BIG DATA PIPELINE FOR COVID-19 ANALYSIS

Aim of the Project

The goal of this project will be to investigate the patterns of the pandemic taken from two sources: the count of COVID-19 cases and the count of vaccinations in the selected countries in order to determine the relationship between the vaccination effort and subsequent changes in case numbers. The project leverages data from two sources: daily new cases as well as global vaccination trends. By integrating and visualizing these datasets, we aim to:

1. Give an analysis of the evolution of the pandemic over time and at global and country levels.

2. Consider the measures and steps governments have taken through mass vaccination.

3. To diagnose any unusual behavior in case development rates and vaccinations that can help in shaping more effective public health interventions.

Specific Objectives:

Often known as the ‘coronavirus data,’ processes information on daily new cases and deaths up to the country level and date.

Integrate case data with vaccination status to make it easier to evaluate the premises for pandemic progression.

Derivable, and therefore changeable, metrics include Daily New Cases for short-term fluctuations in the rate of infection.

Make major trends easy to visualize with specific plot types so as to allow those who are charged with decision-making to easily analyze the results.

Dynamic and scalability are the key themes of this project: the additional datasets (such as hospitalization and testing rates) were not included because the current work can be easily extended to include them in the future.

This project emphasizes a modular and extensible approach to data processing, ensuring that it can incorporate additional datasets (e.g., hospitalizations, testing rates) for more nuanced analyses in the future.

Results of the Project

The analysis and visualization conducted in this project provided the following key findings:

1. Daily Case Trends:

The graph presenting new cases every day helped to define trends in COVID-19 increasing day after day during some days or weeks. For instance, there were observable fluctuations of infections such as the ones experienced in Argentina country.

2. Vaccination Impact:

If the total immunizations curve was overlaid on the new cases curve, the analysis suggested that there was a combination of greater overall vaccination coverage leading to lower new cases over roughly similar timescales. For instance, in several provinces that administered vaccination exercises more aggressively, the progression of fresh infections began slowing down once the number of vaccinations met a certain level.

3. Country-Specific Insights:

Vaccination data was also collected and supplemented with case statistics to enable a detailed analysis of the pandemic situation in the countries. This brought out some disparities in the manner in which diverse countries handled the pandemic, and this was in regard to local policies, healthcare systems, and vaccines available.

4. Visualization Effectiveness:

Vaccination and infection rate trends were well depicted in the produced line charts. These visualizations proved especially helpful in cases where one could overlay the timelines of the vaccinations with the timelines of the infection rates.