BIG DATA PIPELINE FOR COVID-19 ANALYSIS

Aim: - The analysis will involve observing the trends of the pandemic counting two sources the number of COVID-19 cases and the number of vaccinations, from the countries of one's interest, in order to establish the relationship between the efforts made in vaccinating and subsequent variations in the case numbers. The work will use data from the two sources: new cases added every day along with global trends in vaccination. An integration of these different sources should help us achieve the following objectives:

1. Give an analysis about the evolvement of the pandemic with time at global and country levels.

2. Consider measures and steps governments undertook through mass vaccination.

3. Diagnose any atypical behaviour in the case development rates and vaccinations, which could help in shaping better public health interventions.

Specific Objectives:

Commonly termed as `coronavirus data`, it registers information regarding daily new cases and death tally up to the national level and date.

Integrate case data with vaccinal status such that it facilitates the easier evaluation of assumptions regarding distance to pandemic progression.

Daily New Cases derive, and are thus modifiable, metrics for short-term fluctuations in the infection rate.

Visualize broad trend data in very specific plot types to facilitate easy visualization of inputs by those with decision-making responsibilities.

They usually cover short data sets that report on new cases on a daily basis, which include deaths until it involves the country level and the date.

So that case data can be integrated with vaccination status to facilitate easier evaluation of assumptions of distance to pandemic progression.

Derivable, and therefore changeable, are measures such as Daily New Cases for short-term fluctuations in the infection rate. Specific plot types should be used to simplify visualization of broad trend data for easy inputs from people with decision-making responsibilities.

This project is dynamic and scalable by nature: additional datasets, such as hospitalization and testing rates, have not been included for the time being because it has been designed to be easily extended for adding future datasets to it. In its design this project advocates a modular and extensible approach, which leaves it open to integrating new sets of data- hospitalization, testing rates-for deeper analyses in the future.

Result: -

Findings from the data analysis and visualization in this project have included:

1. Daily Case Trend:

The daily new-case graph helps determine trends in COVID-19 increases on certain days or weeks. An example is the fluctuations of infections like those observed in Argentina.

2. Impact of Vaccinations:

If one plotted the overall immunization curve against the new cases curve, the inference would be higher aggregate vaccination leading to fewer new cases over a roughly similar timeframe. Some provinces where vaccination drives were more aggressive started slowing new infections when the number of vaccinations reached a certain threshold.

3. Country-Specific Insights:

Vaccination data was collected and supplemented with case statistics to facilitate a detailed analysis of the pandemic situation in the individual countries. It brought out such disparities with regards to local policies, health system, and vaccine availability in dealing with the different countries with the pandemic.

4. Visualization Efficacy:

It was well reflected in the produced line graphs with respect to how vaccination and infection rates are trending over time. These were particularly useful in cases where one could superimpose the timelines of the vaccination and infection rates.

Submitted by,

Aparna P Santhosh

233304

DA -Bio AI