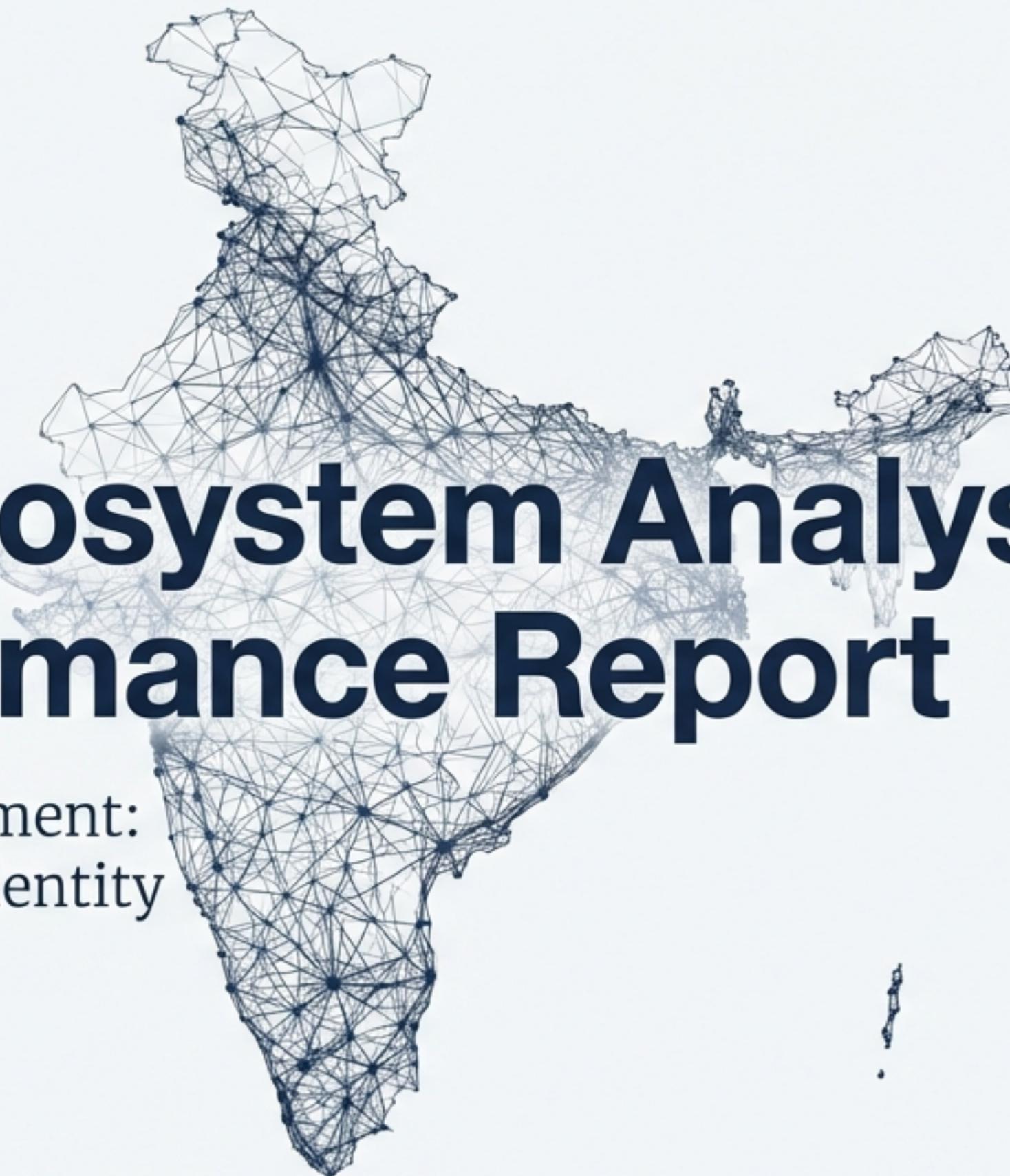


# Aadhaar Ecosystem Analysis: 2025 Performance Report

From Acquisition to Sustainment:  
Insights from 54.6 Million Identity  
Records



Analysis Period: January – December 2025. Data Source: UIDAI API Logs (Enrollment, Demographic, Biometric)

[https://github.com/predictor450v/adhar\\_analysis\\_hackathon\\_UIDAI/blob/main/analysis.ipynb](https://github.com/predictor450v/adhar_analysis_hackathon_UIDAI/blob/main/analysis.ipynb)

# *Problem Statement*

As the Aadhaar ecosystem approaches near-universal coverage, the nature of its operational challenges has fundamentally changed. Traditional analyses that focus primarily on new enrollments no longer reflect the dominant workload of the system. Instead, Aadhaar now operates as a mature digital identity infrastructure where the majority of activity is driven by demographic updates, biometric verification, and lifecycle maintenance.

The absence of a consolidated, data-driven understanding of this transition limits effective policy planning, infrastructure scaling, and regional resource allocation. This project addresses this gap by analyzing large-scale Aadhaar enrollment, demographic update, and biometric update datasets from 2015 to quantify system maturity, identify lifecycle-driven usage patterns, and derive actionable insights for managing Aadhaar as a long-term national digital public infrastructure.

Code link(git hub)

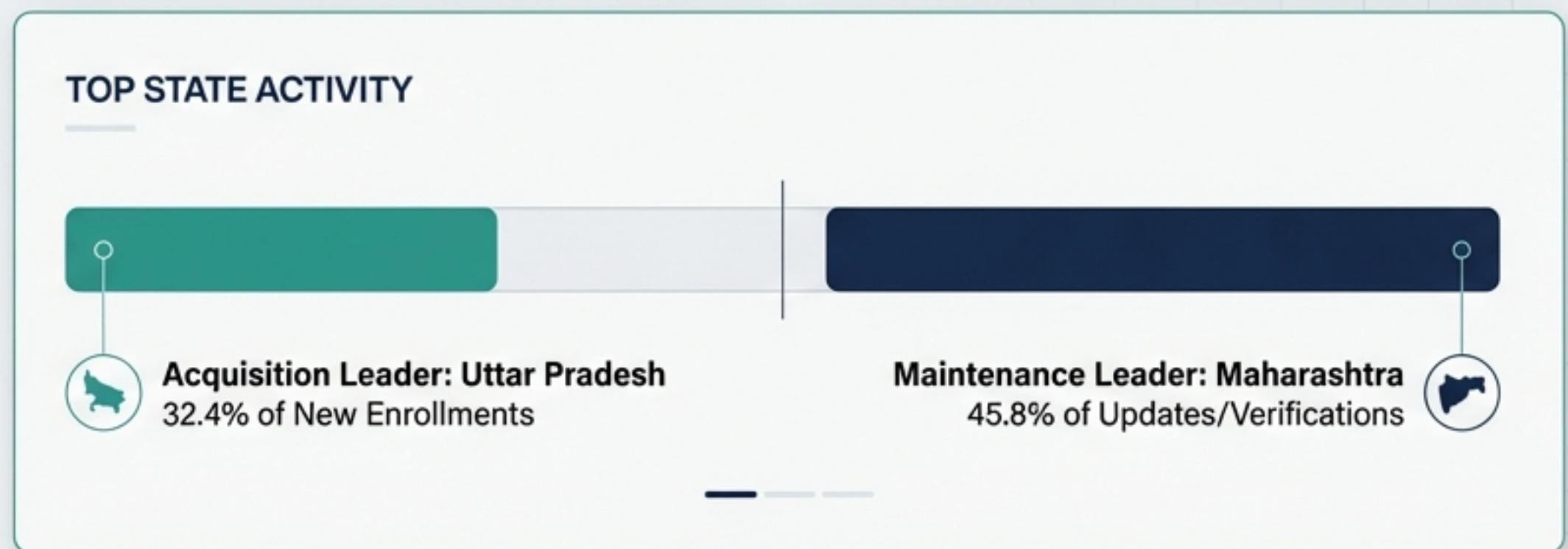
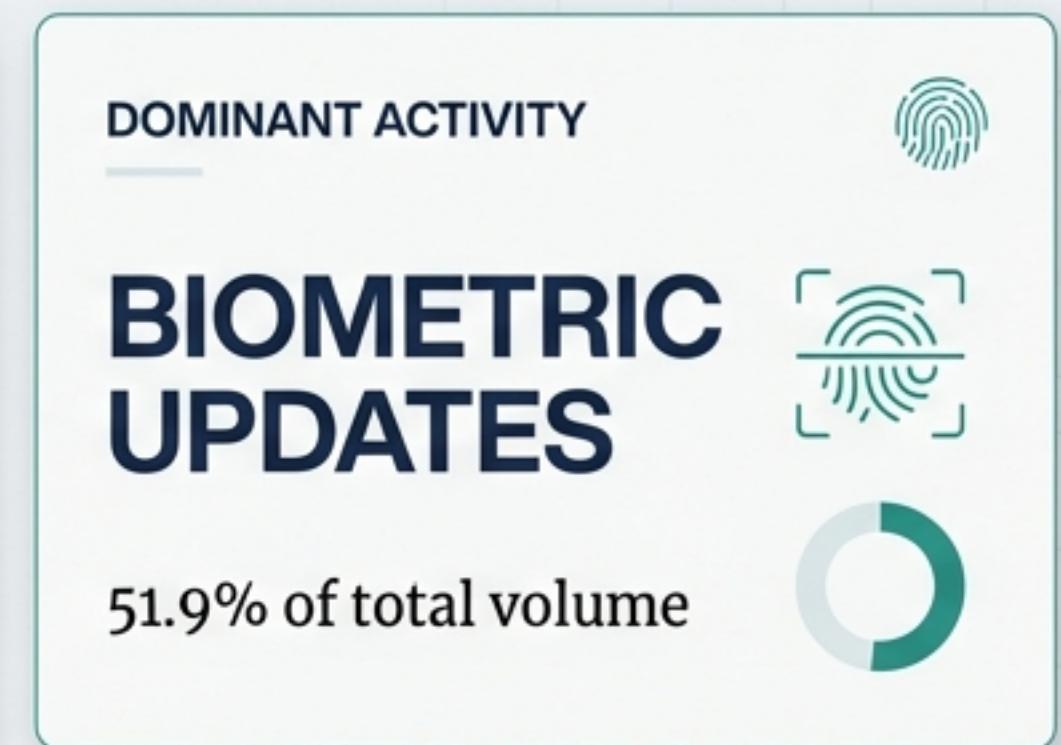
# The Defining Trend of 2025 is System Maturity

The Aadhaar ecosystem has transitioned from an ‘acquisition phase’ to a ‘sustainment phase’.

For every single new enrollment in 2025, the system processed nearly 20 identity updates.

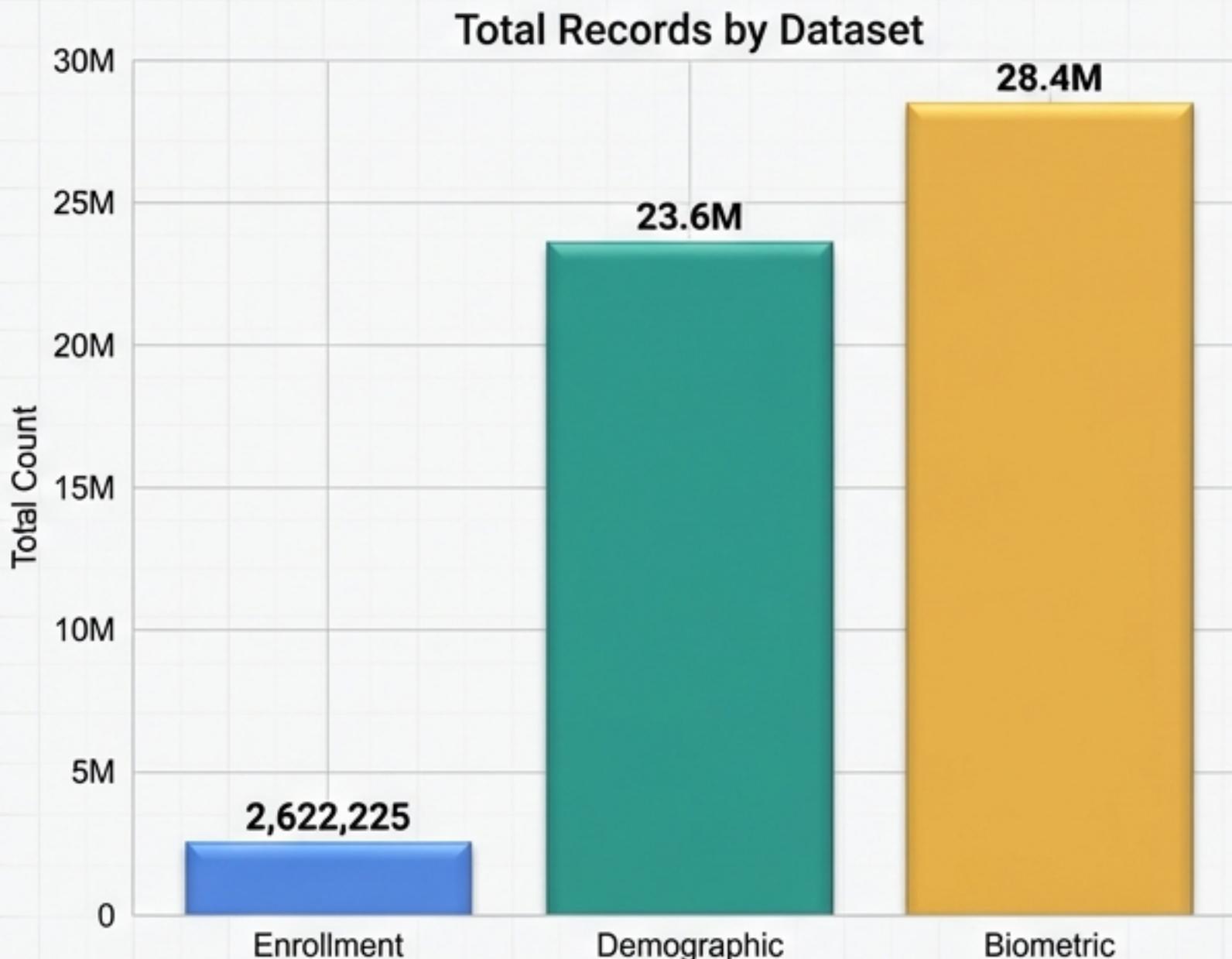
- New enrollments account for less than 5% of total system activity.
- The infrastructure operational load is now overwhelmingly driven by verification (Biometrics) and data correction (Demographics), requiring a pivot in resource allocation strategy.

## Key Statistics



# Operational Load is Driven by Identity Verification, Not Creation

Volume distribution across the three primary datasets (Jan-Dec 2025).



## Biometric Updates: 28.4 Million

The heavy lifter. Driven by mandatory age-based updates and authentication security.



## Demographic Updates: 23.6 Million

The maintenance layer. Driven by address and name corrections for KYC.

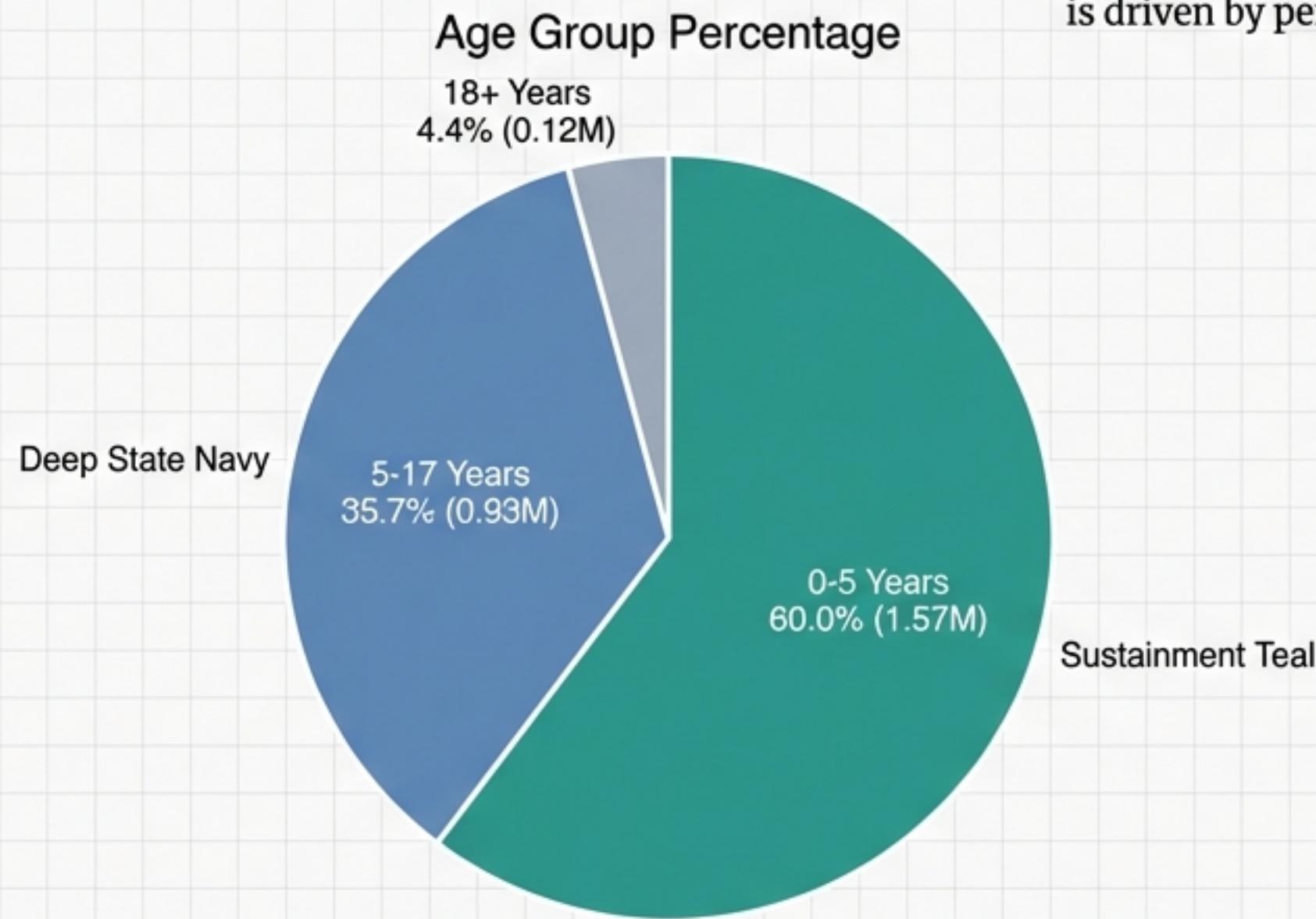


## Enrollments: 2.6 Million

The entry point. Now a fraction of the total operational volume.

# Enrollment Infrastructure Now Functions Primarily as a Birth Registry

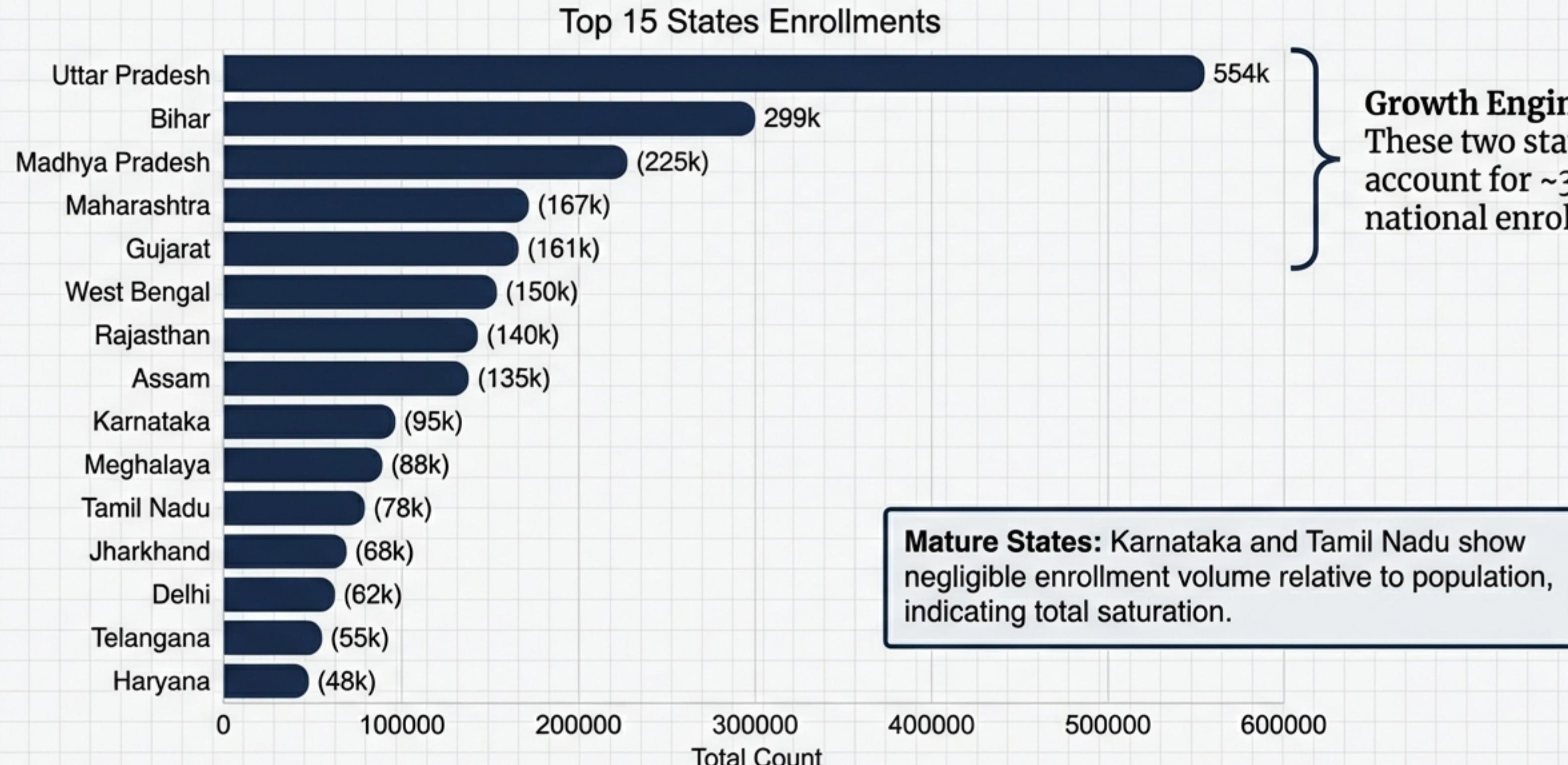
Adult saturation is near-total. The residual adult enrollment represents edge cases, while the core intake is driven by peri-natal and early childhood registration.



**Key Insight:** 60% of all new enrollments (1.57M) are children aged 0-5. Only 4.4% are adults.

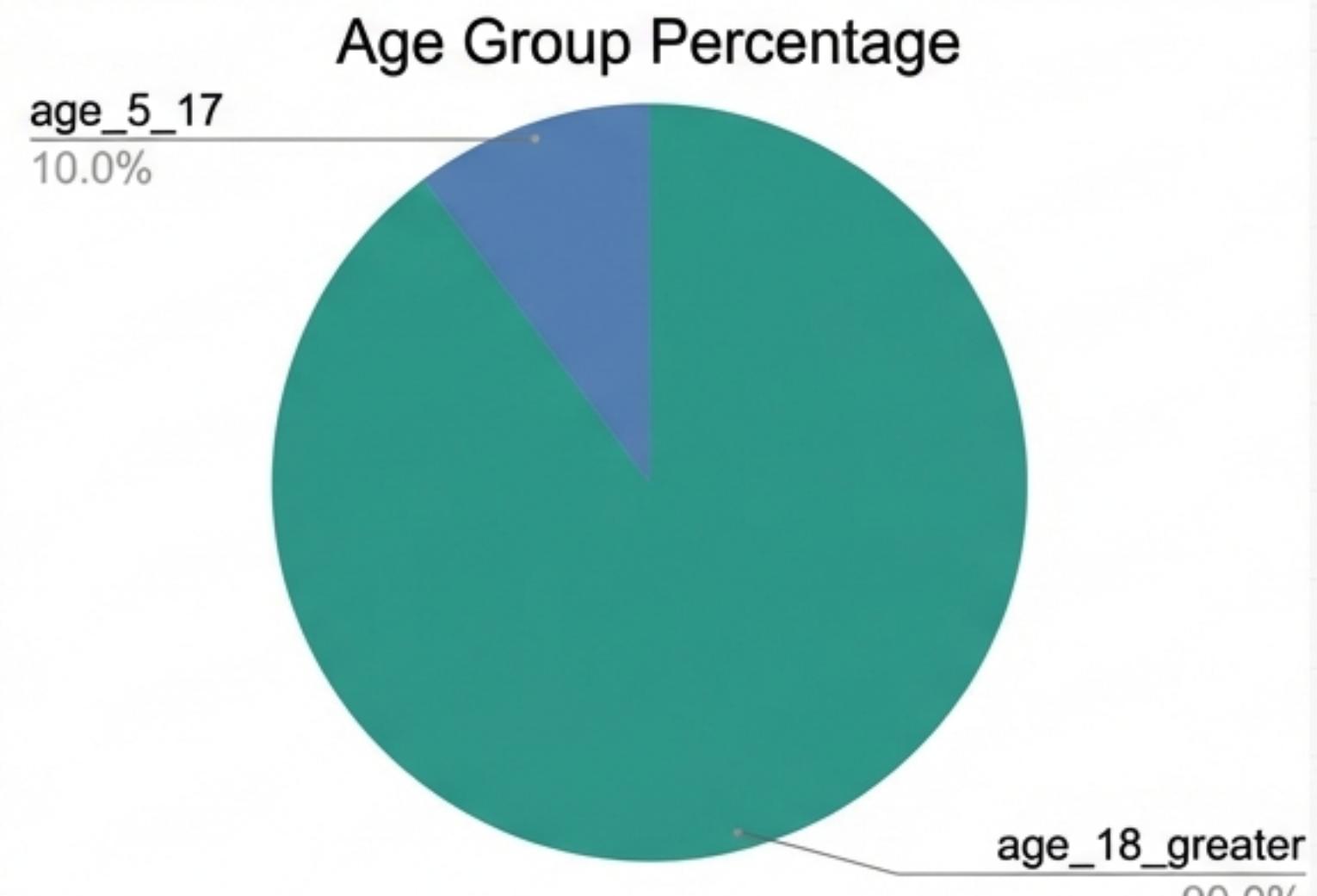
# The ‘Hindi Belt’ Drives Identity Acquisition

Merriweather: Geographic distribution of new enrollments.



# Demographic Updates Reflect an Active, Mobile Adult Workforce

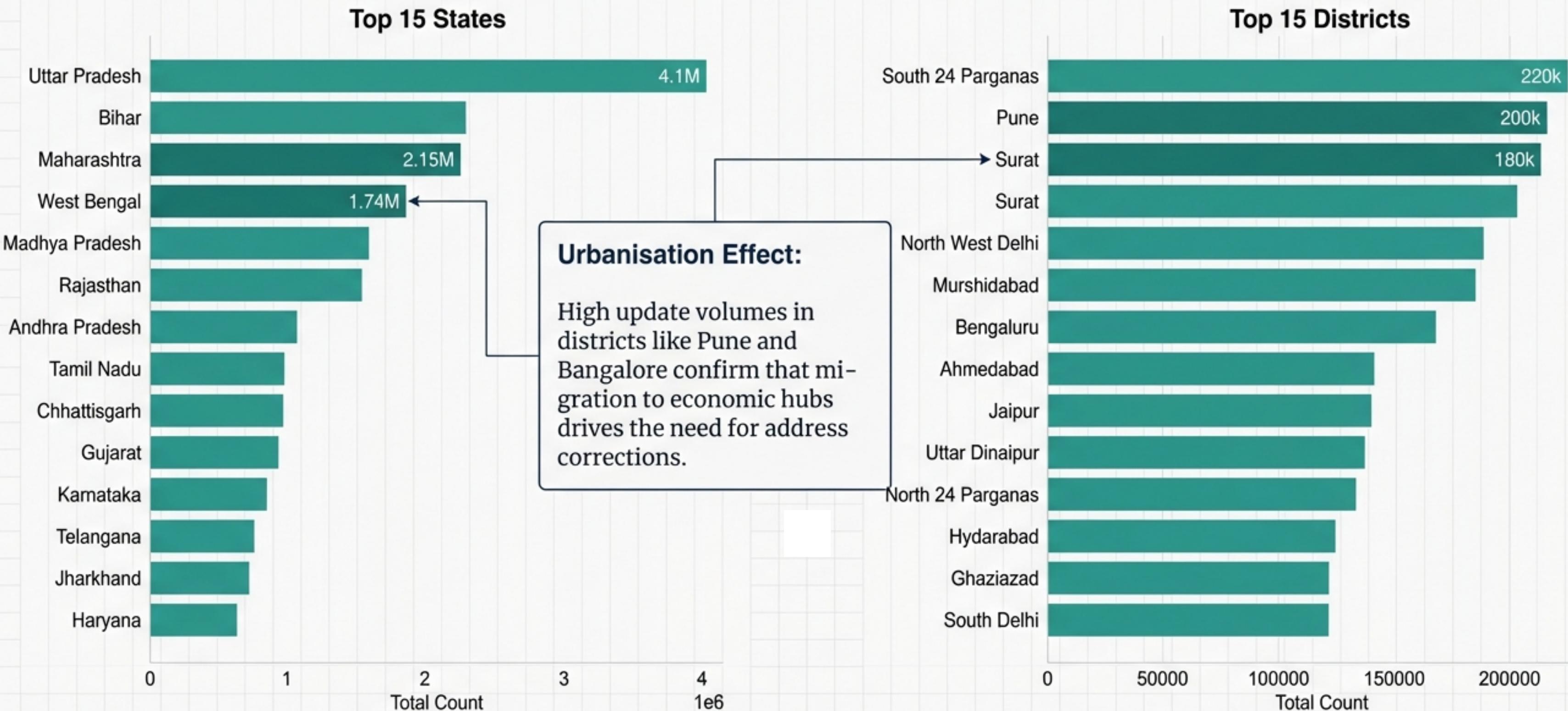
Unlike enrollment, the demographic update = demographic update stream is dominated by adults. This high volume of address and name changes serves as a proxy for internal internal migration and economic activity, as citizens update records to satisfy banking KYC and welfare delivery requirements.



**Key Insight:** 90% of demographic updates (21.2M records) come from the 18+ age group.

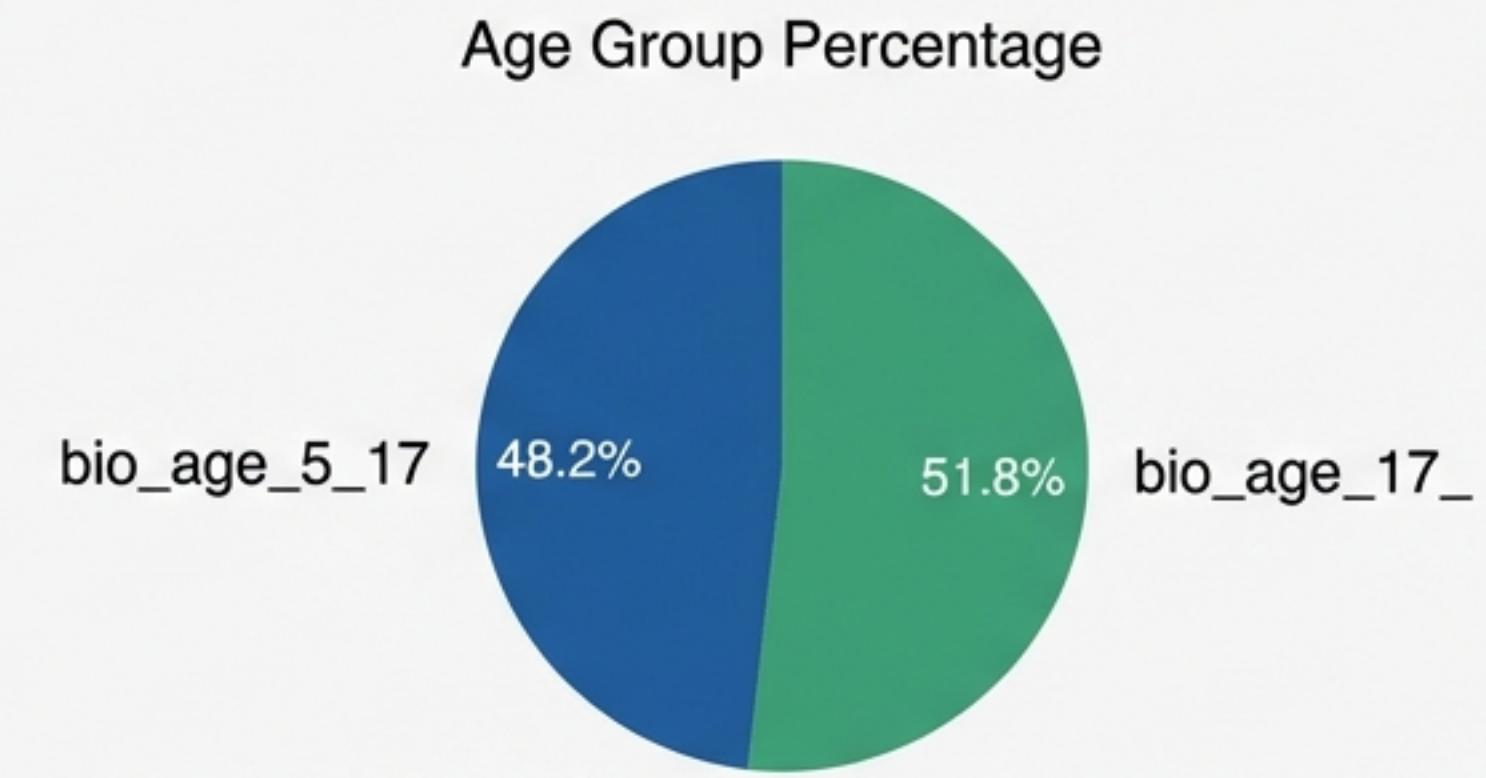
# Economic Centres Drive the Demand for Data Maintenance

## Top States and Districts for Demographic Updates.



# Mandatory Checkpoints at Ages 5 and 15 Drive Biometric Volume

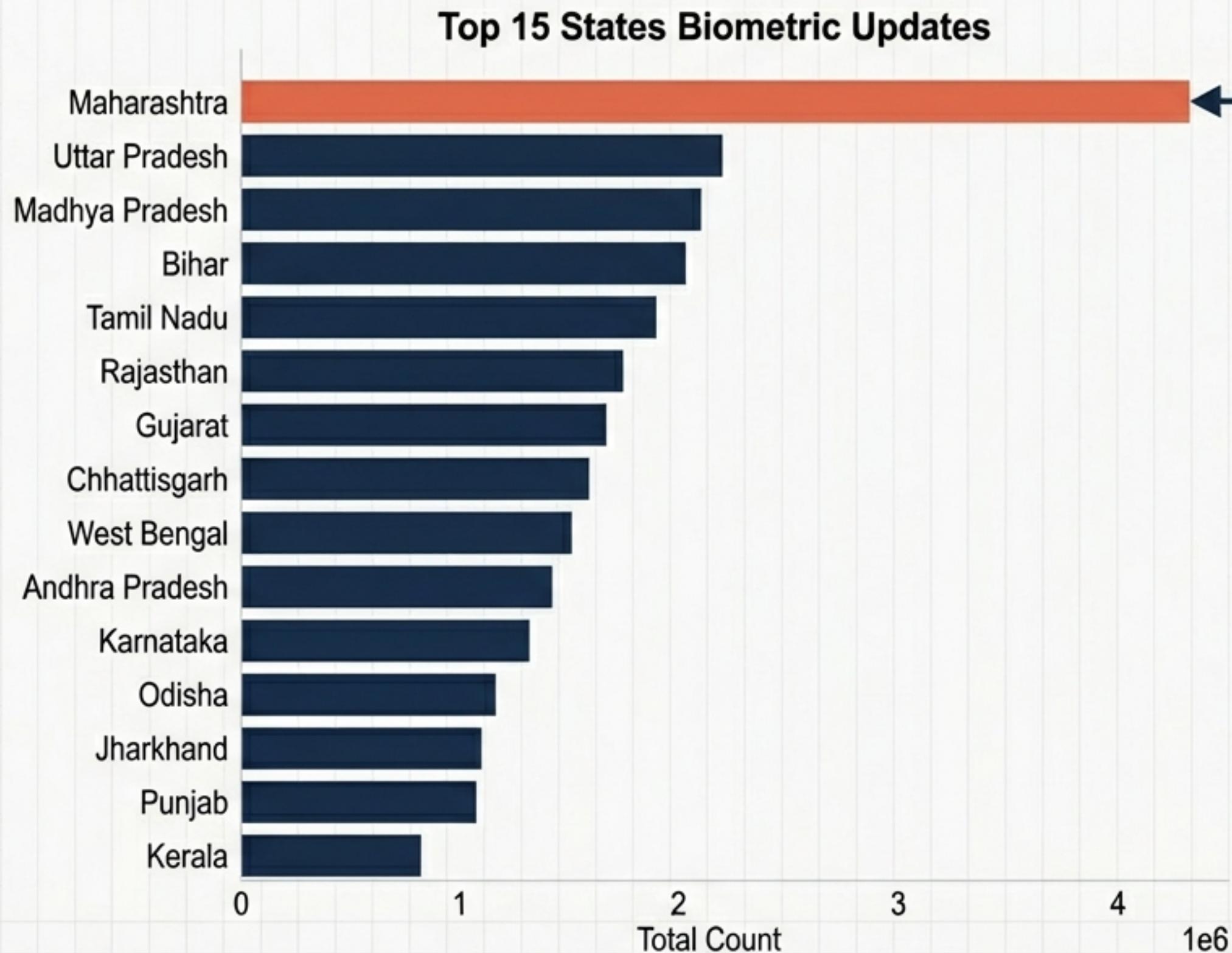
This stream represents the highest security tier of the ecosystem, processing 28.4 million records. The volume is split evenly between students (mandatory updates) and adults (authentication correction).



**Operational Insight:** The 5-17 age block accounts for 48% of volume, driven by mandatory biometric locking policies at age 5 and 15.

# Maharashtra as the Prototype of a ‘Mature’ Aadhaar State

## Biometric update volume by state.



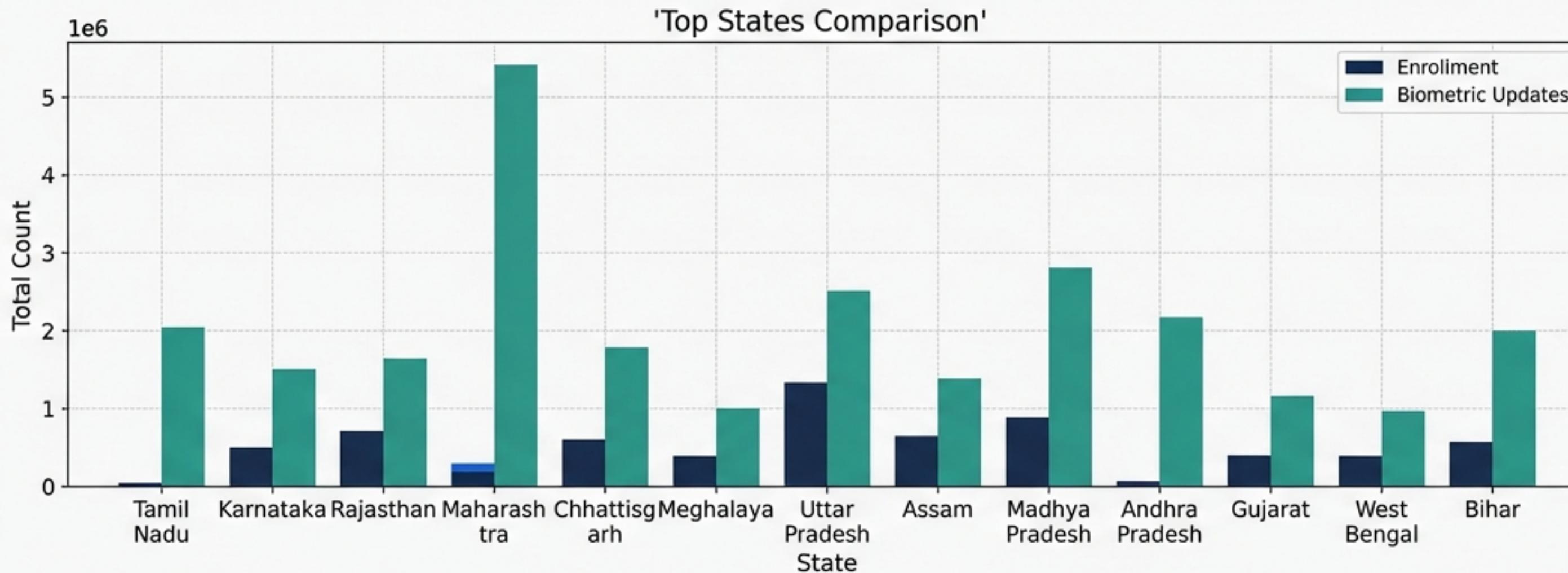
### The Maharashtra Anomaly:

Recording 4.4 Million biometric updates, Maharashtra's volume is nearly double that of Uttar Pradesh.

This represents the future state of the ecosystem:  
**Low Enrollment + High Maintenance.**

# Divergent State Behaviours: Acquisition vs. Maintenance

## Comparative analysis of Enrollment vs. Biometric Updates.



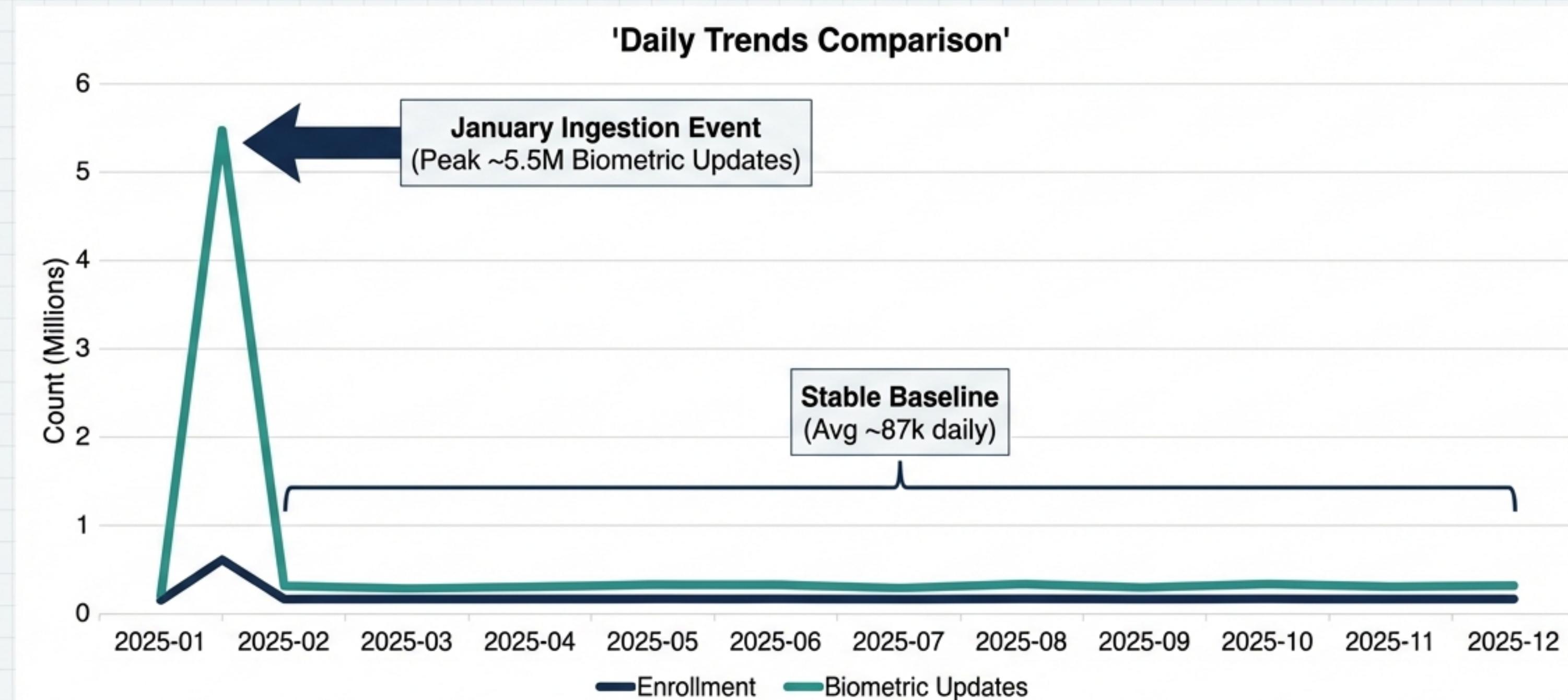
### The Data Divide:

1. The Acquisition Model (Bihar): High Enrollment relative to updates.
2. The Maintenance Model (Tamil Nadu & Andhra Pradesh): Enrollment is negligible, but update volumes are massive.

Strategic Implication: A 'one-size-fits-all' operational strategy is inefficient. Resources must be specialized by region.

# The January 2025 Surge: An Operational Anomaly

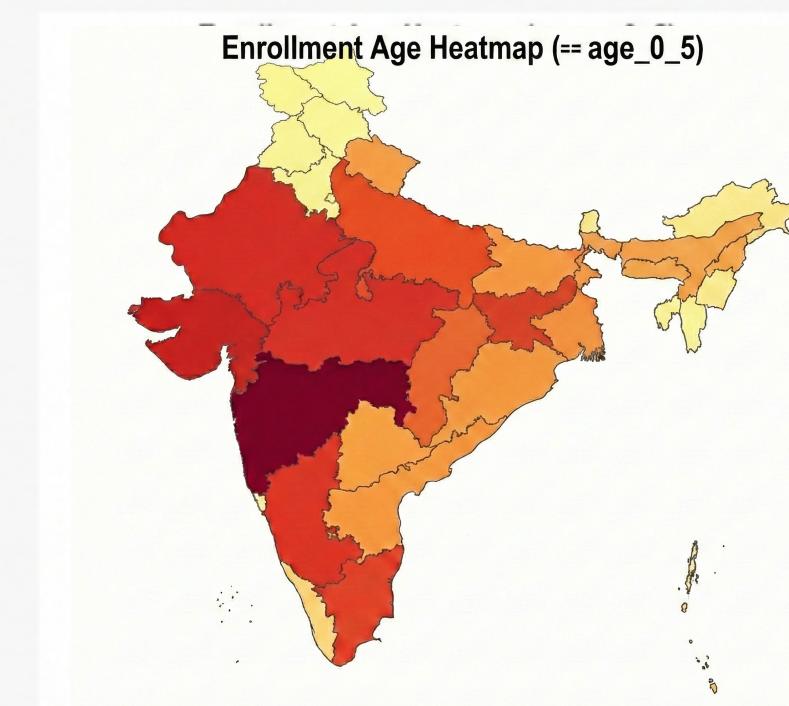
All three datasets show a massive outlier spike in January 2025 before stabilising to a consistent baseline. This suggests a backlog clearance event or a specific government deadline-driven campaign.



# The Citizen Identity Lifecycle

Tracing the user journey across the three datasets.

## Entry (0-5 Years)

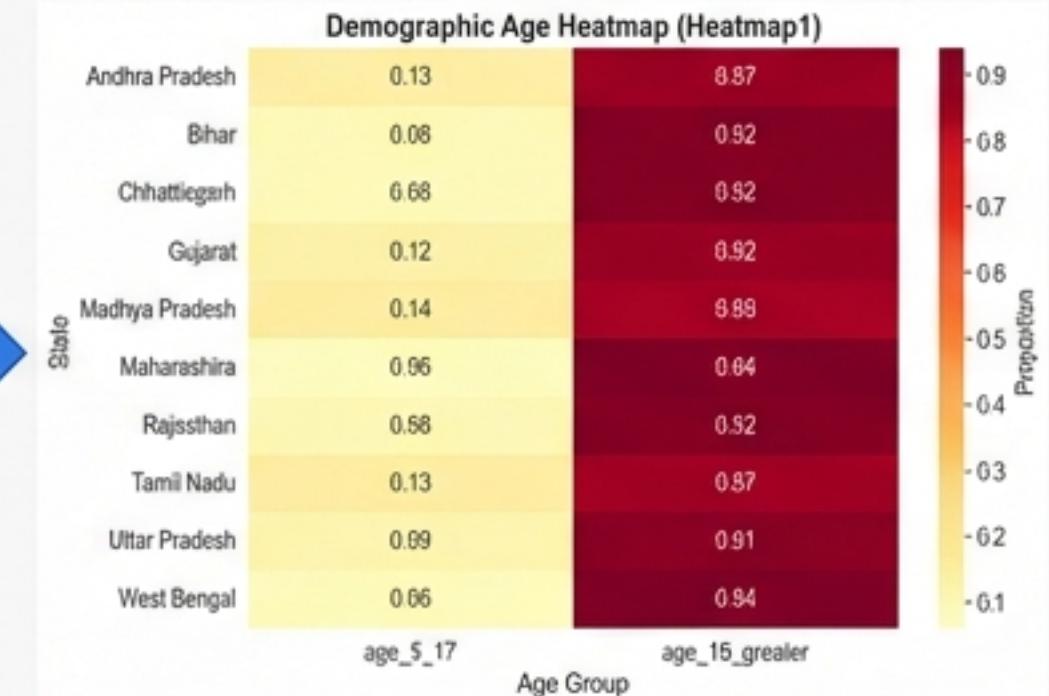


## Verification (5-17 Years)



**Domain: Enrollment.**  
The system captures identity at birth.

## Sustainment (18+ Years)

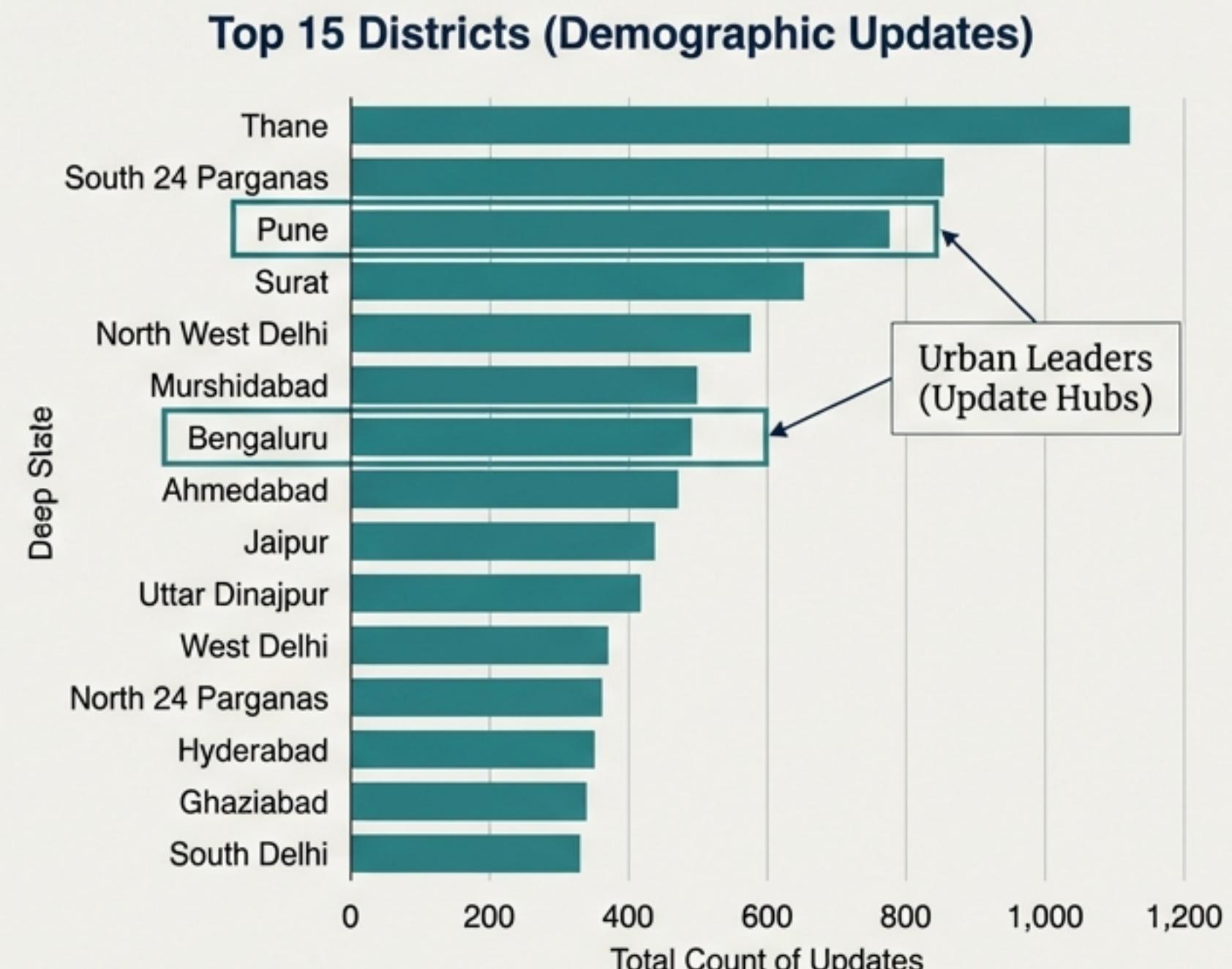
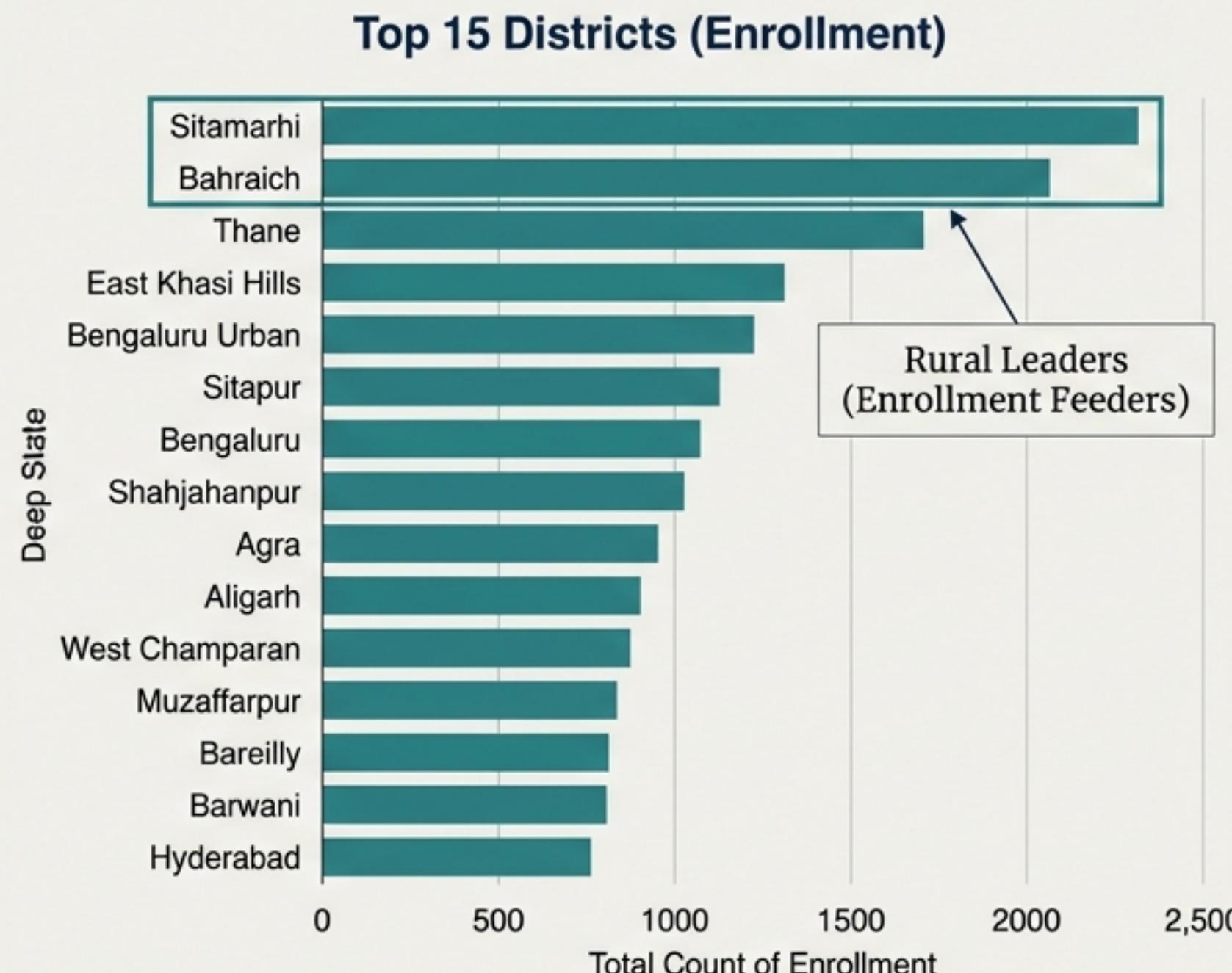


**Domain: Biometrics.**  
Identity matures through mandatory locking.

**Domain: Demographics.**  
Citizens maintain identity to reflect migration and employment.

# District-Level Hotspots Reveal Urban-Rural Divides

## Top 15 Districts Analysis



# Strategic Recommendations for the Sustainment Phase

**01**

## Asymmetric Resource Allocation

Shift budget and personnel from general enrollment centres to dedicated "Update Hubs" in Maharashtra, Karnataka, and Tamil Nadu to handle the verification load.

**02**

## Anganwadi Integration

With 60% of enrollments in the 0-5 age bracket, decouple enrollment from general banks and integrate exclusively with early childhood care centres (Anganwadis).

**03**

## Self-Service Expansion

The 21M+ adult demographic updates suggest a highly literate user base. Expand self-service digital tools to reduce footfall at physical centres.

# Appendix: Methodology & Data Scope

## Data Sources & Timeframe

- Data Sources: UIDAI API Logs (Anonymised).
- Timeframe: 03 January 2025 – 12 December 2025.

## Data Scope

- Enrollment: 2.6M records
- Demographic Updates: 23.6M records
- Biometric Updates: 28.4M records

## Definitions

- 0-5 Cohort: ‘Bal Aadhaar’ (no biometrics captured).
- Biometric Update: Includes both mandatory (5/15 years) and voluntary updates.

**Disclaimer:** January 2025 data contains outlier spikes attributed to backlog processing; trends should be viewed with this anomaly in mind.

# Structured Methodology

The analysis follows a structured, reproducible data analytics pipeline implemented in Python:

## • Data Ingestion

- Loaded multiple UIDAI-provided CSV files for enrollment, demographic updates, and biometric updates
- Merged partitioned datasets into unified data frames

## • Data Cleaning

- Removed duplicate records
- Handled missing values in critical geographic fields (state, district)
- Filtered invalid or incomplete records

## • Data Standardization

- Normalized age-group columns across datasets (0–5, 5–17, 18+)
- Standardized date formats and derived temporal features (day, month, year)

## • Feature Engineering

- Computed total activity metrics per record
- Generated state-wise, district-wise, and age-wise aggregates

## • Exploratory Data Analysis

- Conducted temporal trend analysis (daily, monthly, cumulative)
- Performed geographic distribution analysis
- Compared enrollment, demographic, and biometric volumes

## • Visualization

- Developed charts and infographics using Matplotlib and Seaborn
- Highlighted anomalies, dominant patterns, and comparative trends

