# A. Title Page

Lewis University  
CPSC 50900: Database Systems   
Spring 2025 Term Project

MEDICAL CLINIC

[SushmaChowdaryPamu@lewisu.edu](mailto:SushmaChowdaryPamu@lewisu.edu)

# B. Initial Proposal

*Description of Data aimed to store*

The aim of this proposal is to create a robust data storage and analysis system focused on excess death statistics stemming from various health crises, particularly COVID-19. By gathering and synthesizing comprehensive mortality data from multiple countries, the proposal seeks to provide critical insights into the true impact of the pandemic and other significant health events on population health. We will be storing excess death statistics from various countries. This dataset includes several key metrics, such as total deaths, COVID-19 related deaths, expected deaths, non-COVID deaths, and derived metrics such as excess deaths per 100,000 people and percentage change in mortality rates. This comprehensive approach allows for a nuanced understanding of mortality trends over time and in different contexts.

*Interest over the Selected Data*

The main focus on this dataset derives from its essential use for measuring COVID-19 and other health emergencies' effects on population deaths. The analysis of public mortality figures above typical levels helps identify complete pandemic fatalities so that public health strategies can be developed to handle these challenges moving forward.

*Importance of Data*

Public health policymakers together with health researchers and health organizations depend on excess death data as their essential resource. This data reveals the healthcare system burdens and helps authorities decide how resources will be appropriated while developing intervention approaches. Researching excess deaths allows stakeholders to advance their knowledge about healthcare results as well as analyze public health performance and create plans for overcoming future health issues.

*Data Come From*

Data will be obtained from CSV files showing excess death statistics compiled for Austria and Belgium as well as Britain and Germany and various other nations. The data collection happens through national health agencies together with statistical offices and research institutions which survey public health trends and mortality rates.

*Data Uses By*

Similar kind of data will serve epidemiologists together with public health officials and researchers and policymakers as its key end users. Participants using the data analysis will examine mortality patterns to find death-related risks for above-average mortality rates while developing foundation-based health strategies and responses.

*Planning to Build Application*

Users will access an interactive analysis platform which enables them to study and understand excess death data. The application enables users to compare data across countries along with regional information while it allows temporal analysis of death rate changes and informant users about health intervention impacts. The accessible format of the data SETS user’s up to make purposeful findings while enabling them to contribute to better public health decisions.

# C. Data Source

The project incorporates multiple COVID-19 excess death data files in CSV format that has been collected to present statistics about mortality from different countries. The COVID-19 Excess Death Tracker on The Economist website grants researchers access to comprehensive mortality trend data which enables them to evaluate pandemic consequences. The data files contain four key groups of information which include country names and total deaths and COVID-19 related deaths as well as expected deaths and excess deaths (Friedman, 2019). Every data file possesses systematic order which supports consistent datasets through rows that show unique observations while time-spanned rows exist alongside multiple demographic and mortality-specific variables in the columns.

Data originates from a meticulously developed tracker at The Economist that analyzes recent mortality statistics by comparing them against historical data-based expected totals. A data collection from The Economist includes approximately 20 countries' death statistics to survey both immediate COVID-19 effects as well as underlying public health trends. The defined methodology enables researchers together with policymakers to examine historical along with present-day health events which affect population mortality levels through a straightforward analysis system that functions as a fundamental record for continued evaluation.

I will create independent tables in the relational database setup that both contain the structures for handling country demographics and death statistics. A Country table contains unique identifiers combined with national names while the Death Statistics table references these identifiers for documenting multiple death statistics. The database will use constraints for maintaining data integrity through identifier non-null checks and format standardization. The structured database with data cleanliness procedures and consistency checks provides critical tools for public health strategy development and decision making which expands the accessibility of vital mortality information. The data files will become accessible through a public GitHub repository which provides both transparency as well as easy access to stakeholders who wish to analyze excess death metrics.

DataSource Link: https://data.world/liz-friedman/covid-19-excess-death-tracker-from-the-economist

# D. Alternative Ways to Store the Data

**1. Hierarchical Database Model**

The hierarchical database model operates through a tree system to store data by connecting records as parent-child relationships with multiple child records possible under each parental record. The data hierarchy for excess death statistics would have countries as top nodes which contain sub-nodes displaying total deaths alongside COVID-19 deaths and expected deaths and excess deaths. We should add one more node which will show dates and time series data to provide detailed exploration from country to death statistics across different periods.

*Implementation:*

A DBMS suitable for hierarchical models would include systems such as IBM's Information Management System and Microsoft’s Windows Registry for implementation (Alrhaimi, 2024). The data storage system defines hierarchical records with each level containing its own schema that maintains parent-child relations to the next level. This model enables easy and effective data retrieval along these relationships to provide fast metric access without requiring SQL complexity.

*Advantages:*

The hierarchical model renders application design simpler since it stores data using organizational relationships which represent realistic hierarchies (Huang et al., 2024). This data structure provides easy access to the related data because all statistics related to a specific country appear together. Trans-versing its hierarchical structure enables the system to execute complex searches which leads to boosted performance of data retrieval for specific user queries.

*Disadvantages:*

The hierarchical model displays various drawbacks because making changes to a data structure requires extensive redesign work for new data categories. Relational databases surpass the hierarchical system by providing better querying functionality especially regarding the representation and effectiveness of many-to-many relationship analytics. This includes the process of comparing data between different countries and their health metrics.

**2. NoSQL Document Store (JSON)**

You can implement modern data storage through the MongoDB NoSQL system as a document store. The model allows data storage through JSON-like documents that support flexible as well as nested data structures (Kumar, 2024). Each single document in this database would represent a distinct country while containing total death statistics along with COVID-19 data and expected and excess death figures along with date information. By using a NoSQL document store any structural changes to dataset metrics can be dealt with effortlessly through data structure modifications which do not need schema adjustments.

*Implementation:*

The storage of our data in a NoSQL document store happens through an implementation that uses MongoDB to create one document per country inside a collection.

*Advantages:*

One main benefit of the NoSQL document store is its adaptable approach. The database enables flexible structure definition and quick development speed alongside adaptable data capacity which expands without hierarchical limitations. JSON format is supported by the document store because it facilitates easier web technology integration in contemporary application development.

*Disadvantages:*

# Using a NoSQL document store system brings the complication of working without a standardized querying language while making it difficult to perform complex queries. The ability of NoSQL databases to expand their capacity by adding more servers comes with the disadvantage of weak support for transactions and inconsistent data management compared to relational systems. This combination presents risks for maintaining data integrity unless proper management measures are established. Newcomers need to overcome a longer learning curve because the documentation quality of these systems is usually not as advanced as established relational database management systems.

# R. Activity Log

**Activity Log**

| **Week** | **Day** | **Activities** |
| --- | --- | --- |
| Week 1 | Day 1 | On Day 1 I searched for information on excess death statistics and selected the metrics that would make up the dataset. I reviewed The Economist’s COVID-19 Excess Death Tracker as my main data source. |
| Day 3 | Studied CSV file structures and tabular organization. The analysis of dataset fields helped identify structural relationships which would determine how the database should be implemented. |
| Day 5 | I developed a first draft proposal during Day 5 to describe project boundaries together with the essential components such as data resources and analytic importance of examining excess mortality numbers. |
| Week 2 | Day 1 | I defined our relational database schema with specifics about both Country and Death Statistics tables during Day 1. These data types as well as constraints have been defined according to their specifications. |
| Day 3 | Researched SQL best practices for data integrity and normalization. The database schema received refined structural changes to remove duplicate data entries. |
| Day 5 | I started preparing and cleaning CSV files for the purpose of maintaining comprehensive and uniform data structures. |
| Week 3 | Day 1 | Implementation of cleaned data occurred on Day 1 through transfer to a MySQL database followed by initial database querying assessment. Testing of initial SQL queries allowed me to retrieve and analyze important statistical data. |
| Day 3 | Investigated NoSQL alternatives, particularly MongoDB, for flexible data storage. |
| Day 5 | Designing a JSON-based NoSQL document schema took place on Day 5 to ensure both scalable growth potential and effective retrieval capabilities. |
| Week 4 | Day 1 | Implemented MongoDB database structure. The system involved database data insertion followed by query retrieval testing. |
| Day 3 | Day 3: An assessment of relational and NoSQL database advantages and disadvantages took place on this day. |
| Day 5 | Day 5: The researcher compiled recorded information about storage alternatives and project implications into documentation on 5th day. |
| Week 5 | Day 1 | On Day 1 I programmed Python scripts that handled CSV file parsing together with database entry into both relational and NoSQL systems. |
| Day 3 | Day 3:  On Day 3 of the project I started developing the interactive data visualization application by drawing designs for main application features and interface building blocks. |
| Day 5 | Day 5:  On day 5 the team brought together database queries with basic front-end display of mortality trend information. |
| Week 6 | Day 1 | The team refined the data visualization dashboard with new functionality to filter data according to countries selected together with date ranges on Day 1. |
| Day 3 | Implemented dynamic data retrieval from both SQL and NoSQL databases. |
| Day 5 | The team performed debugging and performance optimization for data querying efficiency on 5th day. |
| Week 7 | Day 1 | The first day of this project was dedicated to uploading data files alongside project documentation to GitHub to enhance both accessibility and transparency. |
| Day 3 | A final project review took place on Day 3 to validate conformity with the established rubric criteria. |
| Day 5 | On fifth day I executed final assessments while developing a presentation that included project targets together with research techniques and results. |
| Week 8 | Day 1 | Finalized project report and interactive visualization features. |
| Day 3 | I submitted all necessary materials through GitHub repository and provided access to the structured database along with application demonstration on Day 3. |
| Day 5 | The last day was devoted to reviewing own work and documenting essential project learning’s along with space for future enhancement ideas. |

**Reference**

Friedman, L. (2019). COVID-19 excess death tracker from The Economist. Data.World. <https://data.world/liz-friedman/covid-19-excess-death-tracker-from-the-economist>

Alrhaimi, S. A. (2024). Innovative aspects of designing and managing database systems for modern companies in the energy sector. In E3S Web of Conferences (Vol. 541, p. 04008). EDP Sciences.

Huang, K., Liu, D., Chen, T., Wang, Y., Wang, C., & Shi, W. (2024). Real-time map rendering and interaction: a stylized hierarchical symbol model. International Journal of Digital Earth, 17(1), 2367728.

Kumar, T. V. (2024). A Comparison of SQL and NO-SQL Database Management Systems for Unstructured Data.