

COVID-19 SQL Analytics Project — Overview

1. Introduction

The COVID-19 pandemic created an urgent need for accurate data analysis to understand how the virus spread across countries. This project focuses on analyzing COVID-19 case trends using structured data stored in an SQLite database. By performing SQL-based analytics, we uncover patterns such as daily new cases, growth rate, monthly summaries, and total cases per country. The project demonstrates how SQL can be used effectively for real-world data analysis and reporting.

2. Abstract

This project analyzes COVID-19 cases for different countries using a custom-designed relational database built in SQLite. It includes tables for country-level information and day-wise case statistics. The analytics involve calculating trends, 7-day moving averages, highest daily spikes, monthly aggregates, and cases per million population. SQL queries such as joins, aggregations, window functions, and subqueries are used to extract insights. The project demonstrates core principles of data modeling, ETL, querying, and real-world data analytics using SQL.

3. Tools Used

- SQLite – Lightweight relational database for storing COVID-19 data
- DB Browser for SQLite – GUI tool for database creation, editing, and query execution
- SQL – For writing analytical queries and implementing views
- Dataset Source – Sample

4. Steps Involved in Building the Project

Step 1: Designing the Database Schema

Created two tables:

- country: stores country details such as name, continent, and population
- cases: stores day-wise total cases, new cases, and deaths for each country
- A foreign key relationship ensures proper linking.

Step 2: Inserting COVID Data

Inserted sample COVID-19 data for countries like India and USA, covering multiple dates.

This forms the dataset for analysis.

Step 3: Data Cleaning

Performed basic data validation:

Checking NULL values

Making sure date formats are consistent

Ensuring case numbers are non-negative

Step 4: Running Analytical SQL Queries

Developed and tested essential analytics:

- Latest total cases per country
- Daily new cases trend
- Highest daily cases
- Monthly case and death summaries
- Cases per million population
- 7-day moving average (using window functions)
- Summary view for reporting
-

Step 5: Creating Views

- Created a view named `country_summary` to quickly fetch:
- Latest total cases
- Latest total deaths
- Country and population details
- This improves reusability and simplifies complex reports.

5. Conclusion

This project successfully demonstrates how SQL can be used to analyze real-world pandemic data. By building and querying a structured SQLite database, meaningful insights such as trends, peaks, and case distribution can be extracted. The project highlights essential analytical SQL skills such as joins, aggregations, window functions, grouping, and view creation. It provides a strong foundation for anyone working with data analysis, SQL development, or database reporting.