

COVID VACCINE ANALYSIS

2.1 INTRODUCTION:

The objective of the Covid Vaccine Analysis is to comprehensively assess the global progress of COVID-19 vaccinations. This analysis takes into account crucial factors such as the total number of vaccinations, the count of individuals fully vaccinated, the types of vaccines available in each country, and the respective sources along with their websites.

2.2 DATASET :

Utilizing the dataset available at Kaggle - Covid World Vaccination Progress, we aim to derive insights in the following ways:

- (a) Rank countries based on the total number of people fully vaccinated.
- (b) Rank countries based on the number of people fully vaccinated per hundred.
- (c) Compile a list of vaccines used globally.
- (d) Determine the number of each vaccine type used in the European Union.

Data Preprocessing:

Rigorous data cleaning and preprocessing are essential to ensure data accuracy. Steps include handling missing data, standardizing formats, and addressing outliers.

2.3 COLUMNS:

The data (country vaccinations) contains the following information:

- **Country**- this is the country for which the vaccination information is provided;
- **Country ISO Code** - ISO code for the country;
- **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations** - this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;
- **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;
- **Daily vaccinations (raw)** - for a certain data entry, the number of vaccination for that date/country;

- **Daily vaccinations** - for a certain data entry, the number of vaccination for that date/country;
 - **Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country;
 - **Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country;
 - **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;
 - **Number of vaccinations per day** - number of daily vaccination for that day and country;
 - **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;
 - **Vaccines used in the country** - total number of vaccines used in the country (up to date);
 - **Source name** - source of the information (national authority, international organization, local organization etc.);
 - **Source website** - website of the source of information;
- There is a second file added recently (country vaccinations by manufacturer), with the following columns:
- **Location** - country;
 - **Date** - date;
 - **Vaccine** - vaccine type;
 - **Total number of vaccinations** - total number of vaccinations / current time and vaccine type.

2.4 LIBRARIES USED:

Numpy: Numpy provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.

Installation: *pip install numpy*

Pandas: Pandas is a data manipulation and analysis library. It provides data structures like Series and DataFrame, which are designed for efficient and intuitive data manipulation.

Installation: *pip install pandas*

Matplotlib: Matplotlib is a 2D plotting library for Python. It produces high-quality graphs, charts, and figures, making it suitable for data visualization tasks.

Installation: *pip install matplotlib*

Seaborn: Seaborn is built on top of Matplotlib and provides a high-level interface for drawing attractive statistical graphics. It comes with several built-in themes and color palettes to make data visualization more appealing.

Installation: *pip install seaborn*

After installation, you can **import** these libraries in your Python scripts or Jupyter Notebooks using “**import**” statements:

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

2.6 EXPLORATORY DATA ANALYSIS:

Data Exploration:

Exploring the dataset and finding various insights about the vaccines, total number of people vaccinated and vaccine types.

Feature Engineering:

Create new features, e.g., vaccine coverage rates, vaccine efficacy scores, and demographic variables, to enhance the analysis's depth.

Visualizations:

Using visualizations such as line charts, heatmaps, and geographic maps to uncover trends, correlations, and disparities in vaccine distribution and effectiveness.

Machine Learning Algorithms:

Consider exploring advanced machine learning techniques like clustering or time series forecasting to uncover hidden patterns in vaccine distribution and adverse effects data.

CONCLUSION:

The Covid Vaccine Analysis offers a comprehensive exploration of global vaccination data, presenting insights through meaningful visualizations and graphs.