

Name- Dawit Habte Tekeste

Msc GPH -

Matriculation No 22404947

Report on R Project: Maternal Health Risk Analysis and Visualization

1. Introduction and Context

This project focuses on analyzing and visualizing maternal health risk data using R. The goal was to clean a real-world dataset, perform exploratory data analysis, and create an interactive visualization using Shiny. The dataset, obtained from Kaggle, contains maternal health indicators such as age, blood pressure, and blood sugar, with an associated risk level classification (low, mid, high).

The key user persona is public health professionals and clinicians aiming to understand patterns in maternal risk factors and improve risk stratification through interactive visual tools. They seek to quickly identify risk groups and explore key variable distributions to inform healthcare interventions.

2. Data Cleaning and Preparation

The raw dataset contained some inconsistencies and irrelevant records, including extreme ages and character-type inconsistencies. The cleaning process involved:

- **Removing invalid records** such as ages below 15 or above 50, which are unlikely for maternal health.
- **Converting relevant variables** into appropriate data types, including converting the `RiskLevel` variable from character to factor.
- **Creating a numeric risk score variable** to map risk levels (`low=1`, `mid=2`, `high=3`) for correlation and visualization purposes.
- Ensuring consistent encoding and formatting for all variables.

This cleaning was scripted in `cleaning.R` and applied to produce `cleaned_maternal_data.csv`.

3. Exploratory Data Analysis and Visualization

Using `Visualization.R`, exploratory analyses were performed including:

- **Bar plots** showing count distribution of patients by risk levels, with color palettes adjusted for clarity.
- **Boxplots** comparing numerical variables (age, systolic blood pressure, blood sugar) across risk levels to identify significant differences.
- **Correlation heatmap** analyzing numeric variables in relation to the numeric risk score, highlighting positive and negative associations.

Visualizations were generated using `ggplot2` and `corrplot`, facilitating insights into key risk factor patterns.

4. Interactive Visualization via Shiny App

The core deliverable is an interactive Shiny app (`app.R`) that integrates the cleaned dataset and visualizations. Features include:

- **Dynamic filtering and selection** by risk levels and variables.
- Real-time updates of plots based on user input.
- A user-friendly interface designed for public health analysts and clinicians.

The app was developed following best practices in Shiny UI and server logic separation and deployed to

5. Summary and Impact

This project successfully demonstrates the end-to-end process of data cleaning, analysis, and interactive visualization using R and Shiny. The modular approach with separate scripts (`cleaning.R`, `Visualization.R`, and `app.R`) ensures reproducibility and ease of maintenance.

The Shiny app offers stakeholders an accessible tool to explore maternal health risks visually, supporting data-driven decision making in healthcare settings.

SUBMISSION LINKS

Shiny app link https://dawithabte.shinyapps.io/my_project/.

Github Repo <https://github.com/wite38/maternal-health-shiny-app>