Global Life Expectancy Analysis Project Proposal

Caleb Damron
Department of Electrical Engineering and Computer Science
The University of Tennessee,
Knoxville Tennessee
cdamron2@vols.utk.edu

Clayton Tucker
Department of Electrical Engineering and Computer Science
The University of,
Knoxville Tennessee
ctucke24@vols.utk.edu

Cole Price
Department of Electrical Engineering and Computer Science
The University of Tennessee,
Knoxville Tennessee
yhb368@vols.utk.edu

Danyil Chuprynov
Department of Electrical Engineerng and Computer Science
The University of Tennessee,
Knoxville Tennessee
dchupryn@vols.utk.edu

Dhruv Patel
Department of Electrical Engineerng and Computer Science
The University of Tennessee,
Knoxville Tennessee
dpate122@vols.utk.edu

***Abstract*— This paper proposes an interactive data visualization platform to examine the connection between life expectancy and social, economic, and environmental factors across nations. We will integrate publicly available datasets from the World Bank Group and Kaggle, arranging them by country and year to report our findings through an interactive leaderboard and statistical charts. Our goal is to identify the factors that have the most influence on life expectancy and provide an intuitive interface for identifying these correlations. This proposal outlines our sources of data, methodology, team responsibilities, timeline, and deliverables.**

I. OBJECTIVE

The purpose of this project is to investigate and visualize what influences life expectancy in countries. We will examine economic figures like GDP and average income, social indicators like happiness and freedom scores, and environmental figures like air quality, healthcare expenditure, and access to clean water. By correlating and analyzing multiple datasets, we hope to determine which variables most strongly impact life expectancy worldwide. Our main product is an interactive web page with a sortable data table and dynamic visualizations that allow users to explore this data in a helpful manner.

II. MOTIVATION

Life expectancy is a straightforward measure of the health and well-being of a nation, but it varies significantly between countries. It is also very useful for public health to know why this variation exists. We aim to make sophisticated international health statistics easier to comprehend through an interactive webpage where users can explore correlations between life expectancy and each of the factors. Our first hypothesis is that healthcare spending, air quality, and GDP should all positively correlate with life expectancy. We also expect to find countries where people live long lives despite having limited economic resources. In addition to our research goals, the project offers direct experience with the whole data science pipeline, from raw data gathering to developing an actual working visualization tool.

III. DATA

A. Data Sources and Collection

We will draw our primary data from public sources, particularly the World Bank Open Data and Kaggle datasets, focusing on country-level statistics. The World Bank offers comprehensive information on life expectancy, GDP per capita, income classifications, and healthcare spending. Kaggle provides additional datasets covering happiness indices, freedom metrics, and air quality measurements. We may supplement these with data from WHO or UN databases as we continue building out the application.

B. Data Processing and Integration

The data integration process involves downloading CSV/Excel files from identified sources, cleaning to handle missing values and standardize country names, and merging datasets using country names and years as keys to create a complete dataset. Each observation will represent a country-year with associated metrics including life expectancy, GDP per capita, income group, happiness index, freedom index, air quality index, healthcare expenditure, and access to clean water where available. The final cleaned dataset will be exported to CSV format so they can be utilized by the frontend. Key challenges include missing data across countries and years, inconsistent country naming conventions across datasets, and ensuring data quality from multiple sources. These will be addressed through careful cleaning measures and thorough documentation of our limitations.

C. Methods and Analysis

The analytical approach involves computing descriptive statistics and Pearson correlation coefficients between life expectancy and each independent variable to quantify relationships. Countries will be grouped by income level and region to identify patterns. On top of this, we will identify nations that deviate significantly from expected trends.

The platform will feature multiple visualization types:

- Interactive Data Table: A sortable table displaying all countries with their metrics, allowing users to sort by any column
- Dynamic Scatter Plot: Life expectancy fixed on the y-axis while users select different factors from a dropdown menu for the x-axis
- Correlation Heatmap: Matrix visualization showing relationships between all variables
- Boxplots: Compare life expectancy distributions across income groups

The system will be built using HTML, CSS, and JavaScript with Chart.js for visualizations. Pre-processed CSV or JSON data files will be loaded directly into the frontend, requiring no backend server infrastructure. The application will be locally hosted for demonstration purposes.

IV. Responsibilities

To ensure efficient project execution, responsibilities will be divided among team members as follows. All team members will work on data collection, identifying and downloading datasets from World Bank, Kaggle, and other sources. Dhruv Patel and Caleb Damron will handle data cleaning and integration, merge datasets, handle missing values, and standardize formats. Dhruv Patel, Clayton Tucker, and Cole Price will perform statistical analysis, including correlation analysis, descriptive statistics, and interpretation of findings. Caleb Damron, Danyil Chuprynov, and Cole Price will lead visualization development, implementing the interactive table, dynamic graphs, and other visualizations using Chart.js. Danyil Chuprynov, Dhruv Patel, and Cole Price will focus on frontend development, designing and implementing the overall website structure, layout, and user interface. Clayton Tucker, Caleb Damron, and Danyil Chuprynov will manage documentation, maintain project records, and write sections of the final report. All team members will work on testing, debugging, and the preparation of the final report and presentation. Regular team meetings will promote our coordination and encourage optimal problem-solving.

V. Timeline

Milestone	Deadline
Identify and download initial datasets	October 7
Complete data cleaning and integration	October 14
Finalize data validation and export processed dataset	October 16
Complete correlation analysis and preliminary findings	October 23
Develop interactive data table	October 28
Implement dynamic scatter plot with dropdown	October 30
Create additional visualizations	November 4
Design and implement website frontend	November 6
Integration testing and bug fixes	November 11
Draft final report sections	Mid November
Complete final report	Mid-Late November
Prepare and deliver class presentation	Late November

VI. Expected Outcome

The primary deliverable will be a locally hosted web application featuring an interactive sortable data table where users can explore country rankings across all metrics, a dynamic scatter plot with dropdown selection for examining correlations with life expectancy, and statistical visualizations including correlation heatmaps and boxplots by income group. The interface will be designed for intuitive use by non-experts.

The secondary deliverable will be a comprehensive final report (4-6 pages in IEEE format) documenting data sources and processing methods, statistical analysis results including correlation coefficients, discussion of findings and identification of outlier countries, limitations of the analysis, and suggestions for future work. We expect to find positive correlations between life expectancy and GDP per capita, healthcare expenditure, air quality, and happiness index scores, with high-income countries showing significantly higher average life expectancy. We expect to identify outlier countries that achieve high life expectancy despite moderate GDP due to efficient healthcare systems or strong social factors.

VII. Conclusion

This project addresses which social, economic, and environmental factors most significantly influence life expectancy across countries. By integrating multiple public datasets and creating an interactive visualization platform, we aim to make complex global health data accessible and interpretable. The project provides valuable experience in the complete data science workflow, from acquiring data through analysis to visualization and presentation. The resulting web application and comprehensive report will serve as both an educational tool and a demonstration of our team's technical capabilities in data science and web development.