Problem : Covid-19 Vaccines Analysis

Analyzing COVID-19 vaccine data reveals vaccination trends, distribution patterns, and the impact of immunization campaigns, offering insights crucial for public health strategies and global pandemic management.

Dataset Origin

The dataset was obtained from Kaggle, a popular platform for datasets and data science competitions. Kaggle serves as a valuable resource for diverse datasets that cater to various domains.

Data Preprocessing:

Before applying the process I verified there are no null values in the dataset and I grouped data by country(India) for specific plots.

Libraries Used:

I utilized key libraries like Pandas for data manipulation, sci-kit-learn for machine learning tasks, and Seaborn/matplotlib for data visualization.

Selected Columns:

To perform clustering and time series forecasting, I focused on numerical features such as '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_vaccinations_per_million.'

Visualization:

For visualizing the clusters and time series forecasting, I created a scatter plot using Seaborn and Barchart-like charts.

Clustering and Evaluation

*   I used the k-means clustering algorithm, a popular choice for its simplicity and efficiency. The number of clusters, a crucial parameter in k-means, was set to 3 based on experimentation.

*   The silhouette score measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation). Higher silhouette scores indicate better-defined clusters.