COVID-19 Vaccination Database Project — Scenario Document

This document describes the scenario, motivation, and implementation steps for the COVID-19 Vaccination Database project completed by Sai Teja Bathula. The project was designed to demonstrate data collection, normalization, relational schema design, and SQL-based data analysis.

Data Source

The dataset for this project was obtained from the publicly available repository maintained by Our World in Data (OWID): https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations. This dataset contains daily updated information on COVID-19 vaccinations administered across multiple countries, including metrics such as total vaccinations, people vaccinated, people fully vaccinated, and vaccine types used.

Project Implementation

- Downloaded and updated vaccination data in CSV format from the OWID GitHub repository.
- Designed an Entity Relationship Diagram (ERD) to model the vaccination dataset.
- Applied normalization rules (up to Third Normal Form, 1NF–3NF) to reduce redundancy and improve consistency.
- Implemented the database schema in SQLite, creating tables to store vaccination data by country and date.
- Populated the database with the cleaned and structured data from the CSV files.
- Wrote and executed SQL queries for analysis, including identifying top countries by vaccination rate and daily/weekly statistics.
- Tested and validated database functionality to ensure referential integrity and accuracy of the data.

Learning Outcomes

- Practical understanding of data extraction and transformation from real-world public datasets.
- Hands-on experience in relational database design and applying normalization principles.
- Ability to implement and query a SQLite database effectively.
- Skills in documenting and presenting technical solutions in a professional manner.

Conclusion

This project highlights the ability to translate a real-world public health dataset into a structured database system capable of supporting analytical queries. It demonstrates key skills in SQL, database design, and data governance practices which are directly applicable in professional data management and cybersecurity contexts.