## **COVID VACCINES ANALYSIS**

**PHASE-2:** Document submission

**PROJECT:** *COVID VACCINES ANALYSIS*

**Problem Definition:**

The problem is to conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. This project involves data collection, data pre-processing, exploratory data analysis, statistical analysis, and visualization.

**Design Thinking:**

This project proposes to conduct an in-depth analysis of COVID-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. The project will involve the following steps:

**1. Data Collection:**

The project will collect COVID-19 vaccine data from various sources, including government agencies, public health organizations, and research institutions. The data will include information on vaccine efficacy, distribution, and adverse effects.

**Dataset link:** <https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

**2. Data Cleaning:**

Check data for inconsistencies, missing values, and outliers. Validate that the data is in a usable format.

**Tools:** Using Python libraries like pandas for cleaning and removing inconsistencies in the data.

**3. Data Analysis:**

Clustering algorithms like K-Means and Hierarchical clustering have been employed in order to reveal patterns in the data. Comparing the accuracy score of the algorithms , we use the algorithm with more accuracy.

**Exploratory data analysis**: Exploratory data analysis will be performed to identify patterns and trends in the data. This will help to generate hypotheses about the relationships between vaccine efficacy, distribution, and adverse effects.

**Descriptive Analysis:** Calculating basic statistics, such as mean, median, and standard deviation, to understand the central tendencies and variabilities in the data.

**Hypothesis Testing:** Using statistical tests to compare different variables, for instance, to determine if there are significant differences in infection rates between regions.

**Time-Series Analysis:** Analysing the data over time to identify trends and patterns, such as seasonal variations or the impact of public health measures. We use ARIMA models to capture time-dependant patterns and predict future adverse effects.

**Tools:** Using Python libraries like NumPy and SciPy for statistical analysis and modelling.

**4. Data Visualization:**

Creating 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.

**Time series plot:** We create time series line plots for each country to visualize the daily cases and decide about deployment of vaccines.

**Bar Charts**: We generate bar charts to compare the mean values of daily cases and vaccine production for different countries.

**Error Bars**: We use error bar charts to visualize standard deviations, showing the variability around the mean for cases and death.

**Tools:** Using Python libraries like Matplotlib, Seaborn, and Plotly for data visualization.

**5. Report Compilation:**

We use document preparation tools like Microsoft Word, Google Docs to compile the report and present it to the stake holders.