COVID-19 Data Analysis and Visualization (SQL + Python)

This project explores the global and country-level trends of the COVID-19 pandemic using real-world data from Our World in Data, analyzed with SQL and Python (Pandas/Matplotlib). The goal is to understand the spread, severity, and the impact of testing and vaccination efforts through interactive visualizations and insightful metrics.

Key Features

Data Sources:

- CovidDeaths and CovidVaccinations datasets merged on location and date

- Cleaned using SQL and loaded into Python via Google Colab

SQL-Based Analysis (Microsoft SQL Server)

Using SQL, we explored:

- Global death rate vs. case count

- Percentage of population infected per country

- Highest infection and death rates

- Vaccination rollouts over time (using CTEs and temp tables)

- 7-day rolling death averages

- Creation of views like PercentPopulationVaccinated and GlobalNum for easy insights

These insights formed the foundation of our Python visualizations.

Python Visualizations (Google Colab)

Using Matplotlib and Seaborn, we visualized:

1. COVID-19 Waves in the U.S. → Multiple waves visible, with a major spike in late 2020–early 2021

2. Deaths Per Million (U.S.) → Mortality peaked with major case surges, declining post-vaccine

3. Vaccinations vs New Cases (U.S.) → Clear inverse trend showing the impact of mass vaccination

4. Top 5 Populous Countries Comparison → India showed a massive spike in early 2021, while China and Africa maintained lower rates

Tools & Skills Used

- SQL (Joins, CTEs, Aggregations, Temp Tables, Views)

- Python (Pandas, Matplotlib, Seaborn)

- Excel (data validation & prep)

- Google Colab (cloud-based Jupyter environment)

Takeaway

This project demonstrates how data storytelling can uncover powerful insights hidden within large health datasets. It showcases the value of combining SQL-driven backend analysis with front-end Python visualizations to make public health trends more understandable.