Biometric Registration Data Analysis Report

Introduction

This notebook presents an in-depth analysis of biometric registrations across India for March–July 2025 using Apache PySpark. It explores demographic trends, geographic distribution, age group patterns, and district-level variations.

Dataset Overview

- **Dataset:** Biometric_update_data_March-July.csv
- **Source:** Government biometric registration dataset
- **Total Records:** 109,743 records covering 6 columns: Date, State, District, Pincode, Bio_age_5_17, Bio_age_17+
- **Columns:** 6 attributes including age-specific counts and location identifiers
- **Data Quality:** Clean dataset with numeric age values; minimal missing values

1. Dataset Description & Cleaning

- Verified data types and converted age columns to numeric
- Checked for missing or anomalous values
- Calculated total population and child-to-adult ratios at district, state, and monthly levels

2. Descriptive Analytics

- **Age Group Distribution:** Total children (5–17): 20,170,712; total adults (17+): 22,371,122; combined population: 42,541,834
- **Monthly Trends:** Registrations varied across months with highest combined registrations in July 2025 (9,794,178)
- **State-wise Distribution:** Maharashtra, Bihar, and Madhya Pradesh consistently show high populations
- **Child-to-Adult Ratio:** Average ratio across all states is ~1.12, indicating slightly higher child population relative to adults in some districts

3. Advanced Analysis

- **District-level Patterns:** Districts with high total population and high child-to-adult ratios identified for targeted updates
- **State-level Trends:** States with consistent increases/decreases in registrations can help prioritize resources
- **Visualization Insights:** Heatmaps, bar charts, and trend lines reveal regional disparities and age group distributions

4. Key Insights

- Adults (17+) dominate the population, but several districts have higher child-to-adult ratios
- Maharashtra, Bihar, and Madhya Pradesh are the most populous states in the dataset
- Monthly trends show growth in registrations with fluctuations across months
- Child-to-adult ratios help highlight districts requiring focused initiatives

5. Visualizations

- Age group distribution pie chart
- Monthly total population trend line
- Child-to-adult ratio histogram
- Top districts by total population (bar chart)
- State-wise average child-to-adult ratio (bar chart)
- District-level heatmaps for child-to-adult ratio

Conclusion

The analysis of biometric registrations for March–July 2025 highlights demographic and geographic patterns across India. While adults form the larger portion of the population, districts with higher child-to-adult ratios are crucial for targeted initiatives. Monthly trends reveal fluctuations in registration counts, emphasizing the need for focused outreach in under-registered regions.

States like Maharashtra, Bihar, and Madhya Pradesh consistently lead in registrations, serving as key areas for government programs. District-level insights, visualized through heatmaps and bar charts, allow policymakers to allocate resources effectively and monitor progress.

Future work could include predictive modeling to forecast registration trends, assess outreach program effectiveness, and explore socio-economic factors influencing registration coverage. These insights provide actionable guidance for government agencies and organizations aiming to improve biometric registration rates and ensure accurate population tracking.