COVID-19 Data Analytics Using SQL

Introduction

This project focuses on analyzing the global COVID-19 pandemic using SQL. By leveraging cleaned datasets, we aim to uncover patterns and trends related to confirmed cases, deaths, and recoveries across different countries.

Abstract

We used two key datasets: one providing daily confirmed COVID-19 statistics by country, and another summarizing the latest per-country snapshot. SQL queries were used to analyze trends, calculate moving averages, determine death/recovery rates, and identify high-risk countries based on new and cumulative cases.

Tools Used

- SQLite
- DB Browser for SQLite
- Datasets from Kaggle
- Python (for preprocessing and PDF generation)

Steps Involved in Building the Project

1. Imported the datasets into SQLite after inspection and cleaning.
2. Created two tables: 'covid_daily' and 'country_latest'.
3. Designed and executed SQL queries to:
   - Rank countries by confirmed cases.
   - Calculate 7-day moving averages.
   - Analyze global daily trends.
   - Identify countries with the highest death and recovery rates.
4. Used window functions and aggregate queries for in-depth analysis.

Conclusion

This project demonstrates how SQL can be used to extract insights from pandemic-related data. We highlighted the power of queries like GROUP BY, window functions, and ranking to understand trends over time and across countries.