Unlocking Societal Trends in Aadhaar Biometric Activity

Comprehensive Data Analysis and Insights Report

Prepared for: UIDAI – Data Hackathon 2026
Prepared by: Syed Anas Faaiz
Date of Analysis: *14-01-2026*
Dataset Focus: Aadhaar Biometric Update Dataset (Age-wise)
Dataset Period: *March 2025 – December 2025*
Total Records Analysed: 1766212
Geographic Coverage: 57 *States, 974 Districts, and 19707 Pincodes*

Executive Summary

The Aadhaar ecosystem generates extensive enrolment and update data across regions, age groups, and time periods, reflecting how identity services are accessed nationwide. While this data is rich in information, systematic analysis of patterns, trends, and variations is essential to support informed operational planning and service improvement.

This study analyses Aadhaar-related activity data with a primary focus on **age-wise biometric activity**, while using enrolment and demographic datasets for contextual understanding. The analysis examines biometric demand across temporal and geographic dimensions to identify recurring patterns, regional concentration, and periods of demand fluctuation.

Key findings indicate that biometric demand varies significantly across **age groups, regions, and time.** Activity is unevenly distributed geographically, with certain districts consistently contributing a higher share of demand. Temporal analysis reveals a combination of stable demand patterns and short-term spikes, indicating that service requirements are neither uniform nor constant.

Age-wise analysis further shows that biometric demand composition differs across regions, highlighting the importance of **age-aware service planning** rather than a one-size-fits-all approach.

Based on these insights, the study proposes **data-driven improvement opportunities** that do not require major system changes. These include:

- Demand-aligned capacity planning
- Adoption of a hybrid service delivery model combining permanent centres, pop-up locations, and mobile units
- Age-aware planning strategies
- Proactive preparation for temporal demand variations

Collectively, these measures can improve infrastructure utilisation, reduce congestion, and enhance overall citizen experience.

Overall, this analysis demonstrates how existing Aadhaar data can be leveraged to support **more flexible, efficient, and responsive service planning**, enabling incremental improvements in Aadhaar service delivery through informed decision-making.

---

1. Problem Statement

The Aadhaar ecosystem generates large volumes of enrolment and update data across diverse regions, age groups, and time periods. While this data reflects how identity services are accessed, underlying patterns, trends, and variations are not always systematically analysed to support operational planning and service improvement. Identifying meaningful insights from this data can enable better accessibility, improved resource utilisation, and enhanced efficiency in Aadhaar service delivery.

---

2. Objectives of the Study

This study aims to:

- Analyse Aadhaar-related activity data across temporal and geographic dimensions
- Identify meaningful patterns, trends, and variations in service demand
- Understand the influence of age-wise activity on service planning requirements
- Support data-driven improvements in the utilisation of Aadhaar service infrastructure

---

3. Datasets Used

The analysis utilises primary datasets provided by UIDAI, including:

Dataset	Description	Key Fields
Enrolment Dataset	New Aadhaar registrations	Date, State, district,
Biometric Update Dataset	Age-wise biometric activity	Date, State, district, bio_age_5_17, bio_age_17_plus
Demographic Update Dataset	Demographic updates	demo_age_5_17, demo_age_17_plus

Table 1: Summary of datasets used for analysis

4. Analytical Approach

The study follows a data-driven analytical approach focused on identifying trends, variations, and operational insights rather than evaluating system performance.

Key elements of the approach include:

- Analysis of age-wise activity to understand recurring service demand
- Geographic assessment to identify regional concentration of activity
- Temporal analysis to detect seasonal or periodic demand variations
- Comparative analysis to support planning-oriented insights

While multiple datasets are available, the analysis places primary emphasis on age-wise biometric activity, as it directly reflects recurring operational demand. Enrolment and demographic datasets are used selectively to provide contextual understanding.

---

5. Data Understanding and Exploration

Initial exploration confirmed successful data ingestion and verified key attributes such as date, state, district, pincode, and age-wise biometric activity counts. The dataset's scale enables meaningful geographic and temporal analysis. Data type inspection indicated the need for date standardisation prior to time-based analysis.

---

6. Data Cleaning and Preprocessing

The following preprocessing steps were applied to ensure data reliability:

- Standardisation of date formats for temporal analysis
- Validation of geographic identifiers (state, district, pincode)
- Removal of duplicate records to prevent inflated counts
- Verification of numerical consistency across activity fields

These steps ensured a clean and consistent dataset for analysis.

---

7. Feature Engineering

Feature engineering was performed to derive meaningful analytical metrics. A total biometric activity measure was created by combining biometric counts for the 5–17 and 17+ age groups, representing overall service load. An age-wise biometric ratio was also derived to compare child and adult demand.

Time-based features such as year and month were extracted, and data was aggregated at monthly and district levels to support trend and regional analysis. These features formed the basis for exploratory analysis.

---

8. Exploratory Data Analysis and Visualisation

Exploratory analysis was conducted using visualisations to examine:

- Overall age-wise activity distribution
- Geographic concentration of activity
- Temporal trends and seasonal variations
- Differences in demand patterns across locations

8.1 Overall Age-wise Biometric Activity

Biometric demand varies between the 5–17 and 17+ age groups, highlighting the importance of age-specific considerations in service planning.

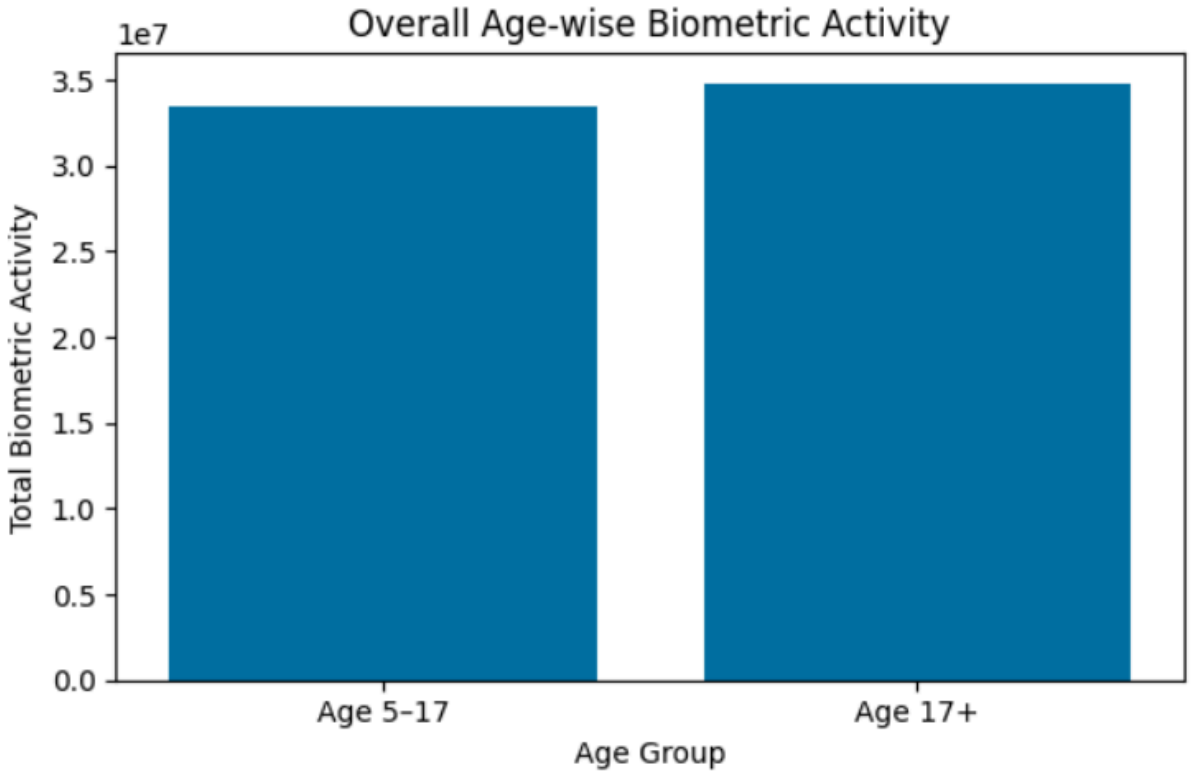

Figure 1: **Overall Age-wise Biometric Activity**

This figure compares biometric demand across the 5–17 and 17+ age groups, highlighting differences in age-wise contribution to overall service demand.

Key Observations:

• Adult biometric activity contributes a higher overall volume

• Child biometric activity remains significant and recurring

8.2 Monthly Trend of Biometric Activity

Monthly aggregation reveals fluctuations in biometric demand, indicating seasonal or periodic patterns relevant to operational planning.

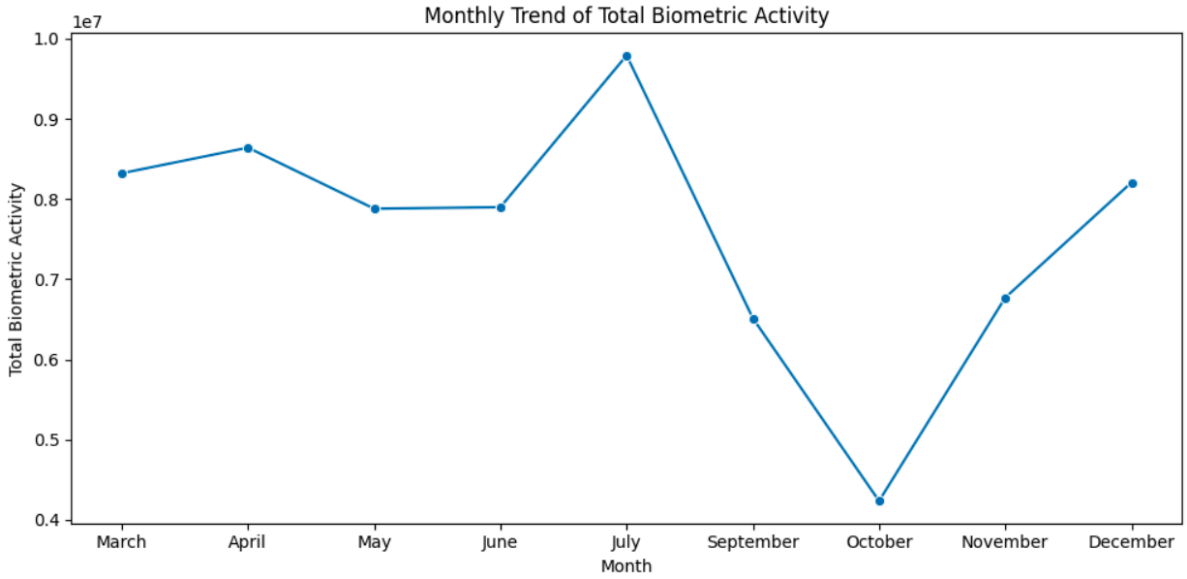

Figure 2: Monthly Trend of Biometric Activity

Key Observations:

- Biometric demand fluctuates across different months.
- Certain months exhibit noticeably higher activity levels.
- Demand does not remain constant over time.

Why this matters:
Monthly variation suggests that biometric service demand follows temporal patterns.
Recognising these trends enables proactive planning instead of reactive capacity adjustments.

8.3 Age-wise Monthly Trends

Comparative trends show that biometric demand does not evolve uniformly across age groups.

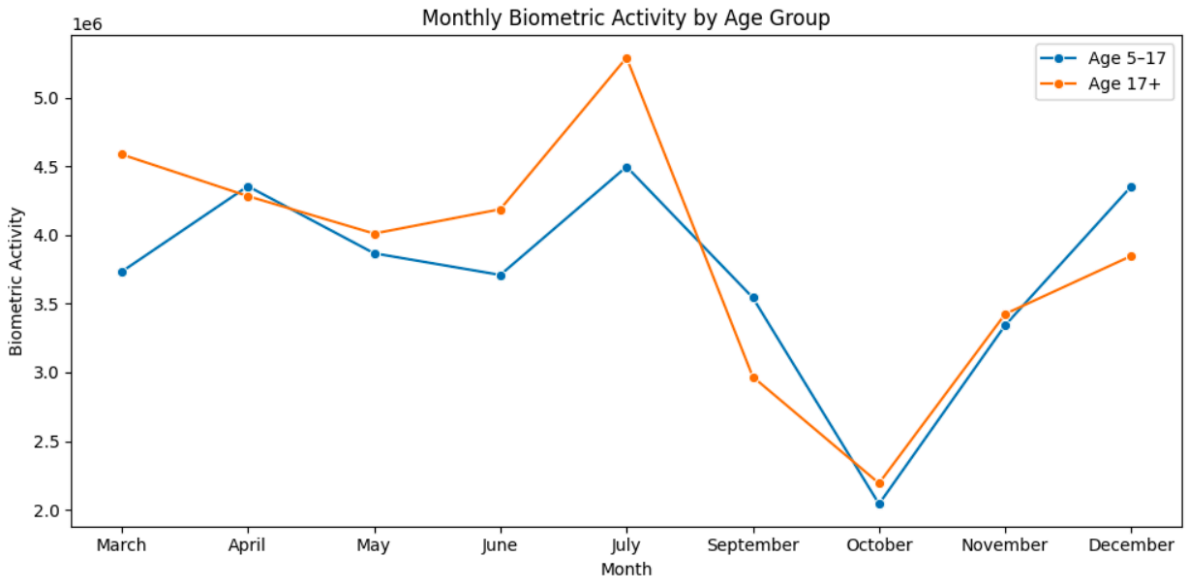

Figure 3: Age-wise Monthly Biometric Activity Trends

Key Observations:

- Both age groups show variation over time, but not always in the same pattern.
- Peaks and declines may differ between child and adult biometric activity.
- One age group may dominate demand during specific periods.

Why this matters:
Differences in age-wise temporal patterns indicate that demand changes do not affect all age groups uniformly. Age-aware service scheduling can therefore improve operational efficiency.

8.4 Regional Concentration

Activity is unevenly distributed across districts, with certain regions contributing a significantly higher share of demand.

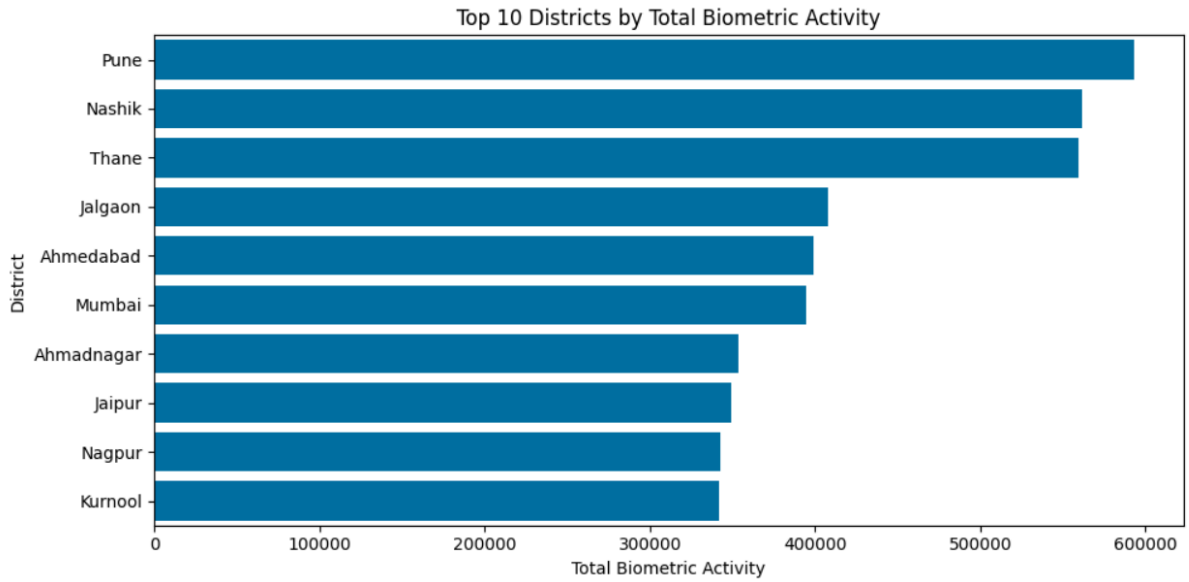

Figure 4: Regional Concentration of Biometric Activity

This figure highlights the top districts contributing the highest total biometric activity.

Key Observations:

- Biometric activity is unevenly distributed across districts.
- A limited number of districts contribute a disproportionately high share of total demand.
- Many districts show comparatively lower activity levels.

Why this matters:
Geographic concentration of demand highlights the need for region-specific planning. Uniform deployment of resources may lead to congestion in high-demand areas and underutilisation elsewhere.

8.5 Age-wise Demand Composition

Child-to-adult biometric ratios vary across regions, indicating differences in demand composition.

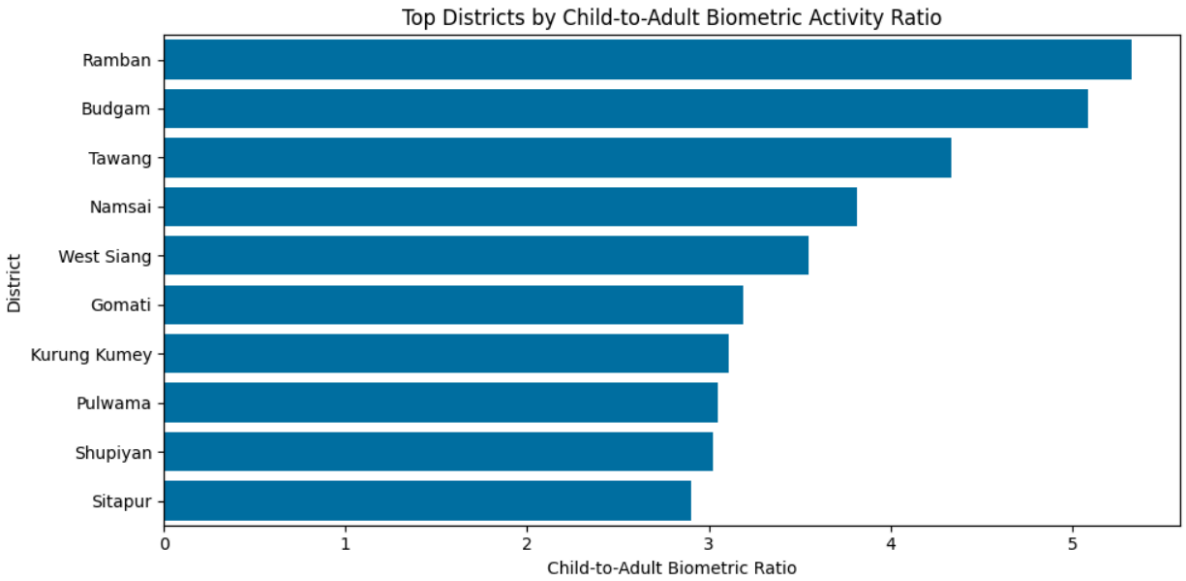

Figure 5: Regional Variation in Age-wise Biometric Demand (Child-to-Adult Ratio)

Key Observations:

- Some regions are relatively child-dominant in biometric demand.
- Other regions are primarily adult-dominant.
- Demand composition varies significantly across districts.

Why this matters:
Understanding demand composition helps tailor service delivery. Regions with higher child-driven activity may require different service setups and timing compared to adult-dominant regions.

8.6 Demand Stability and Spikes

Month-to-month variation shows both stable demand and short-term spikes, suggesting the need for flexible planning.

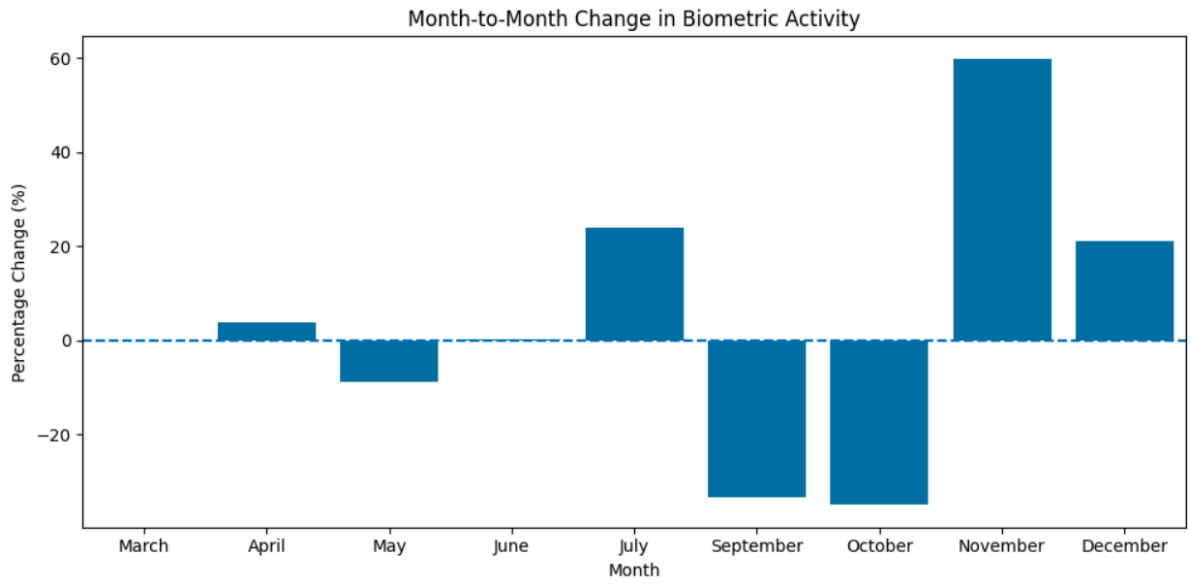

Figure 6: Demand Stability and Temporal Spikes

What this figure shows:
This figure depicts month-to-month changes in biometric activity, highlighting periods of stability and sudden spikes.

Key Observations:

- Certain periods show stable biometric demand.
- Other periods exhibit sharp increases or decreases.
- Demand growth is not always gradual.

Why this matters:
Distinguishing between stable demand and short-term spikes supports decisions on when permanent capacity is sufficient and when temporary measures such as pop-up locations or mobile units are required.

---

9. Key Insights

Key observations from the analysis include:

- Variation in service demand across age groups
- Uneven geographic distribution of biometric activity
- Temporal fluctuations indicating predictable and situational demand changes

These patterns highlight opportunities for anticipatory and region-specific planning.

---

10. Data-Driven Improvement Opportunities

Based on the observed patterns in age-wise, geographic, and temporal biometric activity, several opportunities for improving Aadhaar service planning and delivery are identified. These improvements focus on better utilisation of existing infrastructure, improved responsiveness to demand variations, and enhanced citizen experience, without requiring major system or policy changes.

10.1 Demand-Aligned Capacity Planning

Observation:
Biometric activity is unevenly distributed across regions, with certain districts consistently exhibiting higher service demand.

Improvement:
Service capacity, including staffing levels and operational hours, can be aligned with observed biometric demand rather than uniformly allocated across regions.

Outcome:
This helps ensure that high-demand areas receive adequate support while avoiding underutilisation of resources in low-demand regions.

10.2 Hybrid Service Delivery Model

Observation:
Biometric demand shows a mix of stable patterns and short-term spikes across regions and time periods.

Improvement:
A hybrid service delivery approach can be adopted, where:

- Permanent Aadhaar centres handle steady and predictable demand
- Pop-up locations manage seasonal or periodic surges
- Mobile units are deployed selectively in remote or temporarily overcrowded areas

Outcome:
This approach improves flexibility, reduces congestion during peak periods, and optimises the use of existing infrastructure.

10.3 Age-Aware Service Planning

Observation:
Age-wise biometric demand composition varies across regions, with some areas being child-dominant and others adult-dominant.

Improvement:
Service planning can incorporate age composition insights to tailor service timing, capacity, and setup based on dominant demand groups.

Outcome:
Age-aware planning improves service efficiency and reduces friction for specific user groups.

10.4 Proactive Planning for Temporal Variations

Observation:
Biometric activity fluctuates across months, indicating predictable seasonal or periodic demand changes.

Improvement:
Historical activity trends can be used to proactively adjust staffing levels, operating schedules, and temporary service support ahead of high-demand periods.

Outcome:
Proactive planning reduces last-minute congestion and improves overall service reliability.

10.5 Improving Citizen Experience through Demand Management

Observation:
Demand concentration and temporal spikes contribute to congestion and longer waiting times in certain regions.

Improvement:
Better distribution of service capacity across permanent centres, pop-up locations, and mobile units can be used to manage crowding more effectively.

Outcome:
This leads to reduced waiting times, lower travel burden for citizens, and a more accessible Aadhaar service experience.

---

11. Conclusion

This study demonstrates how systematic analysis of Aadhaar-related data can uncover meaningful demand patterns across age, geography, and time. The findings reinforce the value of data-driven and flexible planning approaches to enhance operational efficiency and service accessibility.

By leveraging historical activity patterns, Aadhaar services can achieve better utilisation of permanent centres, targeted deployment of temporary facilities, and selective use of mobile units—without requiring major system changes.

12. Appendix: Code and Outputs

This section documents the code snippets, intermediate outputs, and visualisations used in the analysis to ensure transparency and reproducibility.

CODE BLOCK 1: Data Loading & Combination

Purpose: Shows you used the official dataset correctly.

```
import pandas as pd

import glob
```

```
# Load all Aadhaar biometric CSV files

files = glob.glob("../datasets/api_data_aadhar_biometric/*.csv")

df = pd.concat([pd.read_csv(f) for f in files],
ignore_index=True)

df.shape
```

CODE BLOCK 2: Data Cleaning

Purpose: Proves data quality handling.

```
# Convert date column to datetime
df['date'] = pd.to_datetime(df['date'], errors='coerce')

# Remove rows with missing key fields
df = df.dropna(subset=['date', 'state', 'district', 'pincode'])

# Remove duplicate records
df = df.drop_duplicates()
```

CODE BLOCK 3: Feature Engineering

Purpose: Shows how insights were derived.

```
# Standardize column name

df = df.rename(columns={'bio_age_17_': 'bio_age_17_plus'})



# Total biometric activity
```

```
df['total_biometric_activity'] = (
    df['bio_age_5_17'] + df['bio_age_17_plus']
)
# Child-to-adult biometric ratio
df['child_to_adult_ratio'] = (
    df['bio_age_5_17'] / (df['bio_age_17_plus'] + 1)
)
```