# UNLOCKING SOCIETAL TRENDS IN AADHAAR ENROLMENT AND UPDATES (2025)


**AUTHOR:** Tenali Anisha

**TEAM ID:** UIDAI_1093

# TABLE OF CONTENTS

## 1. Executive Summary

India's Aadhaar system plays a critical role in identity verification, access to welfare schemes, financial inclusion, and administrative governance. While Aadhaar coverage is extensive, the patterns of enrolment and subsequent updates are not uniform across regions, age groups, or time. Understanding these variations is essential for improving service delivery, ensuring equitable access, and strengthening the overall effectiveness of the system.

This study presents a comprehensive, data-driven analysis of Aadhaar enrolment and update behaviour across India in 2025, with the objective of uncovering meaningful societal patterns, regional disparities, and systemic gaps that can inform policy and administrative decision-making. Using official **UIDAI Aadhaar datasets** on enrolment, biometric updates, and demographic updates, the analysis integrates over **2.9 million records** spanning **March to December 2025** to provide a holistic view of how Aadhaar usage varies across states, districts, time, and age groups.

The project begins with rigorous data consolidation and preprocessing, **merging** three large raw **datasets (enrolment, biometric, and demographic)** using a **full outer join** on **common keys—date, state, district, and pincode**—to ensure no loss of information. Significant effort was invested in data cleaning, including handling missing values, correcting schema inconsistencies, standardizing state names using normalization, manual mapping, and fuzzy matching, and removing invalid records that contained locality-level names in the state column. This resulted in a validated dataset covering **36 official Indian states and Union Territories**, ensuring reliable geographic analysis.

Through exploratory and structured analysis, several key insights emerged. Key findings reveal significant regional and behavioural disparities. At a macro level, **Uttar Pradesh exhibited the highest absolute Aadhaar enrolment and biometric update activity**, reflecting its large population and administrative scale rather than necessarily indicating higher enrolment efficiency. Several Northeast states, particularly **Meghalaya, Assam, Mizoram, and Nagaland, exhibit higher proportions of late Aadhaar enrolment** and larger adult biometric update gaps, suggesting challenges related to accessibility, infrastructure, and historical trust in centralized identity systems.

Temporal analysis indicates that:

- **Demographic updates peak in March -** likely linked to exam and documentation cycles
- **Biometric updates are highest in December** - possibly reflecting year-end administrative processes.
- **Enrolment activity peaks in September** - aligning with admission and employment-related documentation needs.

A critical contribution of this study is the identification of disproportionately high **late Aadhaar enrolment (first-time enrolment at age 18+) in North-Eastern states**, where the late enrolment ratio (14.54%) is more than seven times higher than that of the rest of India (2.06%). The analysis further reveals that this phenomenon is **highly concentrated in Assam and Meghalaya**, which together account for **over 96% of late enrolments** in the region,

indicating state-specific structural and accessibility constraints rather than a uniform regional issue.

In addition, the study highlights an **Adult Biometric Update Gap**, measured as the proportion of adults enrolled but not having updated their biometrics after turning 18; this gap was most pronounced in **select Union Territories** – Lakshadweep, Dadra and Nagar Haveli and Daman and Diu, Ladakh; **Andhra Pradesh**, and **certain North-Eastern states** – Meghalaya, Sikkim and Arunachal Pradesh, suggesting potential limitations in service access, awareness, or administrative reach rather than outright non-compliance.

At a more granular level, **pincode-based analysis** using the coefficient of variation (CV) revealed significant intra-state disparities in enrolment distribution. **States such as Punjab, Uttarakhand, and Haryana showed high spatial inequality**, indicating that Aadhaar activity may be concentrated in urban or administrative centers while peripheral or rural areas remain comparatively underserved.

Overall, this project demonstrates not just technical proficiency in data handling, visualization, and statistical analysis, but also the ability to translate numbers into socially meaningful insights. The findings highlight where Aadhaar implementation is strong, where gaps persist, and where targeted outreach or infrastructure expansion could improve inclusivity and accessibility—making this analysis both analytically rigorous, practical and administratively relevant for governance and public policy.

This study demonstrates that Aadhaar usage is shaped not only by population size but also by accessibility, institutional integration, administrative policies, and societal behaviour. It also bridges data analysis with real-world policy and social implications.

## 2. Problem Statement

Aadhaar has become a foundational digital identity system in India, enabling access to welfare schemes, financial services, education, healthcare, employment, and a wide range of government services. While overall Aadhaar enrolment coverage is extensive, the timing and pattern of enrolment remain critical for ensuring equitable access and effective service delivery. In particular, **late Aadhaar enrolment—defined as first-time enrolment at or after 18 years of age—**raises concerns about delayed inclusion, administrative inefficiencies, and potential barriers to access, especially in certain regions of the country.

This study identifies and examines disproportionately **high rates of late Aadhaar enrolment in North-Eastern states**, highlighting a clear regional disparity when compared to the rest of India. The analysis further reveals that this issue is not uniformly distributed across the region but is **heavily concentrated in Assam and Meghalaya**, suggesting state-specific structural, geographical, and socio-administrative constraints rather than a generalized regional phenomenon.

Complementing this focus, the study also examines an **Adult Biometric Update Gap**, measured as the proportion of adults who were enrolled in Aadhaar but have not updated their biometrics after turning 18. This gap, observed most prominently in certain Union Territories, Andhra Pradesh and select North-Eastern states, indicates potential limitations in service accessibility, awareness, or administrative outreach rather than outright non-compliance.

Overall, this research seeks to understand how patterns of Aadhaar enrolment and update behavior vary across geography and time, with particular emphasis on late enrolment trends, age-wise enrolment behavior, temporal seasonality, and intra-state inequalities at the pincode level. By integrating these dimensions, the study aims to translate raw administrative data into actionable insights that can inform targeted awareness campaigns, infrastructure planning, and evidence-based policy interventions to enhance the effectiveness, inclusivity, and reliability of the Aadhaar ecosystem.

## 3. Objectives

The key objectives of this study are:

1. **To integrate and standardize UIDAI Aadhaar datasets** (enrolment, biometric updates, and demographic updates) into a single, analyzable master dataset while preserving maximum information.

2. **To examine regional patterns in Aadhaar enrolment and updates**, identifying states and districts with high or low activity, with particular attention to North-Eastern states.

3. **To analyze age-wise enrolment behavior**, distinguishing early enrolment (0–5 years) from **late enrolment (18+ years)** and assessing its spatial distribution.

4. **To quantify and analyze the Adult Biometric Update Gap**, identifying regions where adults are enrolled but have not updated their biometrics after turning 18, and interpreting its implications for service access.

5. **To study temporal trends** in Aadhaar enrolment, biometric updates, and demographic updates across 2025, identifying seasonal or administrative patterns.

6. **To assess intra-state inequality using pincode-level dispersion**, identifying states where Aadhaar activity is unevenly distributed within their boundaries.

7. **To translate analytical findings into actionable societal and administrative insights** that can inform targeted awareness campaigns, infrastructure planning, and evidence-based policy interventions, particularly in regions with high late enrolment.

## 4. Dataset Description

This analysis is based on three official UIDAI Aadhaar datasets provided for the hackathon:

1. **Aadhaar Enrolment Dataset -** Tracks new Aadhaar registrations.
2. **Aadhaar Biometric Update Dataset -** Tracks updates to fingerprints, iris scans, and photos.
3. **Aadhaar Demographic Update Dataset -** Tracks updates to name, address, gender, etc.

Each dataset consists of multiple CSV files, which were first merged category-wise (biometric, demographic, enrolment) before being integrated into a single master dataset.

**Common Columns Across All Datasets (Used for Merging):**

- **date** – Date of enrolment or update
- **state** – State/Union Territory
- **district** – District name
- **pincode** – Area postal code

**Additional Columns Used in Analysis:**

**Biometric Dataset:**

- bio_age_5_17 – Biometric updates for individuals aged 5–17
- bio_age_17_ – Biometric updates for individuals aged 18 and above

**Demographic Dataset:**

- demo_age_5_17 – Demographic updates for individuals aged 5–17
- demo_age_17_ – Demographic updates for individuals aged 18 and above

**Enrolment Dataset:**

- age_0_5 – Enrolments for children aged 0–5
- age_5_17 – Enrolments for individuals aged 5–17
- age_18_greater – Enrolments for individuals aged 18 and above

After cleaning and integration, the final master dataset contained approximately **2.95 million records**, covering **36 officially recognized states and Union Territories**, with data spanning from **March 1, 2025 to December 31, 2025**.

## 5. Methodology

The analysis followed a structured, reproducible, and multi-step methodology consisting of data consolidation, cleaning, transformation, and analytical exploration.

**1. Data Consolidation**

- All individual CSV files within each category (biometric, demographic, enrolment) were merged using pandas.concat().
- A **master dataset** was created by performing **full outer joins** across the three datasets using date, state, district, and pincode as common keys.
- This ensured that records present in only one dataset (e.g., enrolment but no updates) were retained.

**2. Data Cleaning and Preprocessing**

Several data quality issues were identified and addressed:

**Handling Missing Values**

- Missing numeric values were filled with **0**, assuming that absence of data indicated no enrolment or update activity on that date in that location.

**Correcting Schema Misalignment**

- A small number of records had misaligned values where numeric entries appeared in the state column (e.g., "100000").
- These records were identified using numeric checks and removed to maintain structural integrity.

**Standardizing State Names**

- The state column contained multiple inconsistencies such as:
    - Case variations (e.g., "west bengal", "WEST BENGAL")
    - Spelling errors (e.g., "West Bangal")
    - Alternative names (e.g., "Orissa" instead of "Odisha")
    - Locality names mistakenly recorded as states (e.g., "Madanapalle", "Nagpur")

The following steps were applied:

- Text normalization (lowercasing, removing special characters, standardizing spaces)
- Manual alias mapping for known variations
- **Fuzzy matching using RapidFuzz** to align near-matches with official state names
- Removal of invalid locality-level entries
- Final validation ensured exactly **36 valid states and UTs**

### 3. Feature Engineering

New analytical variables were created:

- **Total Enrolment:** age_0_5 + age_5_17 + age_18_greater
- **Total Biometric Updates:** bio_age_5_17 + bio_age_17_
- **Total Demographic Updates:** demo_age_5_17 + demo_age_17_
- **Late Enrolment Ratio:** age_18_greater / total_enrollment
- **Adult Biometric Update Gap:** age_18_greater - bio_age_17_ (clipped at zero)
- **Temporal Features:** Month extracted from the date column for trend analysis.
- **Regional Labeling**: A new variable classified each record as "North-East" or "Rest of India" based on state membership to enable direct regional comparisons.

### 4. Analytical Framework

The analysis was conducted across three dimensions:

#### A. Spatial Analysis (State, District, Pincode)

• State-wise comparison of enrolment and updates, with a focused deep dive into North-Eastern states.
• Identification of states with the highest late enrolment and biometric update gaps.
• Pincode-level dispersion using Coefficient of Variation (CV) to measure intra-state inequality and identify areas with uneven Aadhaar access.

#### B. Temporal Analysis (Monthly Trends)

• Monthly aggregation of enrolment, biometric updates, and demographic updates across 2025.
• Identification of peak months and behavioral trends, including seasonality in enrolment and updates.
• Specific examination of monthly late enrolment patterns in North-East and Assam.

#### C. Behavioral Analysis

• Comparison of early (0–5) versus late (18+) enrolment patterns.
• Assessment of adult biometric update compliance and identification of regions with potential accessibility or awareness issues.
• Concentration analysis of late enrolments within North-East, particularly in Assam and Meghalaya.

### 5. Visualization and Interpretation

• Bar charts, line plots, and dispersion graphs were used to communicate insights clearly.
• Each visualization was interpreted in the context of real-world policies, infrastructure, geography, and societal behavior.
• Visuals were designed to highlight both national trends and North-East–specific disparities.

## 6. Key Focus of the Study: Late Aadhaar Enrolment in North-East

While multiple patterns emerge from the analysis of Aadhaar enrolment and update behavior across India, the most significant and policy-relevant finding of this study is the disproportionately high incidence of late Aadhaar enrolment (first-time enrolment at age 18+) in North-Eastern states, and its extreme concentration in just two states—Meghalaya and Assam. This phenomenon has important implications for digital identity access, administrative planning, and national security considerations in sensitive border regions.

**Core Finding (Data-Driven Evidence)**

- The late enrolment ratio in the **North-East is 14.54%**, compared to only **2.06% in the rest of India**, indicating a more than sevenfold disparity.
- Within the North-East, late enrolments are **highly concentrated**, with:
    - Meghalaya: 56.60%
    - Assam: 39.30%
    - Together: ~96% of all late enrolments in the region.
- The remaining six North-Eastern states collectively contribute less than 4%.

This demonstrates that late Aadhaar enrolment is **not a uniform regional issue**, but a **state-specific challenge primarily driven by Meghalaya and Assam**.

**Interpretation (Why this is happening)**

The concentration of late enrolment in these two states suggests the influence of multiple intersecting factors, including:

- **Geographical constraints:** Hilly terrain and remote settlements limit access to fixed enrolment centers.
- **Service accessibility gaps:** Uneven distribution of enrolment infrastructure across districts and pincodes.
- **Socio-cultural trust issues:** Historical apprehensions toward centralized identity systems in certain communities.
- **Border proximity and migration dynamics:** Assam and Meghalaya share international borders with countries such as Bangladesh, Bhutan, and Myanmar, where cross-border movement and migration have historically been sensitive issues. This may have contributed to heightened scrutiny around identity documentation and delayed or irregular enrolment patterns.
- **Administrative and political context:** Assam's unique migration concerns and identity verification policies.

**Policy Context: Assam October 2025 Decision**

In **October 2025, the Government of Assam introduced a policy restricting first-time Aadhaar enrolment for individuals above 18 years of age**, except for specific categories such as Scheduled Castes (SC), Scheduled Tribes (ST), and tea plantation workers. This decision was primarily aimed at curbing **illegal immigration-related identity misuse**, particularly concerning migration from Bangladesh.

The data-driven evidence from this study—showing that Assam alone contributes **nearly 40% of North-East's late enrolments**—provides empirical support for the rationale behind this policy intervention. It highlights that Assam had a disproportionately high burden of late enrolments, making it a focal point for administrative scrutiny and reform.

**Implication (Why this should guide action)**

Given this concentration, policy responses should be **state-targeted rather than region-wide**, with priority given to Meghalaya and Assam. Suggested strategic directions include:

- Expanding **mobile Aadhaar enrolment units** in remote districts.
- Strengthening **last-mile digital identity infrastructure**.
- Conducting **localized awareness programs** in local languages.
- Aligning enrolment drives with school and community institutions to reduce future late enrolment.

This focused approach is likely to yield higher impact than blanket regional interventions.

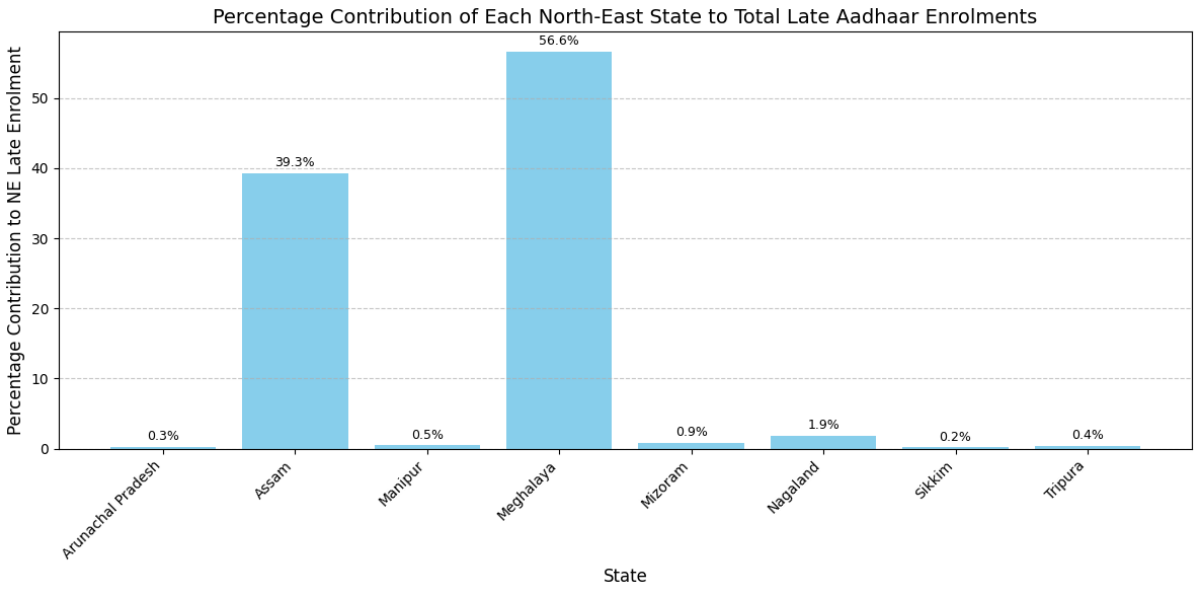


**Figure 6.0:** Percentage contribution of each North-East State to total late Aadhaar enrolments

# 7. Key Insights & Analysis

## 7.1 Descriptive Overview of Enrolment and Update Activity

Two baseline visualizations were included to provide a high-level overview of Aadhaar enrolment and biometric update activity across states. These charts serve as a reference point for interpreting subsequent analyses on late enrolment, update gaps, and regional disparities.

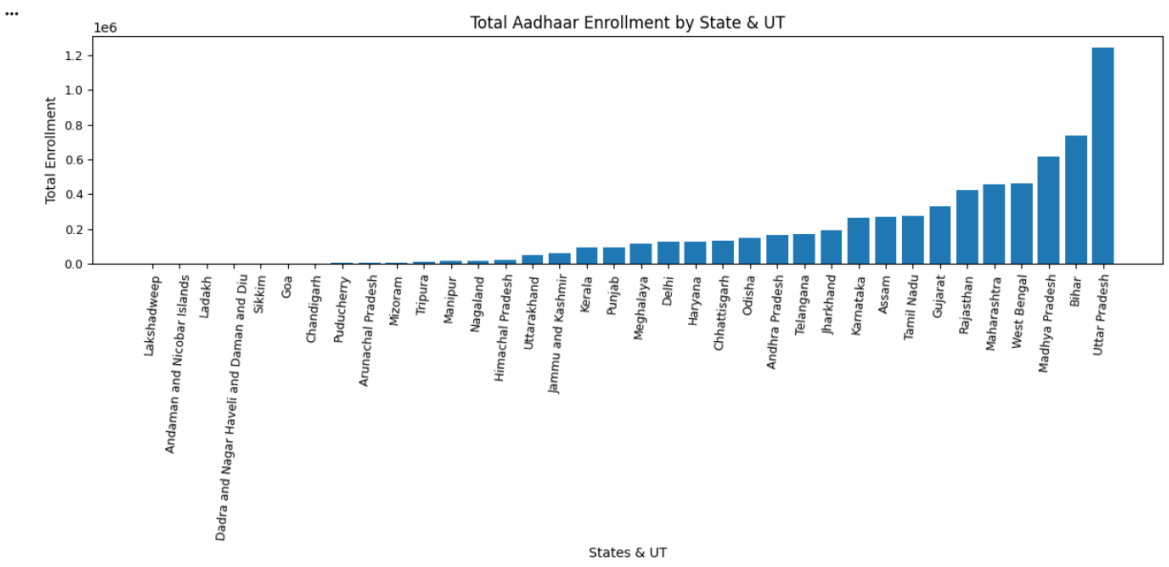


**Figure 7.1.1:** State-wise distribution of total Aadhaar enrolments in 2025, illustrating regional variations in enrolment scale and administrative activity.

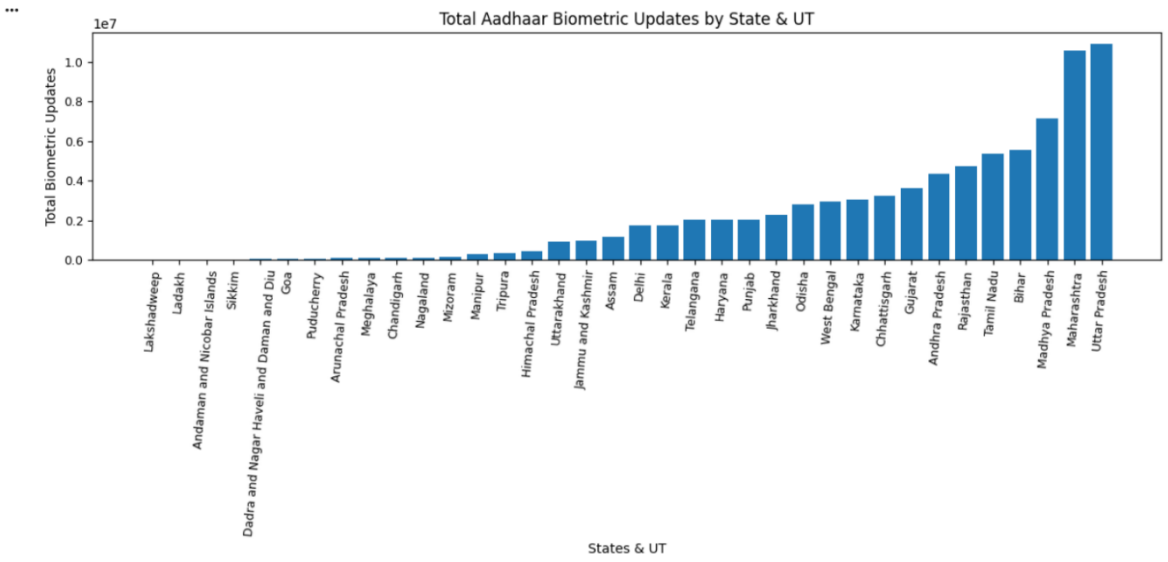


**Figure 7.1.2:** State-wise distribution of Aadhaar biometric updates in 2025, indicating relative levels of update activity across regions.

- State-wise Aadhaar enrolment in 2025, showing that absolute enrolment levels are largely driven by population size and administrative reach, rather than enrolment efficiency.
- State-wise Aadhaar biometric updates in 2025, highlighting regional differences in update behavior and forming the basis for identifying adult biometric update gaps.

### 7.2 Adult Biometric Update Gap – "Who is enrolled but not updated?"

**The Question:**
To what extent are adults (18+) enrolled in Aadhaar but yet to update their biometrics, and how does this vary across states and districts?

**The Finding:**
A measurable proportion of adults were found to be enrolled in Aadhaar but had not updated their biometrics after turning 18. This gap was particularly pronounced in certain Union Territories, Andhra Pradesh and few North-Eastern states, like Meghalaya, Sikkim, and Arunachal Pradesh ranking among the highest in the "adult not updated" ratio.

**The Interpretation:**
This gap does not necessarily indicate inadequate Aadhaar coverage but rather a delay or barrier in biometric updating. Contributing factors may include limited access to enrolment centres, geographic constraints in hilly or remote regions, low awareness about mandatory biometric updates after 18, or historical mistrust of centralized identity systems.

**The Implication:**
Targeted awareness campaigns, mobile Aadhaar update units, and region-specific outreach programs—especially in the North-East and high-gap districts—could significantly reduce authentication failures and improve service accessibility.

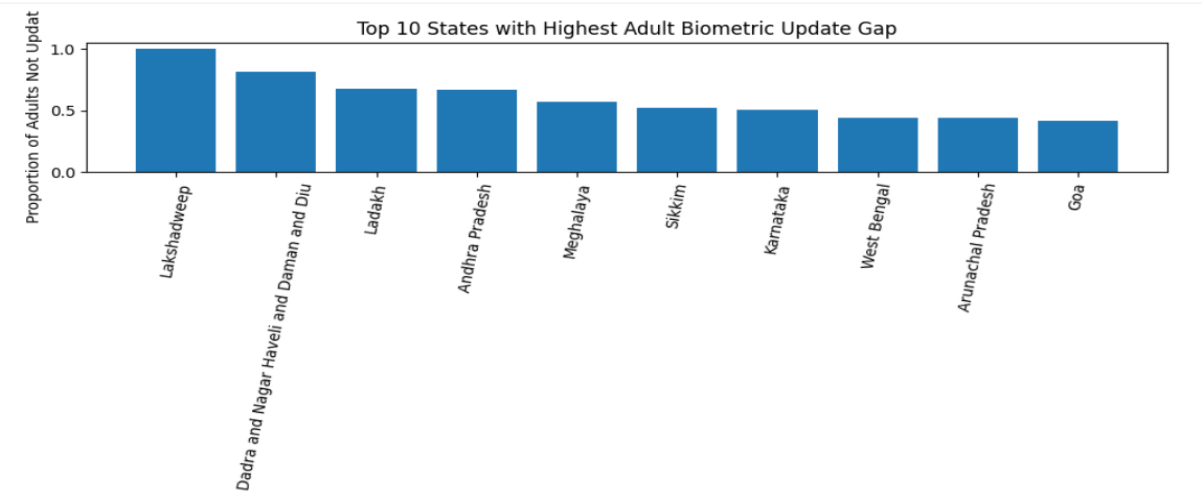


**Figure 7.2:** Adult Biometric Update Gap

- **Highest gap between Aadhar enrollment and biometric update meaning - people have enrolled in larger quantities, but had not updated their biometrics after reaching age 18; Which is necessary.**

### 7.3 Age-wise Aadhaar Enrolment – "At what stage of life are people enrolling?"

**The Question:**
How does Aadhaar enrolment vary across different age groups (0–5, 5–17, and 18+), and what does this reveal about enrolment behavior?

**The Finding:**
Significant variation was observed in age-wise enrolment across states. Some states exhibited higher enrolment among children (0–5), while others showed greater enrolment among adults (18+). Uttar Pradesh had the highest absolute enrolment across all age groups, reflecting its population size and administrative scale.

**The Interpretation:**
States with high 0–5 enrolment indicates early adoption and greater institutional integration of Aadhaar (e.g., in schools, healthcare, and welfare schemes). Conversely, states with higher 18+ enrolment suggest delayed initial registration rather than lack of Aadhaar coverage.

**The Implication:**
States with high late enrolment may benefit from proactive enrolment initiatives targeting young adults, particularly in education, employment, and social welfare contexts.

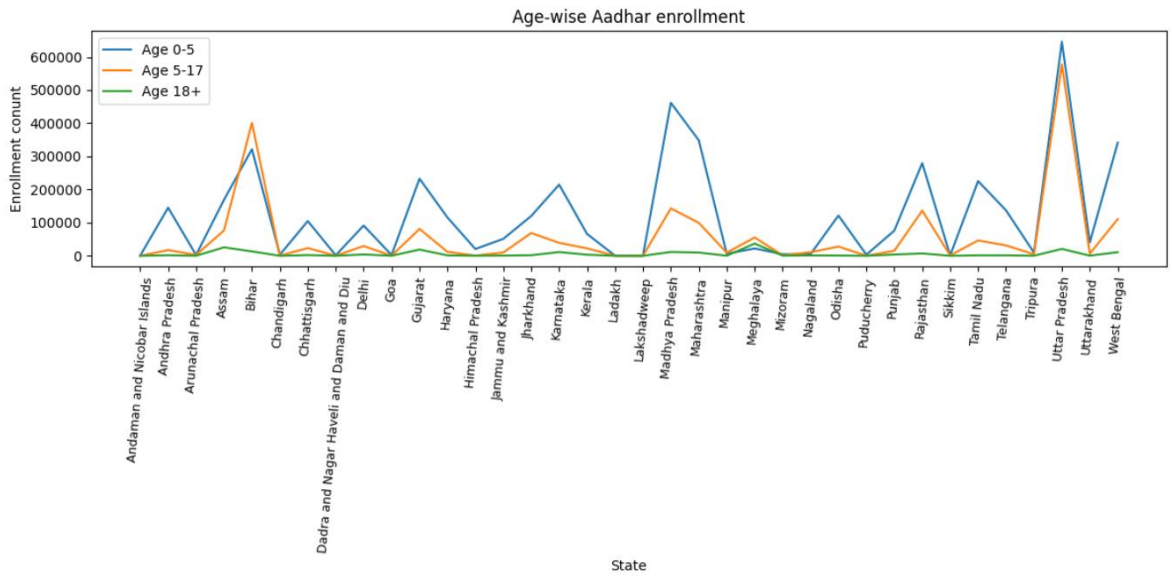


**Figure 7.3:** Age-wise Aadhaar Enrolment

- Blue line highest peak at Uttar Pradesh indicates early Aadhar enrollment i.e. Aadhar enrollment at age 0-5 years is more.
- Green line highest peak at Meghalaya indicates late Aadhar enrollment, i.e Aadhar enrollment after 18 years is more.

This suggests differences in timing of enrolment across states rather than differences in total coverage.

### 7.4 Late Aadhaar Enrolment — "Why are people enrolling after 18?"

**The Question:**
Which states exhibit the highest proportion of Aadhaar enrolment among adults (18+), and what factors might explain this pattern?

**The Finding:**
North-Eastern states — notably Meghalaya, Assam, Mizoram, and Nagaland — show the highest proportions of late Aadhaar enrolment (ages 18 and above). These values stand out prominently compared to most other Indian states.

**The Interpretation:**
High late-enrolment proportions do not necessarily imply poor overall Aadhaar coverage. Rather, they suggest that initial registration is often postponed until adulthood. Multiple structural, administrative, and policy factors contribute to this pattern:

- **Geographical and accessibility constraints:** Hilly terrain and dispersed rural settlements in many North-Eastern regions can limit physical access to enrolment centers, leading individuals to delay enrolment until it becomes essential for services such as banking, education, or employment.
- **Weaker early institutional integration:** In states where Aadhaar is less systematically embedded into early-life touchpoints like birth registration, healthcare, or schooling, enrolments naturally skew toward older age groups.
- **Complex demographic environments and verification protocols:** Some border regions experience complex population mobility dynamics and stricter identity verification procedures, which can lengthen the enrolment process and delay first-time Aadhaar registration.
- **Policy Response in Assam:** In August 2025, the Government of Assam announced that *first-time Aadhaar enrolment for adults (18+)* would be restricted from October 2025, with limited exceptions for Scheduled Castes, Scheduled Tribes, and tea garden communities, as part of efforts to enhance verification standards and address concerns around fraudulent enlistments and undocumented migration. This policy context is relevant when interpreting late enrolment patterns in Assam and neighboring states.

**The Implication:**
To mitigate late enrolment patterns and promote timely Aadhaar adoption, it is important to improve accessibility — especially in geographically challenging regions — and integrate enrolment more thoroughly into early-life institutional processes (e.g., school, healthcare, and childhood welfare registrations). In policy contexts like Assam, balancing robust verification with equitable access remains critical to ensure legitimate residents can obtain Aadhaar in a timely and transparent manner.

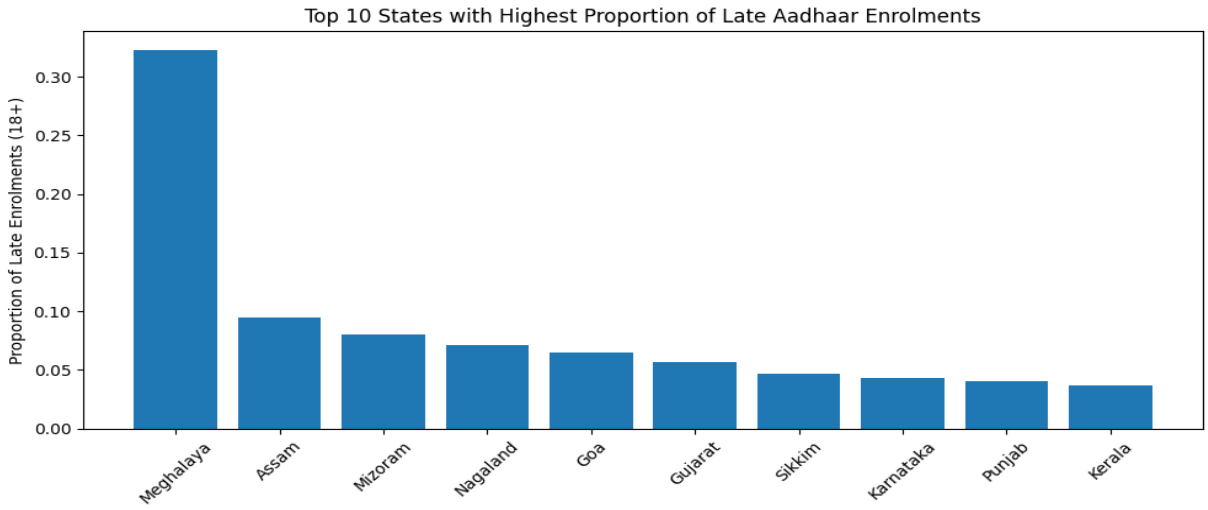


**Figure 7.4:** Late Aadhaar Enrolment

### 7.5 Early Childhood Enrolment – "Are children being enrolled on time?"

**The Question:**
Which states have the highest Aadhaar enrolment among children aged 0–5, and what does this suggest about awareness?

**The Finding:**
Uttar Pradesh exhibited the highest number of 0–5 Aadhaar enrolments, followed by several other populous states. States with strong institutional linkages (e.g., healthcare and school-based enrolment) showed higher early childhood enrolment.

**The Interpretation:**
High early enrolment reflects better administrative integration of Aadhaar into birth registration, healthcare, and welfare programs. It also indicates greater parental awareness of Aadhaar's importance.

**The Implication:**
Best practices from high-performing states could be replicated in low-performing regions to improve early Aadhaar adoption nationwide.

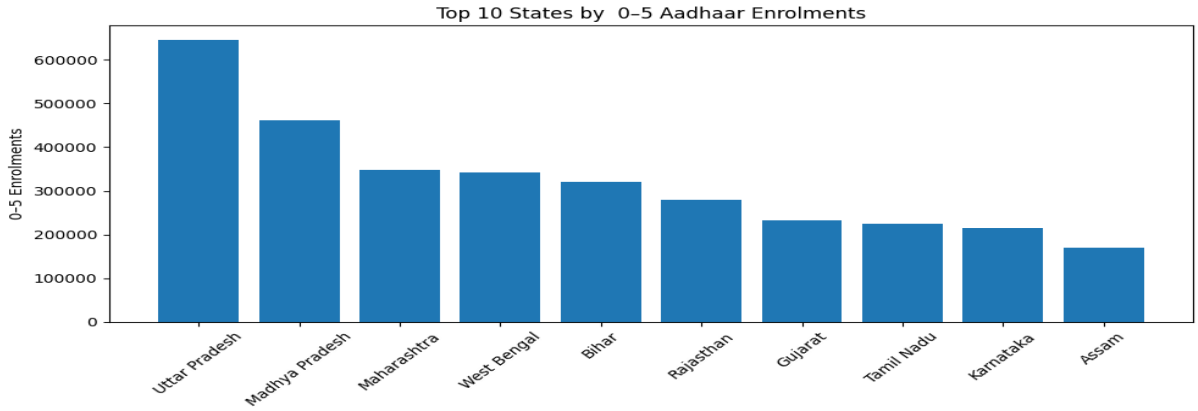


**Figure 7.5:** Early Childhood Enrolment

### 7.6 Intra-State Inequality (Pincode Analysis) – "Is Aadhaar activity evenly distributed within states?"

**The Question:**
How evenly is Aadhaar enrolment distributed across pincodes within each state?

**The Finding:**
Using the Coefficient of Variation (CV), significant intra-state inequality was observed in Aadhaar enrolment distribution. States such as Punjab, Uttarakhand, and Haryana exhibited high dispersion, indicating uneven activity across pincodes.

**The Interpretation:**
High dispersion suggests that Aadhaar services are concentrated in certain urban or administrative hubs, while rural or remote areas may have comparatively lower access or participation.

**The Implication:**
Policymakers should prioritize decentralized Aadhaar infrastructure, such as rural enrolment centers and mobile units, to ensure equitable access across all regions within a state.

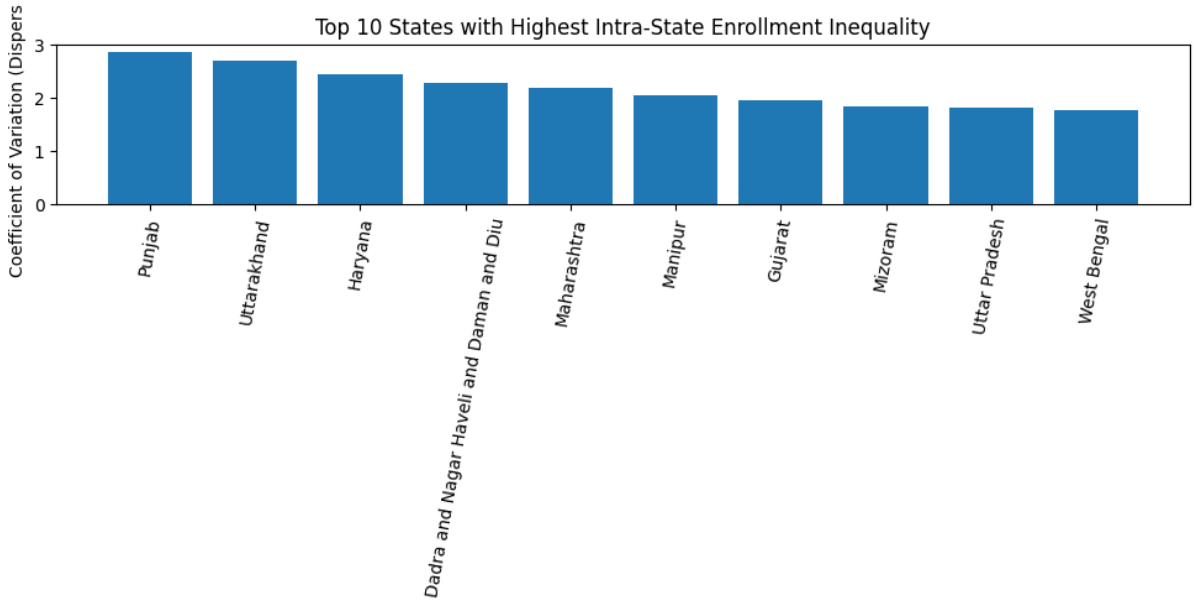


**Figure 7.6:** Intra-State Inequality

### 7.7 Temporal Trends – "When do enrolments and updates peak?"

**The Question:**
How do Aadhaar enrolment, biometric updates, and demographic updates vary over time in 2025?

**The Finding:**

- **Highest enrolments:** September 2025
- **Highest biometric updates:** December 2025
- **Highest demographic updates:** March 2025, followed by a decline and then a resurgence toward September

**The Interpretation:**

- The September enrolment peak likely aligns with admission cycles, employment documentation, and government deadlines.
- The December biometric peak may reflect year-end administrative compliance requirements.
- The March demographic spike suggests links to academic applications (e.g., JEE, NEET, CET), documentation renewals, or welfare scheme deadlines which require accurate Aadhaar details..

**The Implication:**
UIDAI could anticipate demand surges in these months and allocate additional resources accordingly—such as extended service hours, mobile camps, and digital awareness drives—to reduce service bottlenecks.

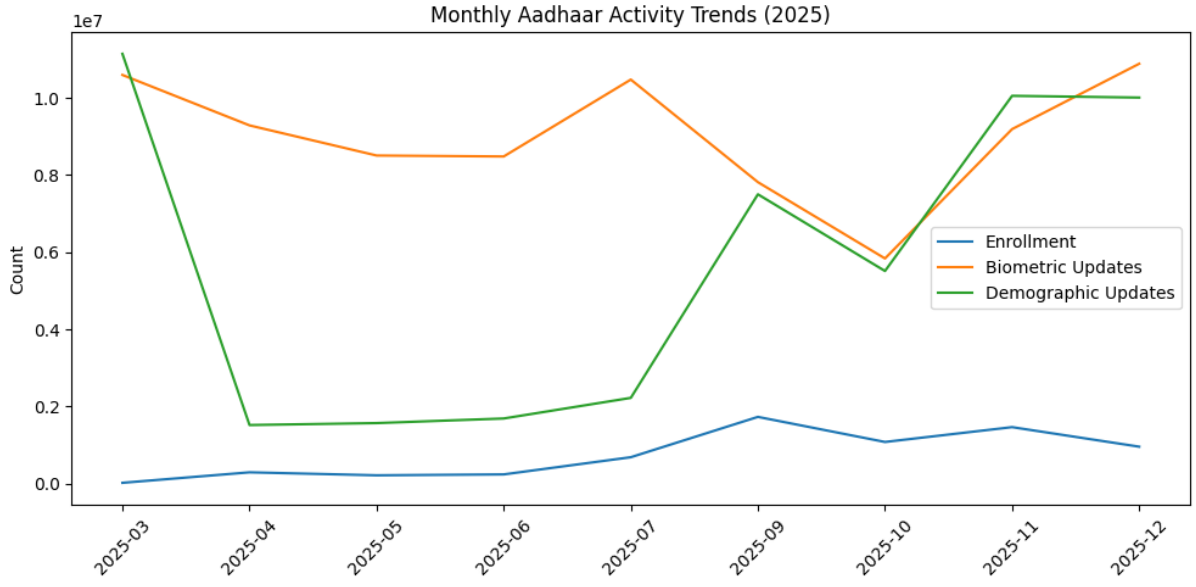


**Figure 7.7:** Temporal Trends

## 8. Impact & Recommendations

The findings of this study have direct implications for policy design, administrative planning, and service delivery within the Aadhaar ecosystem. The evidence—particularly the disproportionately high late Aadhaar enrolment in the North-East (14.54%) compared to the rest of India (2.06%), and the concentration of this phenomenon in Meghalaya and Assam (~96% of North-East late enrolments)—indicates that the challenge is not merely awareness-based but deeply linked to historical, administrative, and geopolitical factors. Accordingly, the recommendations below integrate both data-driven insights and existing policy realities, rather than treating the issue as purely technical or logistical.

### 8.1 Understanding the Drivers of Late Enrolment (Policy Context)

#### (a) Assam – NRC Legacy and Strict Vetting Framework

A major contributor to late enrolment in Assam stems from the 2019 National Register of Citizens (NRC) exercise, during which the biometrics of over 27 lakh individuals were frozen due to pending citizenship verification. This effectively prevented many genuine residents from obtaining Aadhaar for several years, creating a significant backlog of first-time adult applicants in 2025.

In response, the Assam government introduced a more stringent Standard Operating Procedure (SOP) in late 2025, requiring first-time adult Aadhaar applicants to undergo district-level vetting through Block Development Officers (BDOs) or District Commissioners (DCs). While this measure strengthens national security and prevents fraudulent enrolments, it also introduces additional administrative layers that slow down processing and classify many genuine applicants as "late enrollers."

Implication: Policy interventions in Assam must balance security with accessibility by ensuring that verification mechanisms do not unintentionally exclude eligible residents.

#### (b) Meghalaya – Historical Resistance and Gradual Adoption

In Meghalaya, late enrolment is largely socio-cultural rather than administrative. Several tribal organizations historically resisted Aadhaar, fearing misuse by illegal immigrants or potential threats to tribal land rights and identity protections. As Aadhaar became increasingly necessary for banking, welfare schemes, and government services, many adults who had previously opted out began enrolling, leading to a concentrated wave of late enrolments.

However, Meghalaya has also demonstrated best practice through its decentralized digital verification model, where adult enrolment requests are routed to district officials via a state portal, enabling faster approval for genuine residents while filtering suspicious cases.

Implication: Trust-building and community engagement are as critical as infrastructure in reducing late enrolment.

### 8.2 Targeted Policy and Administrative Interventions

#### (a) Mandatory District-Level Vetting with Efficiency Improvements (Assam Model)

- Maintain district-level vetting for first-time adult enrolments in sensitive border regions to prevent identity misuse.
- Simultaneously, streamline procedures by introducing digital workflows, standardized documentation checklists, and time-bound approvals to reduce delays for genuine applicants.
- Establish grievance redressal mechanisms for applicants facing repeated rejections or processing delays.

#### (b) Scaling the Meghalaya Decentralized Approval System

- Replicate Meghalaya's portal-based district verification model in other North-Eastern states with high late enrolment ratios.
- Automate status tracking so applicants can monitor progress rather than relying on manual follow-ups.
- Integrate UIDAI systems with state verification portals to minimize duplication and bureaucratic friction.

#### (c) Special Enrolment Window Periods (Strategic Use)

- Continue periodic, clearly communicated enrolment windows for first-time adult applicants in high-backlog districts, as implemented in Assam.
- Pair these windows with temporary increases in staffing and mobile enrolment units to prevent system overload.
- Use data analytics to identify districts with the highest backlog and prioritize them during window periods.

#### (d) Protection for Marginalized and Vulnerable Groups

- Extend flexible enrolment timelines for Tea Tribes, Scheduled Castes (SC), Scheduled Tribes (ST), elderly individuals, and persons with disabilities in Assam and Meghalaya.
- Provide assisted enrolment support through community volunteers, NGOs, and local administrative offices.

### 8.3 Infrastructure & Last-Mile Access (Operational Solutions)

**Proposed Flagship Initiative: "Aadhaar-on-Wheels for North-East Border Districts"**

To directly address access barriers identified in this study, we propose a targeted mobile enrolment program:

- Deploy Aadhaar-on-Wheels vans equipped with biometric capture and demographic update facilities in remote, hilly, and border-adjacent districts of Meghalaya and Assam.
- Operate on a fixed, publicly communicated schedule so residents can plan accordingly.
- Partner with local administrations, village councils, and community leaders to improve participation and trust.
- Use real-time demand data to dynamically optimize routes and resource allocation.

This initiative would reduce travel burden, minimize administrative friction, and significantly lower late enrolment rates in high-risk regions.

### 8.4 Demand-Based Service Expansion

- Increase permanent Aadhaar enrolment and update centers in districts with consistently high late enrolment ratios.
- Implement demand-based staffing during peak months (March, September, and December), as identified in the temporal analysis.
- Decentralize services at the pincode level in states with high intra-state dispersion to reduce regional inequalities in access.

### 8.5 Awareness, Trust-Building, and Behavioral Interventions

- Conduct localized, multilingual awareness campaigns in Meghalaya and Assam explaining the benefits of timely Aadhaar enrolment and biometric updates.
- Integrate Aadhaar awareness into schools, colleges, and community institutions to encourage early-life enrolment and reduce dependence on late enrolment.
- Introduce automated SMS/WhatsApp reminders for individuals approaching 18 years of age to update their biometrics.

### 8.6 Data-Driven Monitoring and Accountability

- Develop state-level dashboards tracking late enrolment ratios, biometric update gaps, and service availability over time.
- Use district-level data to prioritize resources where the gap is largest rather than applying uniform policies.
- Conduct periodic audits of enrolment backlogs to assess the effectiveness of interventions.

## 9. Limitations

While this study provides valuable insights, it is important to acknowledge certain limitations that should be considered when interpreting the results:

1. **Temporal Scope:** The analysis is based solely on Aadhaar data from 2025 and may not capture long-term trends or historical variations in enrolment behavior.
2. **Lack of Population Baseline:** The study does not incorporate external population data to normalize enrolment rates per capita across states or districts, which could refine regional comparisons.
3. **Urban–Rural Classification:** The dataset does not explicitly distinguish between urban and rural pincodes, limiting direct assessment of rural–urban disparities.
4. **Observational Nature:** The analysis identifies correlations and patterns but does not establish causal relationships between policy, infrastructure, and enrolment behavior.
5. **Policy Interpretation:** Certain interpretations—such as the impact of Assam's October 2025 policy—are based on publicly reported information rather than direct administrative records.

Future research could address these limitations by incorporating multi-year data, demographic baselines, and more granular geographic classifications.

## 10.Conclusion

This study presents a rigorous, data-driven analysis of Aadhaar enrolment and update behavior across India in 2025, with a central focus on late Aadhaar enrolment in the North-Eastern states. By integrating UIDAI's biometric, demographic, and enrolment datasets into a unified master dataset, the analysis uncovers meaningful spatial, temporal, and behavioral patterns with direct policy relevance.

A key contribution of this research is the identification of a significant regional disparity: the North-East shows a 14.54% late enrolment ratio compared to 2.06% in the rest of India, with Meghalaya and Assam alone accounting for approximately 96% of late enrolments in the region. This concentration indicates that the issue is not uniformly regional but shaped by state-specific historical, geographical, and administrative factors. The study situates these findings within real-world contexts such as the legacy of the 2019 NRC in Assam, the subsequent biometric freeze and gradual unfreezing, the October 2025 district-level vetting framework, and Meghalaya's decentralized verification model.

Complementary analyses further strengthen the study. The Adult Biometric Update Gap highlights the need for proactive update reminders, while age-wise enrolment patterns show that stronger early childhood enrolment aligns with better long-term Aadhaar adoption. Temporal trends reveal peaks in March, September, and December, and pincode-level dispersion analysis exposes significant intra-state inequalities in service access.

Building on these insights, the study proposes targeted, data-driven interventions, including: streamlined district-level vetting in Assam, scaling Meghalaya's digital approval model, strategic enrolment windows, safeguards for marginalized groups, demand-based staffing during peak months, and a flagship "Aadhaar-on-Wheels" initiative for remote border districts.

Overall, this research demonstrates how integrated data analysis can guide smarter governance—balancing security, accessibility, and inclusivity—to strengthen the reliability and equity of India's Aadhaar ecosystem.

## 11. Tools & Technologies Used

The following tools, libraries, and platforms were used in this study to ensure reproducibility, efficiency, and analytical rigor:

All analysis was conducted in Python using reproducible notebooks.

**Programming Language and Environment**

- **Python 3.10+** – Primary programming language for data processing and analysis
- **Google Colab** – Cloud-based execution environment for large-scale data handling and visualization

**Data Processing and Analysis Libraries**

- **Pandas** – Data cleaning, merging, transformation, and aggregation
- **NumPy** – Numerical operations and handling missing values
- **RapidFuzz** – Fuzzy string matching for state name standardization

**Data Visualization**

- **Matplotlib** – Creation of bar charts, line plots, and trend visualizations

**Data Storage and Format**

- **CSV (Comma-Separated Values)** – Primary format of UIDAI datasets
- **Google Drive** – Storage and retrieval of large datasets

## 12. Appendix

### Appendix A: Metrics, Formulas, and Definitions

This appendix provides mathematical definitions and explanations of key metrics used in the analysis.

#### A.1 Total Enrolment

This metric represents the total number of Aadhaar enrolments recorded in a given state, district, or pincode.

$$\text{Total Enrolment} = \text{age\_0\_5} + \text{age\_5\_17} + \text{age\_18\_greater}$$

#### A.2 Total Biometric Updates

This measures the total number of biometric updates recorded.

$$\text{Total Biometric Updates} = \text{bio\_age\_5\_17} + \text{bio\_age\_17\_}$$

#### A.3 Total Demographic Updates

This captures all demographic corrections or updates.

$$\text{Total Demographic Updates} = \text{demo\_age\_5\_17} + \text{demo\_age\_17\_}$$

#### A.4 Late Aadhaar Enrolment Ratio

This measures the proportion of adults (18+) among total enrolments.

$$\text{Late Enrolment Ratio} = \text{age\_18\_greater} / \text{Total Enrolment}$$

**Interpretation:**
A higher value suggests a greater share of enrolments occurring in adulthood rather than childhood.

#### A.5 Adult Biometric Update Gap

This measures the number of adults enrolled in Aadhaar who have not updated their biometrics after turning 18.

$$\text{Adult Not Updated} = \max(0, \text{age\_18\_greater} - \text{bio\_age\_17\_})$$

**Interpretation:**
This is not a measure of non-enrolment, but rather a potential **update gap**.

### A.6 Not Updated Ratio

This represents the proportion of adults who have not updated biometrics.

$$\text{Not Updated Ratio} = \text{Adult Not Updated} / \text{age\_18\_greater}$$

### A.7 Pincode-Level Dispersion (Coefficient of Variation - CV)

Used to measure intra-state inequality in enrolment distribution across pincodes.

$$\text{CV} = \frac{\sigma}{\mu}$$

Where:

- $\sigma$= Standard deviation of enrolments across pincodes
- $\mu$= Mean enrolment across pincodes

**Interpretation:**
Higher CV → greater inequality in enrolment distribution within a state.

### A.8 Temporal Aggregation

For monthly analysis, the date column was converted to datetime format and grouped by month:

$$\text{Month} = \text{Extracted from Date}$$

Enrolment, biometric updates, and demographic updates were then aggregated per month.

### A.9 Regional Classification (North-East vs Rest of India)

For comparative analysis, each record was classified into one of two regions based on the state:

$$\text{Region} = \begin{cases} \text{North-East} & \text{if state} \in \{\text{Assam, Arunachal Pradesh, Meghalaya, Mizoram, Manipur, Nagaland, Tripura, Sikkim}\} \\ \text{Rest of India} & \text{otherwise} \end{cases}$$

This classification enabled direct comparison of late enrolment behavior between the North-East and the rest of the country.

### A.10 North-East Late Enrolment Ratio

This metric measures the proportion of late enrolments in the North-East relative to total enrolments in that region:

$$\text{NE Late Enrolment Ratio} = \frac{\sum \text{age\_18\_greater (NE)}}{\sum \text{Total Enrolment (NE)}}$$

Similarly, for Rest of India:

$$\text{ROI Late Enrolment Ratio} = \frac{\sum \text{age\_18\_greater (ROI)}}{\sum \text{Total Enrolment (ROI)}}$$

**Interpretation:**
A significantly higher ratio in the North-East indicates delayed Aadhaar inclusion compared to the national average.

### A.11 State Contribution to North-East Late Enrolment

To assess concentration within the North-East, the percentage contribution of each state was computed as:

$$\text{State Share (\%)} = \frac{\text{Late Enrolments in State}}{\text{Total Late Enrolments in North-East}} \times 100$$

This revealed that Meghalaya and Assam together account for approximately **96%** of total late enrolments in the North-East.

### A.12 Cumulative Concentration (Top-2 States)

To quantify concentration, the cumulative share of the top two states was calculated as:

$$\text{Cumulative Share (Top-2)} = \sum \text{State Share of Assam and Meghalaya}$$

This metric helps determine whether the issue is widespread or concentrated in a few states.

### Appendix B: Data Cleaning Summary

- Missing numeric values filled with 0
- Misaligned records (numeric values in state) removed
- State names standardized using normalization and fuzzy matching
- Locality-level names mistakenly recorded as states were removed

- Final dataset validated to include 36 official states and Union Territories

**Appendix C: Key Assumptions**

1. Missing numeric values were interpreted as **no recorded enrolment/update activity**, rather than true zero demand.
2. Late enrolment was defined strictly as **first-time enrolment at or after 18 years of age** based on available age categories.
3. The North-East classification follows official geographic grouping of eight states.
4. Analysis is descriptive and observational; causality is not inferred from correlations.

**Appendix D: Code Notebook (Google Colab)**

https://colab.research.google.com/drive/11nhjIoosNUzZc5XYpxJGaxYF2Q2M81Wd?usp=sharing