

# **Exploratory analysis of patterns and correlations in STD incidence using large-scale Public Health Data**

## **Objective:**

To analyze large-scale epidemiological datasets on sexually transmitted diseases to identify key correlations, trends, and demographic patterns. The goal is to uncover insights about factors like age, gender, geography, and socio-economic variables that influence STD incidence and prevalence.

## **Data type:**

High volume structured data.

## **Methodology:**

Use Apache Spark to process and clean large multi-year datasets, combining disease incidence and demographic factors.

Conduct exploratory data analysis (EDA) including:

- Time trends and seasonality by region and disease type.
- Correlation analysis between demographic factors and STD rates.