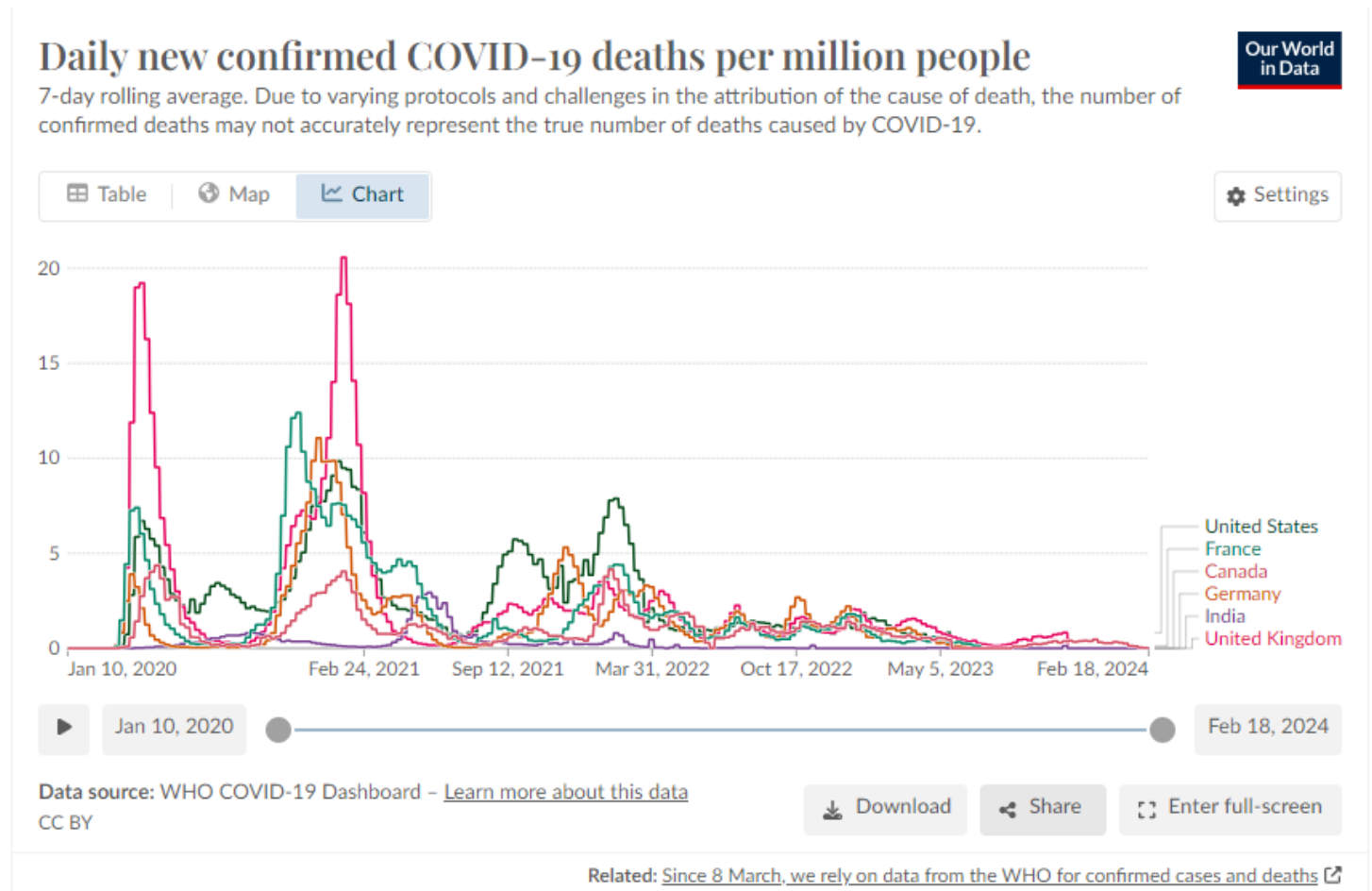


COVID-19 Data ETL Project Documentation

Introduction:

This document describes the process of extracting, transforming, and loading (ETL) data related to COVID-19 deaths and vaccinations from a comprehensive dataset provided by Our World in Data.



Source Data:

- **Location:** <https://ourworldindata.org/covid-deaths>
- **Format:** CSV
- **Date Range:** Jan 10, 2020 - Feb 4, 2024
- **Data Scope:** Countries with population over 10 million

ETL Process:

1. Data Extraction:

- The original dataset, "owid-covid-data_original_ver.csv", is split into two separate CSV files:
 - **CovidDeaths.csv:** Contains data on COVID-19 infections and deaths over time.
 - **CovidVaccinations.csv:** Contains data on vaccination programs.

2. Data Transformation:

- A Python script, "Database_creation.py", utilizes SQLite3 to create a database named "Covid19.db". This script assumes the CSV files are located in the same directory as

the script itself.

3. **Data Loading:**

- The script creates two tables within the database:
 - **CovidDeaths:** Stores data from the "CovidDeaths.csv" file.
 - **CovidVaccinations:** Stores data from the "CovidVaccinations.csv" file.

4. **Optional: Converting to SQL Server/MySQL Compatible Format:**

- Another Python script, "DB_to_SQL_Converter.py", can be used to convert the "Covid19.db" database into an SQL script named "Covid19.sql" compatible with database servers like MySQL and SQL Server.

5. **MySQL Compatibility Adjustments:**

- The generated "Covid19.sql" file requires minor modifications for compatibility with MySQL:
 - Remove the first line containing "TRANSACTION".
 - Replace all single quotes (") with backticks (`).
 - Change the data type of the "population" column from "INT" to "BIGINT".

Output:

- **SQLite3 Database:** "Covid19.db" (suitable for local analysis)
- **SQL Script:** "Covid19.sql" (compatible with MySQL and SQL Server deployments)

Additional Resources:

- The Python scripts ("Database_creation.py" and "DB_to_SQL_Converter.py") are not included in this documentation but are assumed to be part of the project.
- The original dataset is available through <https://ourworldindata.org/covid-deaths>.

MySQL Compatibility:

For users who prefer MySQL, a compatible SQL script named "mysql_Covid19.sql" is available in the "Ready" folder. This script can be directly imported into your MySQL workbench.

Database Functionality Test:

A separate SQL script named "Data_exploration.sql" is also provided in the "ready" folder. This script serves as a basic test to confirm the functionality of the created database. If the script executes successfully without errors, it indicates that the database is working well.