COVID-19 Data Analysis System

Name: Sakshi Paralekar

Roll No: 459

College: Patkar Varde College, Mumbai

Date: 03-06-2025

Project Overview

The COVID-19 Data Analysis System is a SQL-based project designed to analyze global pandemic data. It provides insights such as total confirmed cases, death rates per country, and trends over time for specific regions. The data was sourced from Kaggle (https://www.kaggle.com/datasets/imdevskp/corona-virus-report) and loaded into a SQL database for querying.

ER Diagram (Textual Description)

Table: covid19_data

- id (Primary Key)
- observation_date
- country_region
- confirmed
- deaths

Table Creation

```
CREATE TABLE covid19_data (
   id INT PRIMARY KEY,
   observation_date DATE,
   country_region VARCHAR(100),
   confirmed INT,
   deaths INT,
   recovered INT
);
```

Sample SQL Queries

Total confirmed cases by country

```
SELECT country_region, SUM(confirmed) AS total_confirmed
FROM covid19_data
GROUP BY country_region
ORDER BY total_confirmed DESC
LIMIT 10;
```

Death rate per country

```
SELECT country_region, SUM(deaths) AS total_deaths,
      SUM(confirmed) AS total_confirmed,
      ROUND(SUM(deaths)/SUM(confirmed)*100, 2) AS death_rate
FROM covid19_data
GROUP BY country_region
HAVING total_confirmed > 1000
ORDER BY death_rate DESC;
```

Trend for India by date

```
SELECT observation_date, SUM(confirmed) AS total_cases
FROM covid19_data
WHERE country_region = 'India'
GROUP BY observation_date
ORDER BY observation_date;
```

Conclusion

This project helped in understanding how to work with real-world data using SQL. By analyzing COVID-19 data, we were able to identify key insights such as most affected countries, death rates, and trends over time. This kind of analysis is valuable for health agencies, researchers, and decision-makers.