# SQL-Based Data Analysis Project

## Introduction

This project demonstrates SQL-based data analysis using three different public dataset domains: COVID-19, E-commerce Sales, and HR Analytics. The objective is to showcase data extraction, filtering, aggregation, and reporting capabilities of SQL to derive meaningful business insights.

## 1. COVID-19 Dataset Analysis

Tables: covid\_cases(country, date, confirmed, deaths, recovered), covid\_vaccinations(country, date, total\_vaccinations, people\_vaccinated, people\_fully\_vaccinated).

Queries:

-- Total cases, deaths, and recoveries by country

SELECT country, SUM(confirmed) AS total\_cases, SUM(deaths) AS total\_deaths, SUM(recovered) AS total\_recovered FROM covid\_cases GROUP BY country ORDER BY total\_cases DESC;

-- Daily new cases for a specific country

SELECT date, confirmed - LAG(confirmed) OVER (ORDER BY date) AS new\_cases FROM covid\_cases WHERE country = 'India';

-- Death rate by country

SELECT country, (SUM(deaths) \* 100.0 / SUM(confirmed)) AS death\_rate\_percent FROM covid\_cases GROUP BY country ORDER BY death\_rate\_percent DESC;

-- Top 5 countries by vaccination coverage

SELECT country, MAX(people\_fully\_vaccinated) AS fully\_vaccinated FROM covid\_vaccinations GROUP BY country ORDER BY fully\_vaccinated DESC LIMIT 5;

## 2. E-commerce Sales Dataset Analysis

Tables: orders(order\_id, customer\_id, order\_date, total\_amount), customers(customer\_id, customer\_name, region), products(product\_id, product\_name, category, price), order\_items(order\_id, product\_id, quantity).

Queries:

-- Total sales by region

SELECT c.region, SUM(o.total\_amount) AS total\_sales FROM orders o JOIN customers c ON o.customer\_id = c.customer\_id GROUP BY c.region ORDER BY total\_sales DESC;

-- Monthly sales trend

SELECT DATE\_TRUNC('month', order\_date) AS month, SUM(total\_amount) AS monthly\_sales FROM orders GROUP BY month ORDER BY month;

-- Top 5 best-selling products

SELECT p.product\_name, SUM(oi.quantity) AS total\_sold FROM order\_items oi JOIN products p ON oi.product\_id = p.product\_id GROUP BY p.product\_name ORDER BY total\_sold DESC LIMIT 5;

-- Average order value (AOV)

SELECT AVG(total\_amount) AS avg\_order\_value FROM orders;

## 3. HR Analytics Dataset Analysis

Tables: employees(emp\_id, name, department, salary, hire\_date, gender, age), attrition(emp\_id, attrition\_flag).

Queries:

-- Average salary by department

SELECT department, AVG(salary) AS avg\_salary FROM employees GROUP BY department ORDER BY avg\_salary DESC;

-- Employee count by gender

SELECT gender, COUNT(\*) AS employee\_count FROM employees GROUP BY gender;

-- Attrition rate by department

SELECT e.department, (SUM(CASE WHEN a.attrition\_flag = 'Yes' THEN 1 ELSE 0 END) \* 100.0 / COUNT(\*)) AS attrition\_rate FROM employees e JOIN attrition a ON e.emp\_id = a.emp\_id GROUP BY e.department ORDER BY attrition\_rate DESC;

-- Top 5 highest paid employees

SELECT name, department, salary FROM employees ORDER BY salary DESC LIMIT 5;

## Conclusion

This project highlights how SQL can be applied across multiple domains to extract meaningful insights. COVID-19 dataset analysis provides public health metrics, E-commerce analysis highlights customer and sales trends, and HR Analytics provides workforce-related insights such as attrition and salary trends.