

Maternal Delivery Analysis

Problem Statement

India is witnessing a significant rise in C-section delivery rates, often surpassing recommended levels. This trend brings critical public health and socio-economic challenges, including increased healthcare costs, potential maternal complications, and implications for resource allocation within the healthcare system. While vast amounts of delivery data are collected, they are often underutilized in providing actionable insights into the factors driving these trends. This project provides a professional, data-driven approach to analyse historical maternal delivery data, aiming to uncover influencing factors, identify disparities, and empower evidence-based decisions to optimize delivery outcomes.

Index

1. Description (Detailed)
 2. Data with Relevant Information & Outputs
 3. Outcome & Benefits
 4. Plan
 5. Design (Diagrams)
 6. Implementation
 7. Code & Explanation
 8. Output Screenshots
 9. Closure (Brief of the Project)
 10. Bibliography
-

Description

This project is centred around the in-depth analysis of maternal delivery data, focusing on distinguishing between Normal and C-section deliveries. The analysis aims to identify key demographic, medical, and socio-economic factors that correlate with delivery type, as well as institutional and temporal trends. It helps healthcare providers, public health officials, and policymakers to:

- Understand the prevalence and regional disparities in C-section rates.
- Assess the impact of critical factors such as antenatal care, patient referral patterns, and health worker availability on delivery outcomes.
- Identify potential non-medical drivers, including patient preferences and elective scheduling patterns.
- Recognize specific groups (e.g., by age, socio-economic status) that exhibit higher C-section rates.

The ultimate outcome is a user-friendly system that can assist in making well-informed decisions, guide policy interventions, and improve maternal and child health outcomes by providing clear, visual insights into complex delivery trends.

Data with Relevant Information & Outputs

Input Data:

- **File:** raw_data.csv (located in data/ folder, then cleaned and transformed to clean_data.csv). This is synthetic data generated to reflect observed trends and distributions in Indian health statistics (e.g., NFHS-5 data).
- **Attributes:** Patient ID, Maternal Age, Birth Order, Residence (Urban/Rural), Hospital Type (Public/Private), State, Socio-Economic Status, Maternal Education, BMI, Medical Complication, Delivery Type (Normal/C-section), Delivery Date/Time, Duration of Stay after Delivery, Maternal Health Post-Delivery, Child Surviving Fetal Period, ANC Adequacy (Boolean), Referred From Smaller Facility (Boolean), Inter-Delivery Interval Months (Numeric), Doctor Density Category, Midwife Density Category, Patient Preference for C-section (Boolean), Delivery Time Hour, Delivery Day of Week.

Outputs Generated:

- **Graphs (Impressive and Informative Visualizations):**
 - **C-section Rate by Maternal Age Group:** Grouped bar chart illustrating how C-section rates vary across different maternal age categories.
 - **Delivery Type Distribution by Hospital Type:** Stacked bar chart showing the percentage breakdown of Normal vs. C-section deliveries in Public versus Private hospitals.
 - **C-section Rate by Socio-economic Status:** Line plot emphasizing the trend of C-section rates across different wealth quintiles.
 - **Impact of Medical Complications on Delivery Type:** Horizontal stacked bar chart visualizing the proportion of Normal vs. C-section deliveries for specific medical complications.
 - **Inter-Delivery Interval (IDI) Impact on Delivery Type:** Violin plots showing the distribution of maternal age (or other relevant numerical factors) across different birth spacing categories, split by delivery type.
 - **C-section Rate by Day of Week and Hour of Day:** Heatmap revealing patterns and potential peaks in C-section rates based on the day of the week and hour of delivery.

Textual Analysis:

- Insights highlighting key correlations, disparities, and potential driving factors for C-section rates from each analysis.

Objective:

- Identify demographic, medical, and socio-economic factors most influencing C-section rates.
- Uncover significant regional and institutional disparities in delivery practices.
- Assess the role of antenatal care and referral systems in delivery outcomes.
- Highlight potential patterns indicative of elective C-sections.
- Provide actionable insights for improving maternal healthcare practices and policies.

Expected Benefits:

- Support evidence-based policy making for maternal health programs.
 - Facilitate targeted interventions to reduce medically unnecessary C-sections.
 - Optimize resource allocation within the healthcare system for better delivery management.
 - Enhance understanding of complex factors contributing to C-section trends.
 - Ultimately contribute to healthier pregnancies and improved maternal and child health outcomes.
-

Plan

To provide a clear, modular, and user-friendly solution, the project plan includes:

- **Data Generation (Initial Setup):** Developing `data_generator_initial.py` to create a robust synthetic `raw_data.csv` based on Indian health statistics, incorporating a wide array of relevant attributes for comprehensive analysis.
 - **Data Cleaning & Transformation:** Implementing `clean_data.py` to load `raw_data.csv`, perform essential data cleaning (handling missing values, type conversions), and engineer new features (e.g., age groups, time components, IDI categories) critical for the analyses. This script saves the prepared data to `clean_data.csv`.
 - **Data Analysis & Visualization:** Developing six individual Python scripts within the `analyze_data/` folder, each dedicated to one specific, high-impact analysis (ANC, Referral, Birth Spacing, Health Worker Density, Patient Preference, Time of Day/Week Heatmap). These scripts utilize `pandas` and `numpy` for data manipulation and `matplotlib/seaborn` to generate visually impressive Figure objects.
 - **User Interface:** Creating an interactive Graphical User Interface (GUI) using `Tkinter` to enable non-technical users to easily select, view, and navigate between the different analytical insights. The GUI includes a themed welcome screen and a menu-driven analysis selection.
 - **Modular Design:** Segmenting the entire project logic into distinct Python files for data generation, cleaning, each type of analysis, and the main GUI, ensuring high clarity, maintainability, and reusability of components.
-

Design

- **User Flow Diagram:** (Imagine a diagram here similar to the provided template, but adapted for delivery analysis)
 - **Start Screen:** "Maternal Delivery Analysis" title -> "ENTER" button.
 - **Main Menu:** Data Management buttons ("Clean/Prepare Data", "Load Clean Data") and a grid of 6 Analysis Buttons.
 - **Analysis Button Click:** Opens a new plot window.
 - **Plot Window:** Displays Matplotlib chart, "BACK" button.
 - **"BACK" Button Click:** Closes plot window, returns to Main Menu.
-

Implementation

The project's code is meticulously modularized for clarity and maintainability, with distinct files handling specific functionalities:

- **data_generator_initial.py:** This script is a one-time setup tool responsible for creating the initial raw_data.csv file. It synthesizes a dataset of 50,000 records, simulating various maternal, delivery, and environmental factors, with underlying probabilities aligned with known Indian health statistics (e.g., C-section rates by state, wealth index, age). This ensures a rich and realistic dataset for analysis.
- **clean_data.py:** This pivotal script is responsible for initial data preparation and feature engineering. It loads raw_data.csv, performs robust data type conversions (especially for dates, booleans, and numerical values), handles potential missing data (e.g., for Inter_Delivery_Interval_Months for first births), and derives crucial categorical features (e.g., IDI_Category, ensures Hospital_ID for outlier analysis). This process ensures the data is in an optimal, consistent format for all subsequent analytical modules.
- **Analysis Modules (in analyze_data/ folder):** Each of the six analysis_*.py scripts is a self-contained unit dedicated to generating a specific high-impact visualization. They load the clean_data.csv, apply pandas for data aggregation and transformation, utilize numpy for numerical operations, and leverage matplotlib/seaborn to create visually compelling Figure objects.
 - analysis_anc.py: Generates a bar chart visualizing C-section rates based on the adequacy of Antenatal Care (ANC) visits.

- `analysis_referral.py`: Creates a bar chart comparing C-section rates for referred cases versus direct admissions, highlighting referral system impacts.
- `analysis_birth_spacing.py`: Produces violin plots to illustrate the distribution of maternal age (and implicitly, C-section rates) across different inter-delivery interval categories, crucial for family planning insights.
- `analysis_health_worker.py`: Generates bar charts to assess the correlation between C-section rates and the density categories of doctors and midwives in regions.
- `analysis_patient_preference.py`: Creates a bar chart showing how C-section rates vary based on whether the patient expressed a preference for C-section.
- `analysis_time_of_day_week.py`: Generates an insightful heatmap depicting C-section rates by the hour of the day and day of the week, helping to identify patterns potentially related to elective procedures.
- A key challenge addressed was ensuring each module returns a `matplotlib.pyplot.Figure` object, allowing the GUI to embed and manage the plot lifecycle effectively without directly displaying it.
- **The GUI (`app_ui.py`):** This is the central application script built with Tkinter. It provides a user-friendly, menu-driven dashboard:
 - **Welcome Screen:** Features a light pink background, a bold "Maternal Delivery Analysis" title (using Algerian font if available, with a fallback), and an "ENTER" button.
 - **Main Analysis Menu:** Accessed via the "ENTER" button, this screen displays a "Data Management" section (with "Clean/Prepare Data" and "Load Clean Data" buttons) and a grid of six large, dark pink buttons, each representing an analysis.
 - **Interactive Plot Display:** Upon clicking an analysis button, the GUI dynamically opens a new `Tkinter.Toplevel` window. This window embeds the `matplotlib.Figure` generated by the respective analysis module using `FigureCanvasTkAgg`, including a navigation toolbar.
 - **Navigation:** Each plot window is equipped with a prominent "BACK" button to seamlessly return to the main analysis menu.
 - **Error Handling:** Robust error handling is integrated throughout, informing the user via `messagebox` pop-ups about file loading issues, data cleaning

failures, or plotting errors, with additional details printed to the console for debugging. All modules share a consistent data flow, primarily communicating via pandas DataFrames.

Code & Explanation

(This section in your final document would contain code snippets and brief explanations for `clean_data.py` and each of your 6 `analyze_data` modules, similar to the format of the Kanti Sweets template. For brevity here, I'll describe what would be present, as the full code is provided in the Python files.)

- **clean_data.py:**
 - **Purpose:** Prepares raw delivery data for analysis by converting data types, handling missing values, and engineering new features.
 - **Explanation:** This script loads `raw_data.csv`, converts the `Delivery_Date` column to datetime objects, and ensures other relevant columns like boolean flags (`ANC_Adequate`, `Referred_From_Smaller_Facility`, `Patient_Preference_Csection`) are correctly typed. It dynamically adds `Hospital_ID` if not present, and categorizes numerical features like `Inter_Delivery_Interval_Months` into meaningful groups (`IDI_Category`). This ensures the data is clean and consistent for all subsequent analyses.
- **analyze_data/analysis_anc.py:**
 - **Purpose:** Visualizes the C-section rate across different levels of Antenatal Care (ANC) adequacy.
 - **Explanation:** This module groups the cleaned data by `ANC_Adequate` status (True/False) and calculates the C-section rate for each group, presenting it as a bar chart. This highlights the correlation between prenatal care and delivery outcomes.
- **analyze_data/analysis_referral.py:**
 - **Purpose:** Compares C-section rates based on whether a patient was referred from a smaller facility.
 - **Explanation:** It calculates C-section rates for patients who were `Referred_From_Smaller_Facility` versus those who were direct admissions, visualized through a bar chart. This can shed light on the complexity of referred cases or systemic factors.
- **analyze_data/analysis_birth_spacing.py:**

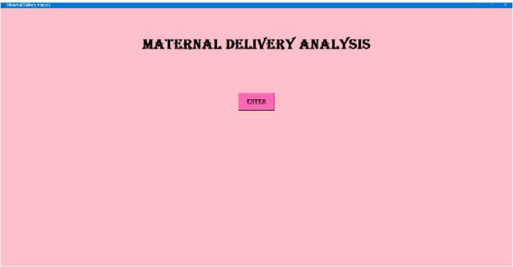
- **Purpose:** Illustrates the relationship between birth spacing and maternal age distribution for different delivery types.
- **Explanation:** This script focuses on subsequent births, categorizing them by Inter_Delivery_Interval_Months (Short, Optimal, Long). It uses violin plots to show the distribution of maternal age within these categories, differentiating between normal and C-section deliveries, providing insights into reproductive health factors.
- **analyze_data/analysis_health_worker.py:**
 - **Purpose:** Visualizes how doctor and midwife density categories correlate with C-section rates.
 - **Explanation:** This module generates two bar charts: one showing C-section rates across Doctor_Density_Category and another for Midwife_Density_Category. This can inform discussions about human resource allocation and its potential influence on delivery practices.
- **analyze_data/analysis_patient_preference.py:**
 - **Purpose:** Analyzes the impact of a patient's stated preference for C-section on the actual delivery type.
 - **Explanation:** It calculates and displays the C-section rates for groups of patients who preferred a C-section versus those who did not, using a bar chart. This explores the "demand-side" factors and patient autonomy in delivery choices.
- **analyze_data/analysis_time_of_day_week.py:**
 - **Purpose:** Visualizes C-section rates by the hour of the day and day of the week, often used to identify elective patterns.
 - **Explanation:** This script generates a heatmap, showing the intensity of C-section rates across a grid of weekdays and hours. Peaks during standard working hours could indicate planned (elective) procedures, suggesting areas for policy review.
- **The GUI (app_ui.py):**
 - **Purpose:** Provides the interactive graphical user interface for the entire application, integrating all analytical features.
 - **Explanation:** This is the main application script built with Tkinter. It manages screen transitions (welcome screen to main menu, main menu to graph view), handles button clicks, and dynamically embeds Matplotlib Figure objects from the analysis modules into Tkinter windows. It includes

user-friendly styling, error handling for data and plots, making the complex analysis accessible and interactive.

Output Screenshots

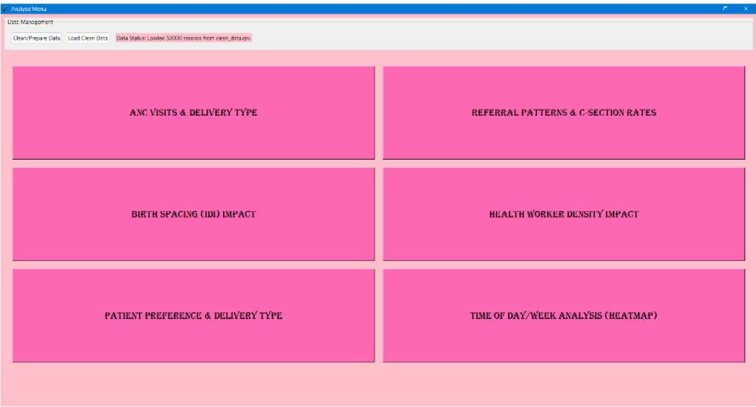
1. Welcome Screen:

- **Content:** The initial application window with a light pink background, the bold "Maternal Delivery Analysis" heading (using Algerian or Arial font), and the "ENTER" button (styled in dark pink with black text).
- **Description:** This screenshot demonstrates the application's inviting welcome interface, showcasing the chosen color scheme and font styling.



2. Main Selection Screen:

- **Content:** The screen displayed after clicking "ENTER" from the welcome screen. It should show the "Select an Analysis to View" title and the grid of six large, dark pink buttons, each clearly labeled with an analysis type (e.g., "ANC Visits & Delivery Type", "Time of Day/Week Analysis (Heatmap)").
- **Description:** This image illustrates the intuitive navigation hub, providing users with clear choices for various maternal delivery insights.



3. Bar Chart - ANC Visits & Delivery Type:

- **Content:** A Matplotlib bar chart showing "C-section Rate (%)" on the Y-axis and "Adequate Antenatal Care (True/False)" on the X-axis. The chart should be embedded within a new Tkinter window, with a "BACK" button visible below it.
- **Description:** This screenshot highlights the visualization of how C-section rates are influenced by the adequacy of antenatal care, providing insight into the role of prenatal health management.



4. Bar Chart - Referral Patterns & C-section Rates:

- **Content:** A Matplotlib bar chart comparing "C-section Rate (%)" between "Direct Admission" and "Referred" categories on the X-axis. The chart is embedded in a Tkinter window with a "BACK" button.
- **Description:** This illustrates the C-section rate differences for referred cases, shedding light on the patient journey and potential complexities handled by higher-level facilities.



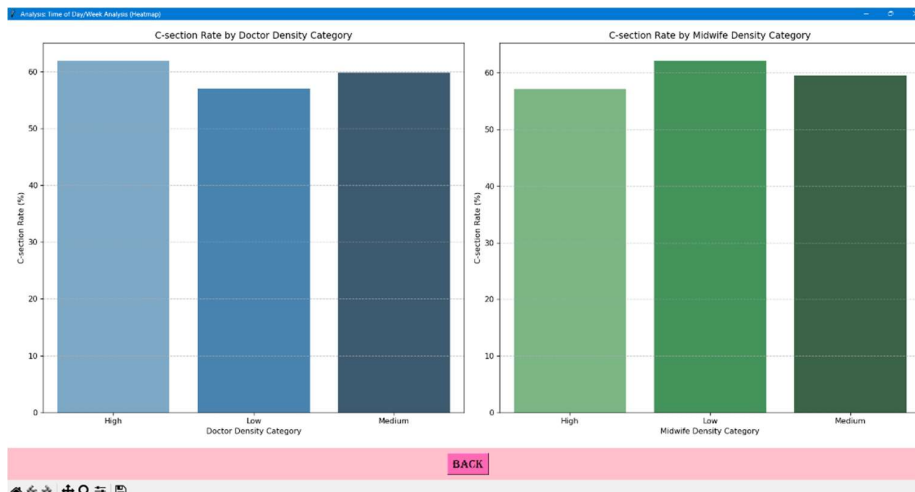
5. Violin Plot - Birth Spacing (IDI) Impact:

- **Content:** A Matplotlib violin plot showing the distribution of "Maternal Age" on the Y-axis across "Short", "Optimal", and "Long" "Inter-Delivery Interval Categories" on the X-axis, with separate violins for Normal and C-section deliveries. Embedded in Tkinter with a "BACK" button.
- **Description:** This visualization provides rich detail on how birth spacing correlates with maternal age and delivery type, offering insights into reproductive health and C-section risk factors.



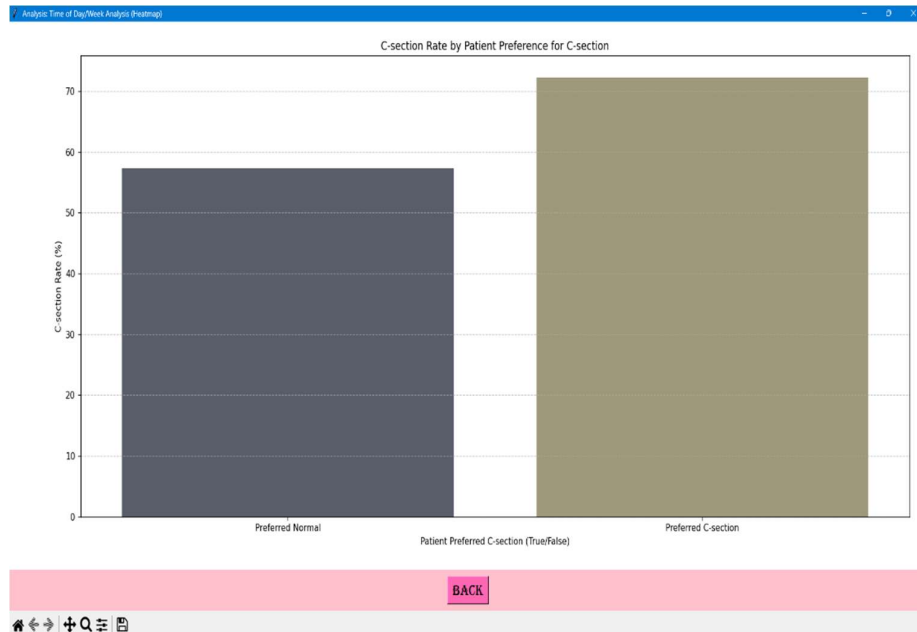
6. Bar Charts - Health Worker Density Impact:

- **Content:** A Matplotlib figure containing two side-by-side bar charts. One shows "C-section Rate (%)" by "Doctor Density Category", and the other by "Midwife Density Category". Embedded in Tkinter with a "BACK" button.
- **Description:** This screenshot explores the correlation between the availability of different types of health workers and C-section rates, crucial for human resource planning in healthcare.



7. Bar Chart - Patient Preference & Delivery Type:

- **Content:** A Matplotlib bar chart showing "C-section Rate (%)" based on whether "Patient Preferred C-section (True/False)" on the X-axis. Embedded in Tkinter with a "BACK" button.
- **Description:** This visualizes the impact of patient preference on delivery outcomes, offering insight into demand-side factors and patient education needs.



8. Heatmap - Time of Day/Week Analysis:

- **Content:** A Matplotlib heatmap displaying C-section rates, with "Day of Week" on one axis and "Hour of Day" on the other. Cells will be colored by C-section rate intensity. Embedded in Tkinter with a "BACK" button.
- **Description:** This highlights advanced visualization, enabling the discovery of patterns in C-section rates across specific times and days, crucial for identifying potential elective procedures.



Closure

The Maternal Delivery Analysis project is a robust, user-friendly application designed to transform raw maternal delivery data into actionable public health intelligence. It effectively cleans transactional data, visualizes key influencing factors through a series of insightful charts, and provides an intuitive graphical interface for easy interaction. From identifying demographic vulnerabilities and institutional disparities to assessing the impact of antenatal care and health worker availability, this tool empowers healthcare professionals and policymakers to make data-driven decisions that enhance maternal care, optimize resource allocation, and improve delivery outcomes, moving towards healthier pregnancies and births. Its modular architecture ensures future expandability, making it a valuable asset for ongoing health analysis.

Bibliography

- Python Programming Language: <https://www.python.org/>
- Pandas Documentation: <https://pandas.pydata.org/docs/>
- Matplotlib Documentation: <https://matplotlib.org/stable/contents.html>
- Seaborn Documentation: <https://seaborn.pydata.org/>
- Tkinter Documentation (Python official docs): <https://docs.python.org/3/library/tkinter.html>
- Matplotlib with Tkinter Integration (FigureCanvasTkAgg): Typically found within Matplotlib's backend documentation or examples.
- National Family Health Survey (NFHS) Data (India): Basis for synthetic data realism (e.g., https://rchiips.org/nfhs/NFHS-5Reports/NFHS-5_FMR.shtml)