

NFHS HEALTH DATA ANALYSIS PROJECT

1. Dataset Description

1.1 Source:

NFHS-5 (National Family Health Survey) Factsheets dataset — collected from government survey data, stored in **NFHS_5_Factsheets_Data.csv**.

1.2 Columns:

- **States/UTs** — Names of Indian States and Union Territories
- **Area** — Urban or Rural classification
- **Number of Households surveyed** — Total households included in the survey
- **Number of Women age 15-49 years interviewed** — Count of women participants
- **Number of Men age 15-54 years interviewed** — Count of men participants
- **Female population age 6 years and above who ever attended school (%)** — Female literacy indicator
- **Population below age 15 years (%)** — Percentage of population under age 15
- **Sex ratio of the total population (females per 1,000 males)** — Gender ratio in total population
- **Sex ratio at birth (females per 1,000 males)** — Gender ratio among newborns
- **Children under age 5 years whose birth was registered (%)** — Birth registration rate *(and several other demographic and health indicators)*

1.3 Data Quality:

- Few numeric columns contain symbols like parentheses and extra spaces (cleaned before analysis)
- Mostly complete data with consistent column structure
- Covers both **Urban and Rural** regions across all Indian States/UTs for a comprehensive demographic overview

2. Operations Performed

2.1 Data Cleaning & Exploration

- Removed extra spaces, parentheses, and formatting errors from numeric fields.
- Converted *Female population age 6 years and above who ever attended school (%)* column to numeric type for accurate filtering.
- Checked for missing or null values — none found.
- Verified unique entries for *States/UTs* and *Area* columns (Urban, Rural, Total).

- Computed descriptive statistics (mean, median, standard deviation) for literacy and population-related columns.

2.2 Descriptive Analytics

- Analyzed **female literacy distribution** across all states and union territories.
- Compared **urban vs. rural** literacy percentages using summary tables.
- Examined **population below age 15 years (%)** to understand demographic balance.
- Visualized **sex ratio trends** and **school attendance percentages** using bar and line charts.

2.3 Relationship Analysis

- Correlated **female literacy rates** with **population below age 15** to assess education impact on youth demographics.
- Compared **sex ratio vs. literacy levels** to identify gender balance trends.
- Analyzed **urban and rural area differences** in literacy and population structure.

3. Key Insights

3.1 Female Literacy Trends

- States with **female literacy above 80%** are primarily urbanized and well-developed regions.
- **Kerala, Delhi, and Goa** recorded the highest literacy percentages among women.
- Rural areas show a noticeable literacy gap of **10–20%** compared to urban counterparts.

3.2 Demographic Patterns

- States with **higher literacy rates** tend to have a **lower population below 15 years**, indicating better education and family planning awareness.
- **Sex ratio at birth** varies significantly, with urban areas generally showing slightly more balanced ratios.
- **Population structure** indicates a younger demographic in rural regions compared to urban ones.

3.3 Social Indicators

- Higher female literacy correlates positively with **mobile phone ownership** and **bank account usage**, reflecting empowerment.
- States with lower literacy percentages also show **higher tobacco and alcohol consumption** among both genders.
- **Birth registration rates** are consistently higher in states with improved female education.

3.4 Regional Comparison

- Southern states outperform northern and eastern regions in most literacy and gender balance indicators.

- Urban areas across all states show **greater access to education and resources** compared to rural zones.

4. Recommendations

4.1 Education & Awareness

- Launch **targeted literacy programs** for women in states with literacy rates below 70%.
- Promote **adult education and skill training** in rural areas to reduce the urban–rural gap.
- Integrate **health and hygiene education** with literacy campaigns to improve social outcomes.

4.2 Healthcare Accessibility

- Strengthen **health infrastructure** in low-literacy and high-population states.
- Encourage **mobile health and telemedicine services** to reach remote areas.
- Improve **maternal and child healthcare awareness** through local community programs.

4.3 Policy & Governance

- Increase **government investment** in female education and empowerment schemes.
- Monitor **sex ratio and child welfare indicators** at the district level for early intervention.
- Establish **data-driven planning tools** for better allocation of resources in underperforming regions.

4.4 Technology & Digital Inclusion

- Expand **digital literacy initiatives** for women, focusing on mobile phone usage and online banking.
- Use **data visualization and predictive analytics** to identify high-priority zones needing intervention.