Executive Summary

This project analyzes global **COVID-19 cases, deaths, and vaccination progress** using MySQL, based on data from **Our World in Data**. The goal was to demonstrate a complete SQL analytics workflow—from loading large real-world datasets to generating meaningful insights.

The project follows a structured approach: large CSV files were ingested into MySQL using command-line tools, data structures were defined upfront, and validation checks were performed before analysis. The analysis focuses on identifying overall trends, high-impact countries, and vaccination progress over time.

A combined view of deaths and vaccination data was created to track **population-level vaccination coverage**, enabling clearer comparisons across locations and time periods. The project emphasizes clean data handling, logical query design, and reproducible analysis.

This repository reflects practical SQL skills commonly used in data analyst roles, including data ingestion, transformation, and time-based analysis.

Key SQL Concepts Used

- `LOAD DATA LOCAL INFILE` for large CSV ingestion
- Database and table schema design
- Data validation using counts and sample queries
- `JOIN` operations across multiple datasets
- Common Table Expressions (CTEs)
- Window functions (`SUM() OVER`) for cumulative calculations
- Derived metrics (e.g., percent of population vaccinated)
- Views to simplify repeated analysis logic