**COVID-19 Data Analysis**

**Abstract**

This project focuses on analyzing COVID-19 statistics using SQL queries and SQLite database tools. A Kaggle dataset was imported, cleaned, and transformed to remove inconsistencies. Proper tables and constraints were created to store daily COVID-19 data by country. SQL analytical queries such as top affected countries, daily global trends, and 7-day moving averages were executed using GROUP BY and window functions. The project demonstrates how SQL can be applied effectively for real-world data analysis, with outputs exported for visualization and reporting.

**Tools Used**

* **SQLite + DB Browser** – For database creation, table design, and SQL query execution.
* **Kaggle COVID-19 Dataset** – Source of global COVID-19 data.
* **Excel/Google Sheets** – For visualization of exported results.

**Steps**

1. **Import Dataset** – Load Kaggle CSV file into SQLite database.
2. **Clean and Transform** – Standardize date format, handle missing values, and compute active cases.
3. **Create Tables** –

countries → Stores country information.

covid\_daily → Stores daily statistics (confirmed, deaths, recovered, active).

1. **Apply Constraints & Indexes** – Ensure data integrity using primary/foreign keys and indexing.
2. **Write Queries** – Use GROUP BY and window functions for trends (top countries, daily totals, moving averages).
3. **Export Results** – Generate CSV/Excel reports for charts and visualization.

**ER Diagram**

# ****COVID-19 Data Model****

## ****Entities and Attributes****

### **1. covid\_daily Table**

* **Primary Keys:**
  + country
  + date
* **Attributes:**
  + confirmed: Total confirmed COVID-19 cases
  + deaths: Total COVID-19 deaths

### **2. location Table**

* **Primary Key:**
  + country
* **Attributes:**
  + region: Geographical region of the country
  + continent: Continental location

### **3. daily\_stats Table**

* **Primary Keys:**
  + country
  + date
* **Attributes:**
  + new\_cases: Daily new COVID-19 cases
  + new\_deaths: Daily new COVID-19 deaths
  + moving\_average: Calculated moving average of cases/deaths

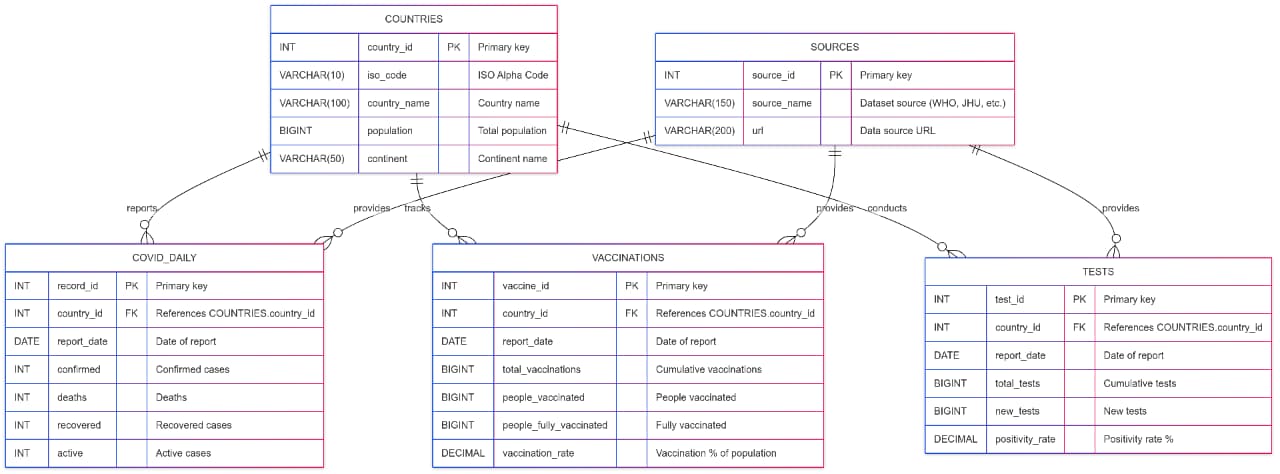
## ****Relationships****

### **1. covid\_daily → location**

* **Type:** One-to-One
* **Description:** Each covid\_daily record belongs to a specific location.
* **Representation:** covid\_daily ||--|| location

### **2. covid\_daily → daily\_stats**

* **Type:** One-to-One
* **Description:** Each covid\_daily record generates corresponding daily statistics in daily\_stats.
* **Representation:** covid\_daily ||--|| daily\_stats



**Conclusion**

The project successfully demonstrated COVID-19 data analysis using SQL. By importing and cleaning Kaggle datasets, creating structured tables, and executing analytical queries, meaningful insights were derived about the pandemic’s trends. Queries such as top countries, daily totals, and moving averages provided a clearer understanding of the data. Exporting results for visualization further highlighted how SQL can support real-world data analysis tasks in a simple and effective manner.