A close-up of a logo

Description automatically generatedA close-up of a logo

Description automatically generated

**University of Westminster**

**Individual Coursework – 5DATA004C Data Science Project Lifecycle**

**NAME** : Mohammed Shakir

**STUDENT ID** : w2053035 / 20222306

**DATASET** : Life expectancy at birth for both sexes combined (years)

**Links for Submission**

* **Video Link**: <https://drive.google.com/drive/u/0/folders/13SZIlxZQeb03w0nO2D4WHCncwphbA2My>
* **Streamlit App URL**: <https://dspl-srilanka-life-expectancy-dashboard-8u8drlfvzfuj6u8szrx6qr.streamlit.app/>
* **GitHub Repository**: <https://github.com/mohammedshakir7/DSPL-SRILANKA-LIFE-EXPECTANCY-DASHBOARD>

Table of Contents

[**Links for Submission** 1](#_Toc197900936)

[1. Dataset Selection and Overview 3](#_Toc197900937)

[Dataset Selection 3](#_Toc197900938)

[Dataset Criteria 3](#_Toc197900939)

[2. Aims and Objectives 4](#_Toc197900940)

[Aims 4](#_Toc197900941)

[Objectives 4](#_Toc197900942)

[3. Development Methodology 5](#_Toc197900943)

[1. Data Preparation 5](#_Toc197900944)

[2. Technologies Used 6](#_Toc197900945)

[3. Application Structure 6](#_Toc197900946)

[4. Functional and Non-Functional Requirements 7](#_Toc197900947)

[Functional Requirements 7](#_Toc197900948)

[Non-Functional Requirements 7](#_Toc197900949)

[5. Testing 8](#_Toc197900950)

[Test log 8](#_Toc197900951)

[Test Cases 8](#_Toc197900952)

[6. Conclusion 9](#_Toc197900953)

# 1. Dataset Selection and Overview

## **Dataset Selection**

For this project, the dataset used is titled *"Life Expectancy at Birth for Both Sexes Combined (Years)*. The data was sourced from the **United Nations Data** platform, and it includes life expectancy values for Sri Lanka from 1950 to 2100. The data is publicly available and can be accessed from the official platform <https://data.un.org/Data.aspx?q=Sri+lanka&d=PopDiv&f=variableID%3a68%3bcrID%3a144>

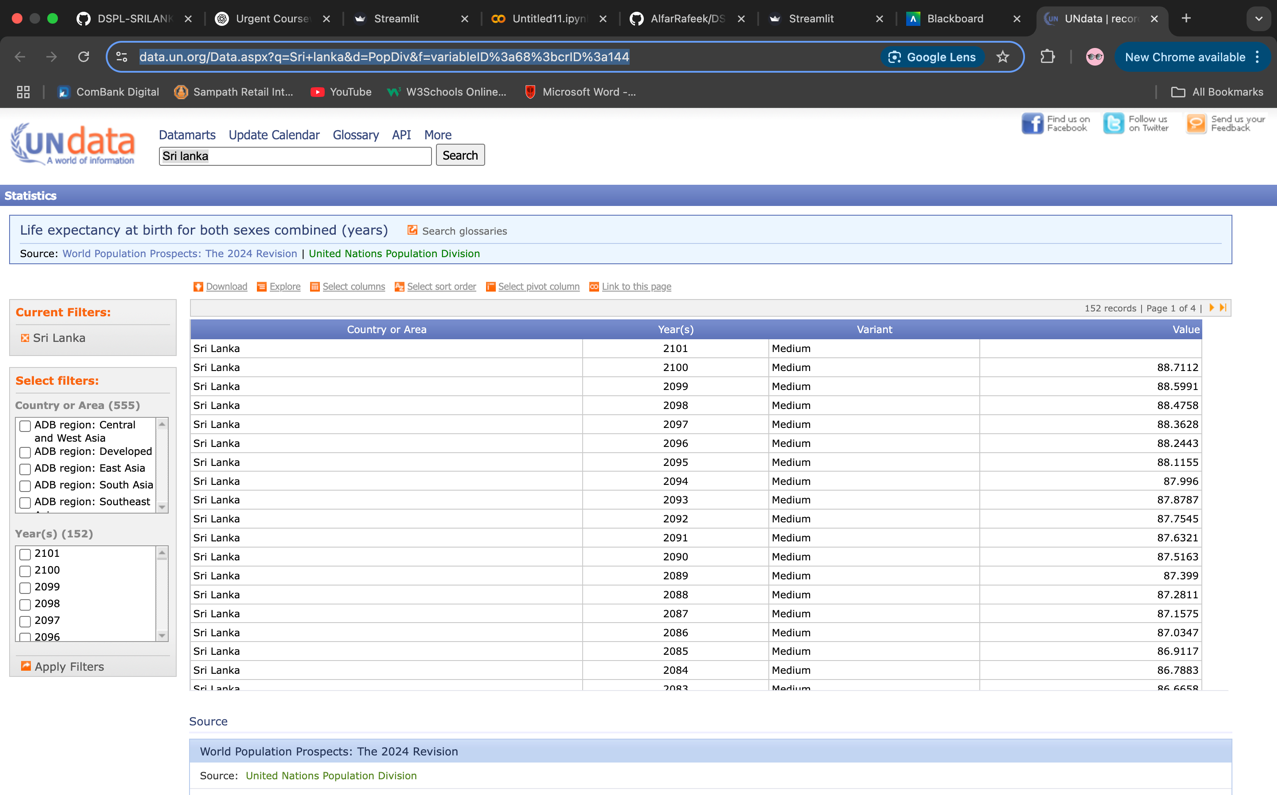
The dataset was retrieved in **CSV format** and then imported into **Google Colab** for further cleaning and processing. The raw data was initialy in a wide format, which was later transformed into a clean, analyzable dataset by removing any discrepancies or missing values.

This dataset meets the following critaria:

* **Publicly Accesible**: The data is freely available for research and analysis through the United Nations Data repository
* **Data Range**: It includes life expectancy data from 1950 to 2100, containing more than 100 data points, fulfilling the minimum requirement of the project.
* **Data Quality**: The dataset was cleaned in **Google Colab**, where I removed any inconsistencies, such as missing values, and transformed it into a usable format for analysis.

## Dataset Criteria

* **Source**: The dataset is sourced from the United Nations Data platform.
* **Size**: The dataset includes multiple years (over 100 data points) from 1950 to 2100, representing life expectancy for Sri Lanka.
* **Variables**: The dataset contains the following key variables:
  + Year(s): The year for which the data is recorded.
  + Variant: The variant of the population data (e.g., "Medium").
  + Value: The life expectancy value for both sexes combined, recorded in years.



# 2. Aims and Objectives

## Aims

The primary aim of this project is to design and develop an **interactive data visualization dashboard** that facilitates the exploration of **life expectancy data** **f**or Sri Lanka, spanning from 1950 to 2100. By presenting this data in an engaging and interactive format, the dashboard will provide valuable insights into the trends, fluctuations, and future projections of life expectancy in Sri Lanka. The goal is to offer an intuitive and user-friendly tool for individuals, researchers, and policymakers who wish to analyze the country’s demographic health trends over a significant period. This platform will enable users to visualize not only historical data but also future projections, enhancing their understanding of how life expectancy has evolved and is projected to evolve.

## Objectives

The specific objectives of the project are:

1. **To Build an Interactive Dashboard**  
   Develop an interactive dashboard using **Streamlit** to present Sri Lanka's life expectancy data, enabling easy exploration and data interaction for both technical and non-technical users.
2. **To Provide Multiple Visualization Options**  
   Offer **bar charts** and **line charts** to visualize life expectancy trends, helping users understand key insights, such as overall trends and anomalies.
3. **To Enable Data Filtering by Year Range**  
   Provide a **slider tool** that allows users to filter the data based on a customizable year range, enabling more detailed analysis of specific periods.
4. **To Offer Customization through a Color Picker**  
   Allow users to customize chart colors via a **color picker** to enhance the visual appeal and user experience of the dashboard.
5. **To Ensure User-Friendliness and Accessibility**  
   Deliver a **user-friendly interface** that is intuitive and easy to navigate, ensuring accessibility for all types of users.
6. **To Provide Insightful Analysis and Future Projections**  
   Enable users to explore **long-term trends and future projections**, helping them identify patterns and understand the factors influencing life expectancy changes.

# 3. Development Methodology

## 1. Data Preparation

The data was loaded and cleaned using Pandas. The cleaning process included:

* **Removing missing values**: Ensuring that the dataset is complete and free from any missing data.
* **Correct data types**: Ensuring that columns like 'Year’ and 'Value' are of the correct data type (numeric, integer for the year, and float for the values).
* **Data Normalization**: Ensuring consistent formatting and handling of outliers to ensure the integrity of visualizations.

## 2. Technologies Used

The project uses the following technologies:

* **Streamlit**: For building the interactive dashboard.
* **Panda**: For data manipulation and cleaning.
* **Plotly**: For creating interactive visualizations (bar and line charts).
* **Matplotlib**: For static charting.
* **AgGrid**: To display the data table interactively.
* **Numpy**: To handle numerical operations (if necessary)

## 3. Application Structure

The Streamlit app is divided into multiple sections:

* **Sidebar**: Allows users to interect with the data by selecting chart types, year ranges, and bar colors.
* **Main Content**: Displays the main title, data table, and visualizations such as bar charts and line charts.
* **Footer**: Provides additional information, such as the project details and developer name.

# 4. Functional and Non-Functional Requirements

## Functional Requirements

1. **Interactive Sidebar Filters**:
   * **Year Range Selection**: A slider to filter the data based on selected year ranges.
   * **Chart Type Selection**: A drop down menu allowing the user to chose between bar and line charts.
   * **Color Picker**: A color picker to customize the chart's color scheme.
2. **Main Content Area**:
   * **Data Table**: Displaying the cleaned data interactively using AgGrid.
   * **Visualizations**: Displaying bar charts and line charts that update based on user selections.

## Non-Functional Requirements

1. **Performance**:
   * The app sholud load the data quickly (within 3-5 seconds).
   * Interactive elements (slider, dropdown, and color picker) should respond in real-time without significant delays.
2. **Usability**:
   * The dashboard should be easy to navigate with intuitive controls (sidebar filters, chart selections, etc.)
   * The app should be responsive, adapting well to both desktop and mobile devices
3. **Scalability**
   * The app should be able to handle larger datasets in the future with minimal performance degradation
4. **Accessibility**:
   * The app should be user-friendly, with clear labels and instructions on how to use the interactive elements.

# 5. Testing

## Test log

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TC** | **Date** | **Executed by** | **Actual Result** | **Pass/Fail** | **Notes** |
| TC1 | 2025-05-07 | Mohammed Shakir | The dashboard loads all charts and elements | Pass | Working link deployed |
| TC2 | 2025-05-07 | Mohammed Shakir | The year filter is working correctly | Pass | Filtered data matches year range |
| TC3 | 2025-05-07 | Mohammed Shakir | Bar chart shows correct life expectancy data | Pass | Data matches the selected years |
| TC4 | 2025-05-07 | Mohammed Shakir | Line chart updates as expected with color filter | Pass | Color picker functionality works |
| TC5 | 2025-05-07 | Mohammed Shakir | Color picker changes the chart colors | Pass | Visual updates are applied |

## Test Cases

* **TC1**: Open the Streamlit app and ensure it loads without errors.
* **TC2**: Select a year range from the slider and ensure the data updates accordingly.
* **TC3**: Switch between bar and line charts and verify correct chart display.
* **TC4**: Use the color picker and ensure the bar chart color updates as per the selected value.
* **TC5**: Ensure the sidebar functions properly and all filters work.

# 6. Conclusion

The primary aim of this project was to develop an interactive dashboard to visualize Sri Lanka's life expectancy data from 1950 to 2100. Using **Streamlit**, I successfully built an engaging and user-friendly platform that allows users to explore trends in life expectancy over the decades. The dashboard includes features like **data filtering by year range**, customizable chart options (bar and line charts), and a color picker for user personalization. The use of **Python**, **Streamlit**, **Plotly**, and **Matplotlib** enabled the cretion of effective, interactive visualizations that make public health data more accessible and understandable.

The manual cleaning and preprocessing of the dataset, which was sourced from reliable public platforms like the **UN DATA**, proved to be a critical task in ensuring that the data was accurate and ready for analysis. Although the process was somewhat time-consuming, it emphasized the importance of data integrity and structure for accurate insights. In future iterations, I would consider automating the data extraction and cleaning process to save time and enhance scalability.

Overall, this project was a valuable learning experience in data wrangling, interactive visualization, and web application deployment. It highlights how data science tools can be utiliized to support public health awareness and decision-making, enabling users to gain valuable insights into the trends and projectons of life expectancy in Sri Lanka. The dashboard has been designed to meet both academic requirements and the needs of general users looking to understand life expectancy trends.