**AIM:**

A COVID-19 analysis report is a comprehensive document that presents data-driven insights and conclusions about the COVID-19 pandemic. It typically covers various aspects, such as the spread of the virus, its impact on public health, economic repercussions, and more. Here's a general structure of such a report:

**Introduction:**

The COVID 19 pandemic caused due to the Corona virus devastated the world by causing several fatalities around the world. This virus originated in Wuhan, China in 2019 and was later spread throughout the world due to human contact in one way or the other. An effort was made to find a cure or vaccine by several health organizations to bring a stop to this pandemic.

In later stages of 2020 several experimental vaccines were developed and was administered to humans. The efforts were successful as the vaccines were helpful in reducing the affects the virus and even if people were infected, they were not in any life threating situation and escaped the illness having only minor symptoms. Many countries later developed their own vaccines and also helped other countries without the resources by providing them with vaccines developed.

**Tools Used:**

Here's an overview of some popular tools for different stages of a data science project:

**1.Programming languages**:

**Python**: Dominant in data science due to its simplicity and a vast ecosystem of libraries.

**2.Integrated Development Environments (IDEs) and Notebooks:**

**Jupyter Notebook/Jupyter Lab:** An open-source tool that lets you create and share documents with live code, equations, visualizations, and narrative text.

**3.Data Cleaning & Transformation:**

**Pandas:** A Python library for data manipulation and analysis.

**4.Visualization:**

**Matplotlib & Seaborn:** Widely used Python libraries for static, animated, and interactive visualizations.

**Plotly:**is a library for creating interactive visualizations in Python.

**5.Machine Learning Frameworks**:

**Scikit-learn:** A Python-based machine learning library for various algorithms.

**TensorFlow:** Open-source libraries for deep learning.

# COVID-19 Vaccine Analysis

For the purpose of analysing COVID-19 vaccines, I will be working with a dataset sourced from Kaggle.

The steps required for Data Analysis:

* Importing all the relevant packages
* Loading the Data
* Transforming data (data cleaning /data wrangling)
* Exploratory data analysis and visualization

## **Step 1 — Importing necessary libraries:**

Before we start analysing data, we need to import the necessary libraries. Some common libraries I used in this analysis include NumPy, Pandas, Matplotlib, Seaborn and Plotly.

**NumPy** is a library for numerical computing in Python. It provides powerful tools for working with arrays and matrices, which are commonly used in data analysis.

**Pandas**is a library for data manipulation and analysis in Python. With Pandas, data is loaded, cleaned, transformed and analysed quickly and easily. It also provides convenient methods for grouping data, filtering data, and handling missing data.

**Matplotlib and Seaborn**are libraries for data visualization in Python. Matplotlib is a general-purpose plotting library that allows you to create a wide range of visualizations, including line plots, scatter plots, bar plots, histograms, and more. Seaborn is a higher-level library that builds on top of Matplotlib, providing additional functionality for statistical data visualization. It includes tools for creating heatmaps, pair plots, and other types of advanced visualizations.

**Plotly** is a library for creating interactive visualizations in Python. It provides tools for creating a wide range of charts and graphs, including line charts, scatter plots, bar charts, and 3D surface plots.

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import plotly.express as px

## **Step 2 — Loading the data**

Once the libraries are imported, I need to load the data (CSV file) into out Python environment. So I will be using pandas for this data = pd.read\_csv(‘www.kaggle.com/covid-world-vaccination-progress’) data.head()

## **Step 3 — Data Wrangling**

Data wrangling, also known as data munging, is the process of cleaning, transforming, and preparing raw data for analysis. It is an important step in the data analysis process as it ensures that the data is accurate, consistent, and ready for analysis. Some common tasks involved in data wrangling:

D**ata cleaning**involves identifying and handling missing data, removing duplicates, and correcting any errors in the data.

**Data transformation**involves converting the data into a format that is suitable for analysis. This may involve scaling the data or converting categorical variables to numerical variables.

**Data reduction**involves reducing the size of the dataset by removing unnecessary variables or observations.

After loading the data and after analysing the data | understood that there are 86512 rows and 15 columns. And in that some of the columns contained null values I have replaced the null values by 0 with the use of replace functions and started working on the data.

data = data.drop(['source\_name', 'source\_website'], axis=1)  
data.isnull().sum()  
data=data.fillna(0) data['date'] = pd.to\_datetime(data.date)

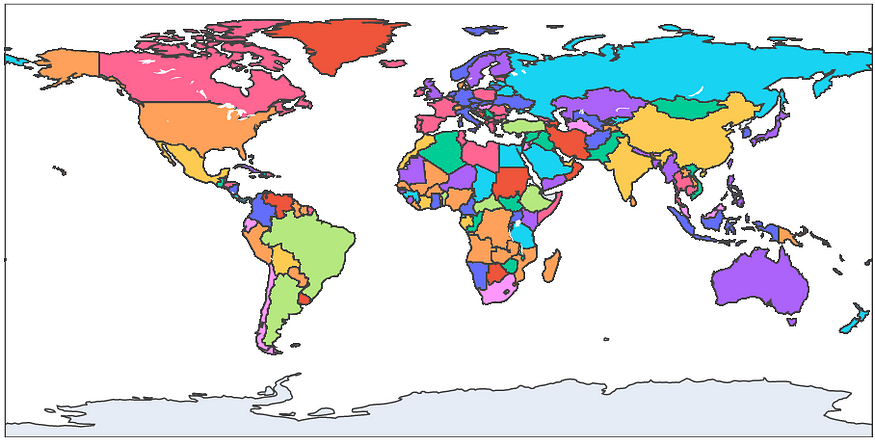
**Step 4 —** **Exploratory Data Analysis and Visualization:**

- **Charts and Graphs:** I will be using various types of charts, such as line graphs for time-series data, bar charts to compare regions, and heatmaps to show the geographic distribution of cases.

- **Geospatial Visualization:** I Use Geographic Information System (GIS) tools like ArcGIS or open-source options like QGIS to create maps that display the spatial distribution of cases.

- **Tools:** I use Python libraries like Matplotlib, Seaborn, and Plotly for data visualization. For geospatial analysis and mapping, I use ArcGIS, QGIS, or Python libraries like Folium and Geopandas.

import plotly.express as px  
import plotly.offline as py  
import plotly.io as pio  
vaccine\_map = px.choropleth(data, locations = 'iso\_code', color = 'vaccines')  
vaccine\_map.update\_layout(height=300, margin={"r":0,"t":0,"l":0,"b":0})



**Insights:**

Here I analysed the top 10 fully vaccinated countries in which India tops the list which indicates that people in the country where showing lots of interests to get vaccinated.

And also I analysed top 5 vaccinated countries here also India tops the list. And then analysed top 5 daily vaccinating countries and here China tops the list.

And also I analyse the sum of daily vaccinating details, fully vaccinating and vaccinating people details.

And our year wise analyse shows that 2021 was the peak year for every vaccination details

**Conclusions:**

In this dataset I came to know that the vaccination process in every country was gone in good pace that indicates we have controlled this disease very soon all over the world.

Based on the analysis of Covid-19 vaccine distribution across different countries, it can be concluded that there is a significant variation in the types of vaccines being used. The Pfizer-BioNTech vaccine is the most commonly used vaccine across the world, followed by the AstraZeneca vaccine. However, some countries have opted for other vaccines such as Sinovac, Sinopharm, and Sputnik V.