Aadhaar Data Analysis Report

Comprehensive Analysis of Enrolment and Update Patterns

Data Period: 2025-03-02 to 2025-12-31

Prepared: 2026-01-05

Executive Summary

This report presents a comprehensive analysis of Aadhaar enrolment and update data,
covering 48 states and 985 districts.
The analysis period spans from 2025-03-02 to 2025-12-31, during which:

- 5,435,702 new Aadhaar enrolments were processed
- 49,295,187 demographic updates were completed
- 69,763,095 biometric updates were recorded

Key findings reveal significant geographic disparities in Aadhaar activity, with the top 5 states
accounting for the majority of enrolments. The adult population (18+) dominates new enrolments,
while youth biometric transitions show concerning variation across states.

Key Insights

• Total of 5,435,702 new Aadhaar enrolments processed during the analysis period.

• Demographic updates (49,295,187) outnumber enrolments by 9.1x, indicating active information maintenance.

• Top 5 states (Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal, Maharashtra) account for majority of enrolments,
showing geographic concentration.

• Adult population (18+) dominates enrolments at 3.1%.

• 71 anomalous days detected in enrolments, potentially indicating system stress or campaigns.

• Youth biometric transitions vary significantly by state, with Daman & Diu leading and The Dadra And Nagar Haveli
And Daman And Diu lagging.

• Cold spots identified in 5 states requiring targeted intervention for Aadhaar coverage.

Section 1: Problem Statement & Approach

Problem 1: Geographic Disparity in Aadhaar Coverage

Which regions show concerning gaps in Aadhaar adoption, and what patterns
indicate underserved populations? This analysis aggregates enrolment data by state and district,
identifies cold spots, and correlates with demographic update patterns to find areas requiring intervention.

Problem 2: Youth Biometric Transition Patterns

How effectively are children transitioning to adult biometrics, and are there
bottlenecks in the system? We analyze biometric update frequency for the 5-17 age bracket,
track temporal patterns as children approach adulthood, and identify states with low transition rates.

Problem 3: Update Behavior Anomalies

What unusual patterns in demographic and biometric updates indicate system stress,
fraud potential, or policy gaps? Anomaly detection is applied to update frequencies, identifying
temporal spikes, geographic clustering, and unexpected demographic patterns.

Section 2: Dataset Description

Enrolment Dataset

Records: 1,006,029
Columns: date, state, district, pincode, age_0_5, age_5_17, age_18_greater
Contains new Aadhaar registrations categorized by age groups.

Demographic Update Dataset

Records: 2,071,700
Columns: date, state, district, pincode, demo_age_5_17, demo_age_17_
Tracks updates to name, address, date of birth, gender, and mobile number.

Biometric Update Dataset

Records: 1,861,108
Columns: date, state, district, pincode, bio_age_5_17, bio_age_17_
Contains fingerprint, iris, and face biometric updates.

Data Quality Assessment

Enrolment: 1,006,029 rows, 22957 duplicates, 199.1 MB

Demographic: 2,071,700 rows, 473601 duplicates, 394.8 MB

Biometric: 1,861,108 rows, 94896 duplicates, 354.5 MB

Section 3: Methodology

Data Preprocessing

1. Date Parsing: Converted DD-MM-YYYY strings to datetime objects
2. PIN Code Validation: Filtered to valid 6-digit numeric codes
3. State Name Normalization: Standardized to consistent naming conventions
4. Missing Value Treatment: Dropped records with null geography fields
5. Feature Engineering: Extracted year, month, quarter, day_of_week; calculated totals

Analysis Techniques

- Temporal Analysis: Time series aggregation, rolling averages, growth rates
- Geographic Analysis: State and district-level aggregation, ranking, hotspot detection
- Demographic Analysis: Age group distribution, transition ratio calculation
- Anomaly Detection: IQR method for identifying outliers
- Comparative Analysis: Cross-dataset metrics, ratio analysis

Tools and Libraries

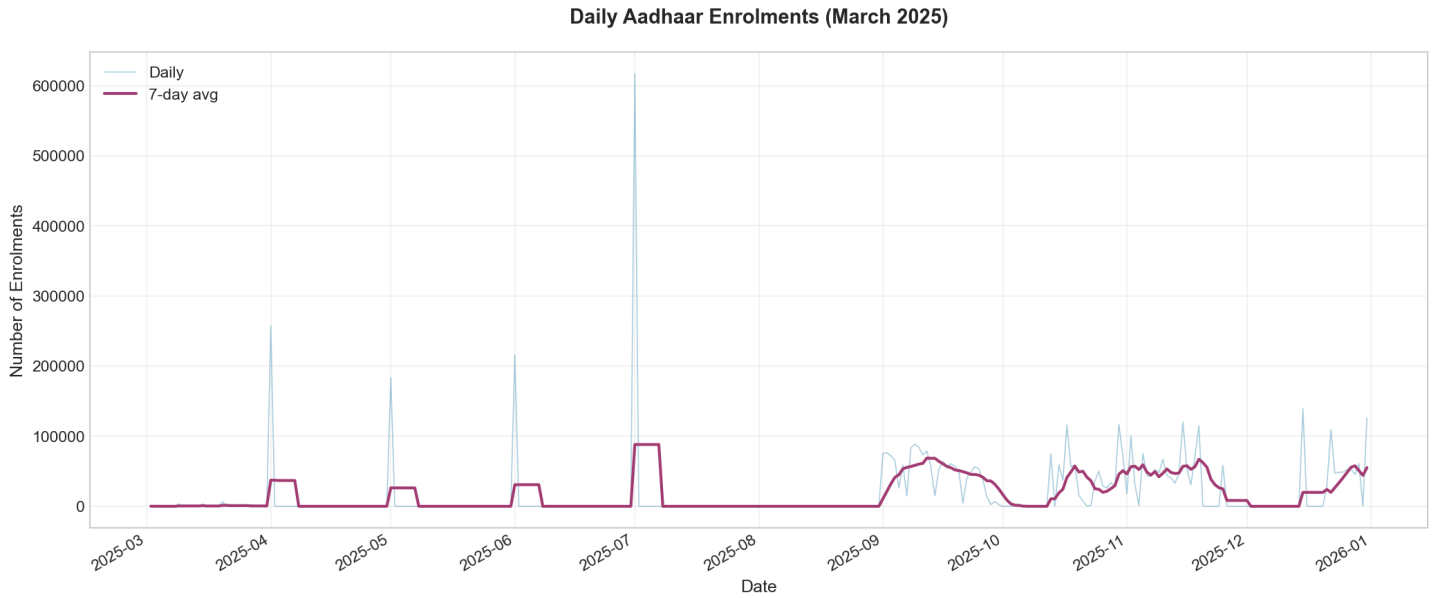Python 3.14 with pandas, numpy, matplotlib, seaborn, scipy, scikit-learn, fpdf2

Section 4: Analysis & Visualizations

A. Temporal Analysis

Average daily enrolments: 17,822
Maximum daily enrolments: 616,868
Minimum daily enrolments: 0

Daily Aadhaar Enrolments (March 2025)
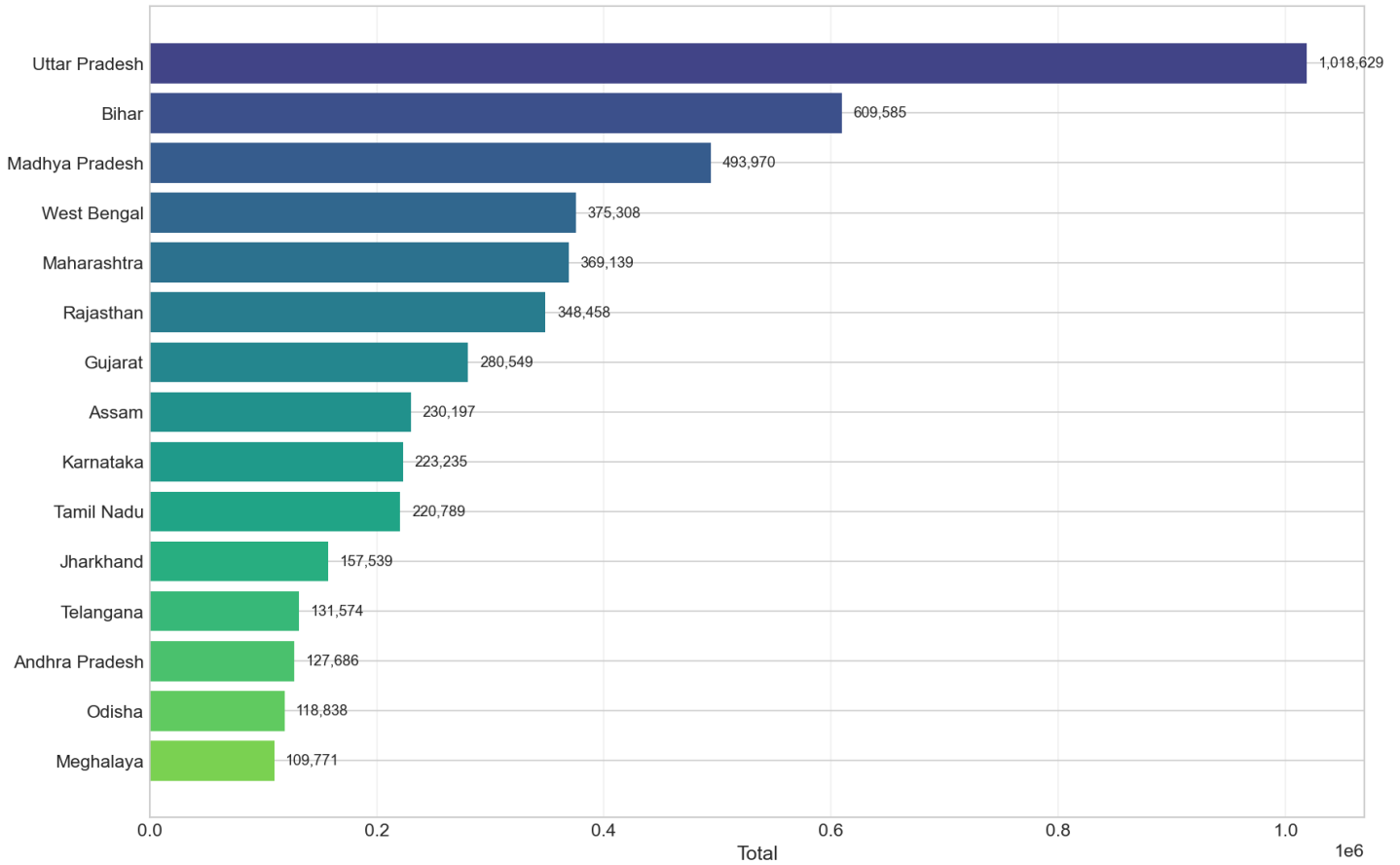
B. Geographic Analysis

Top performing states: Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal, Maharashtra
Bottom performing states: Dadra & Nagar Haveli, Daman & Diu, West Bengal, West Bangal, Westbengal
Hotspots (90th percentile): Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal, Maharashtra
Coldspots (10th percentile): Dadra & Nagar Haveli, Daman & Diu, West Bengal, West Bangal, Westbengal

Top 15 States by Aadhaar Enrolments

State	Total
Uttar Pradesh	1,018,629
Bihar	609,585
Madhya Pradesh	493,970
West Bengal	375,308
Maharashtra	369,139
Rajasthan	348,458
Gujarat	280,549
Assam	230,197
Karnataka	223,235
Tamil Nadu	220,789
Jharkhand	157,539
Telangana	131,574
Andhra Pradesh	127,686
Odisha	118,838
Meghalaya	109,771

C. Demographic Analysis

0-5 years: 3,546,965 (65.3%)

5-17 years: 1,720,384 (31.6%)

18+ years: 168,353 (3.1%)

Aadhaar Enrolment by Age Group
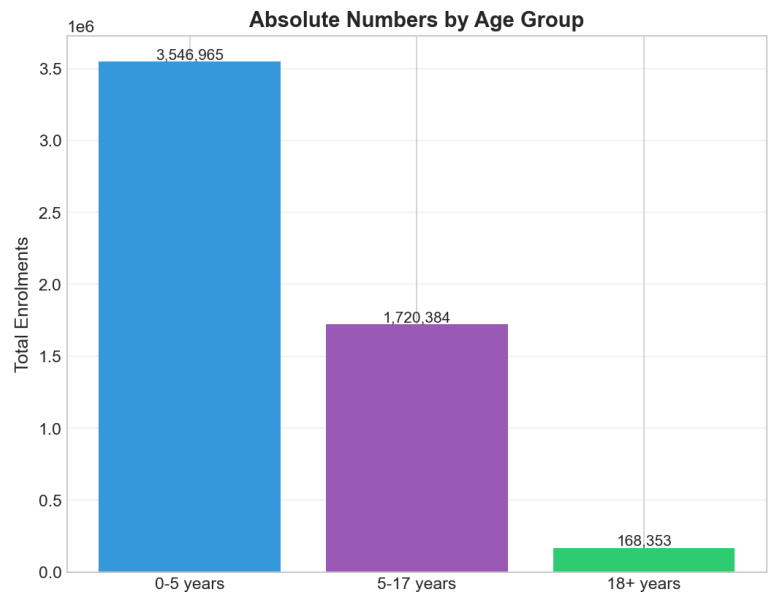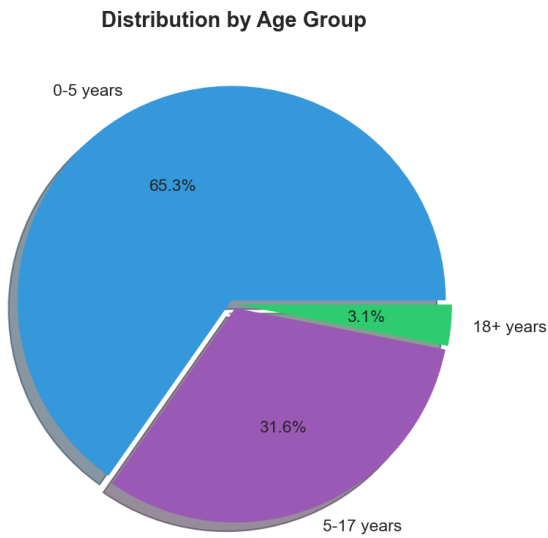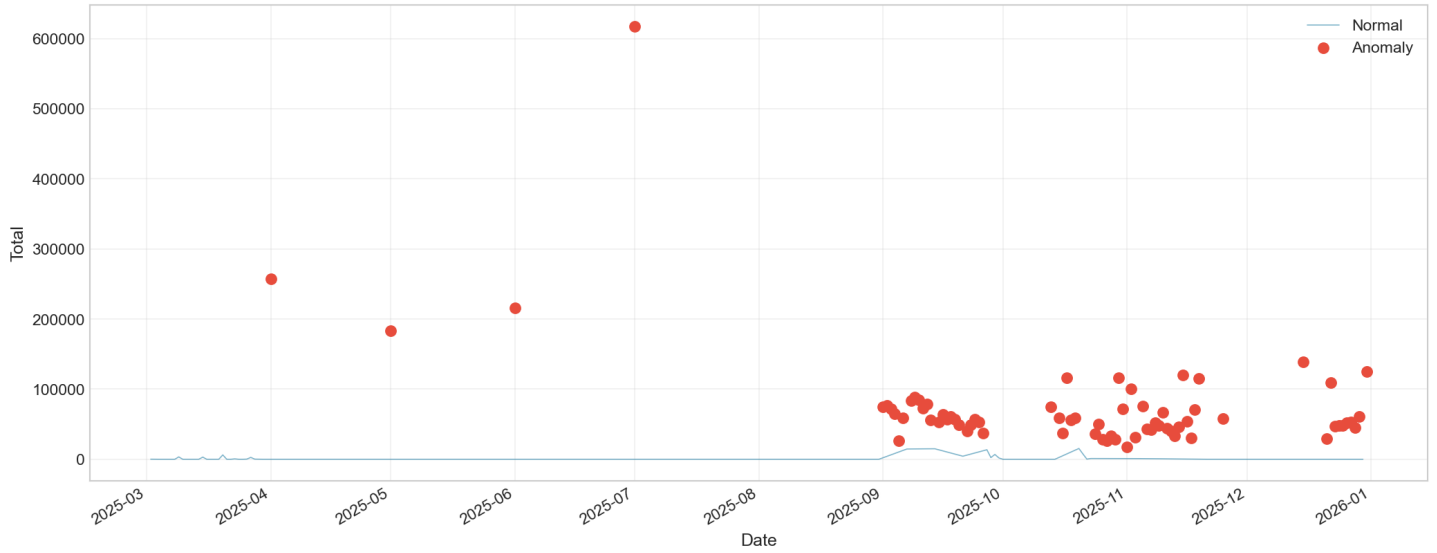
D. Anomaly Detection

Enrolment anomalies detected: 71 days
Demographic update anomalies detected: 9 days

Anomalous days indicate periods of unusually high or low activity that warrant investigation.
These could indicate system issues, special campaigns, or data quality concerns.

Enrolment Anomalies Detected

E. Comparative Analysis

The comparative analysis reveals the relationship between new enrolments
and update activity across states. States with high enrolment but low update rates may indicate
populations with stable information, while high update rates could suggest demographic mobility
or data quality issues.

State-wise Activity Comparison
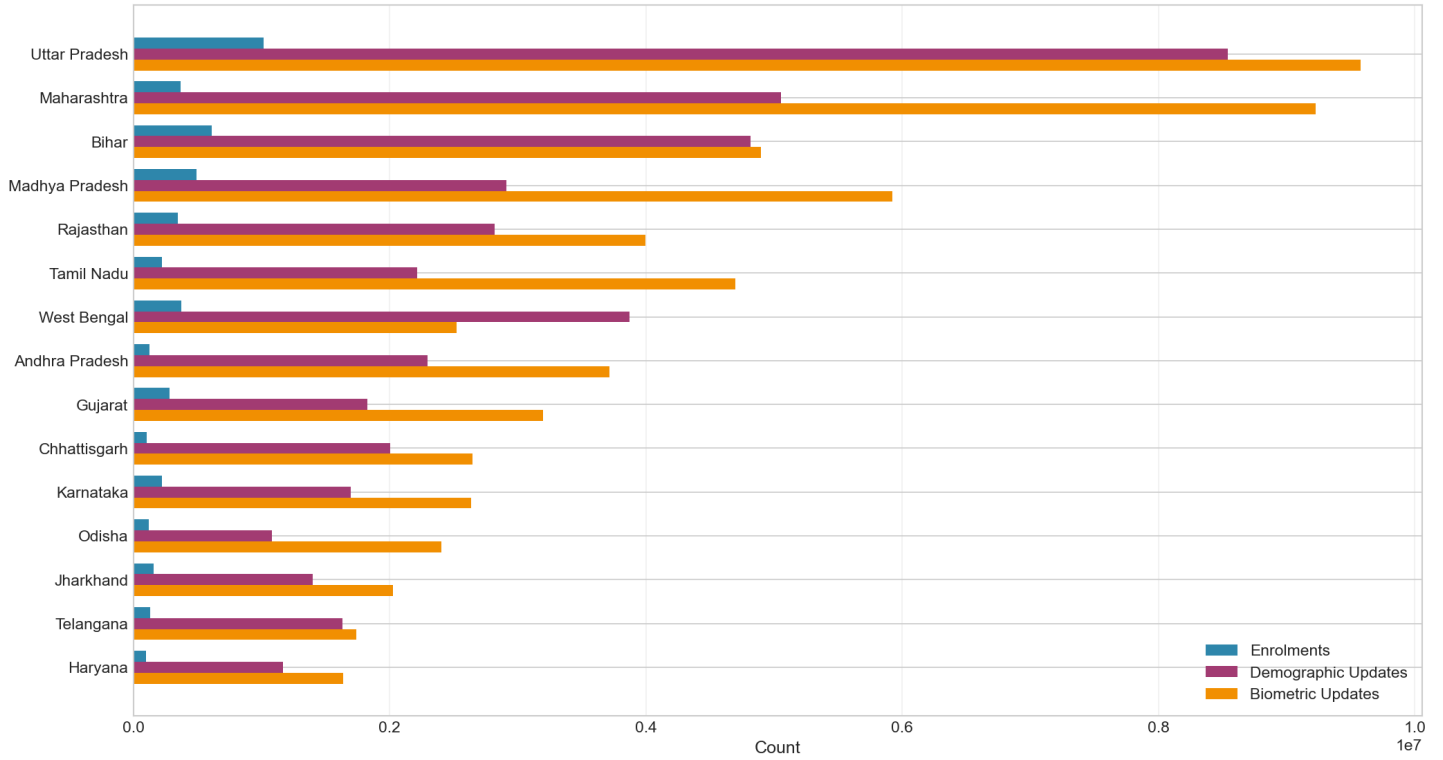
F. Youth Biometric Transition Analysis

Best performing states for youth transitions: Daman & Diu, Andaman & Nicobar, Daman And Diu
Underperforming states: The Dadra And Nagar Haveli And Daman And Diu, 100000, Meghalaya

States with low transition ratios may need targeted campaigns to ensure children
update their biometrics as they approach adulthood.

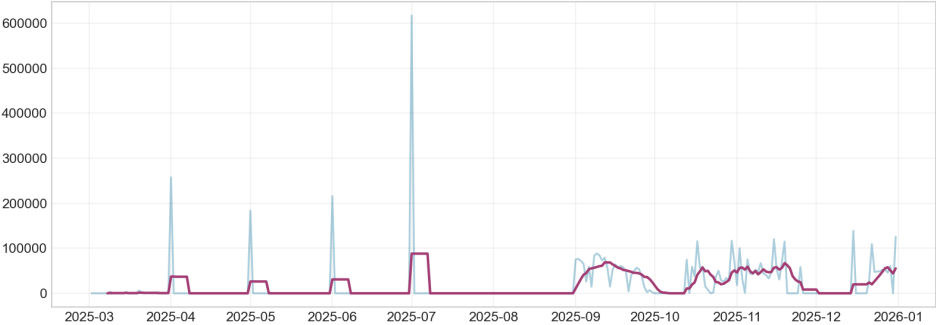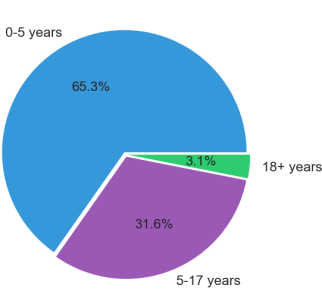**Youth Biometric Transition Rates by State**
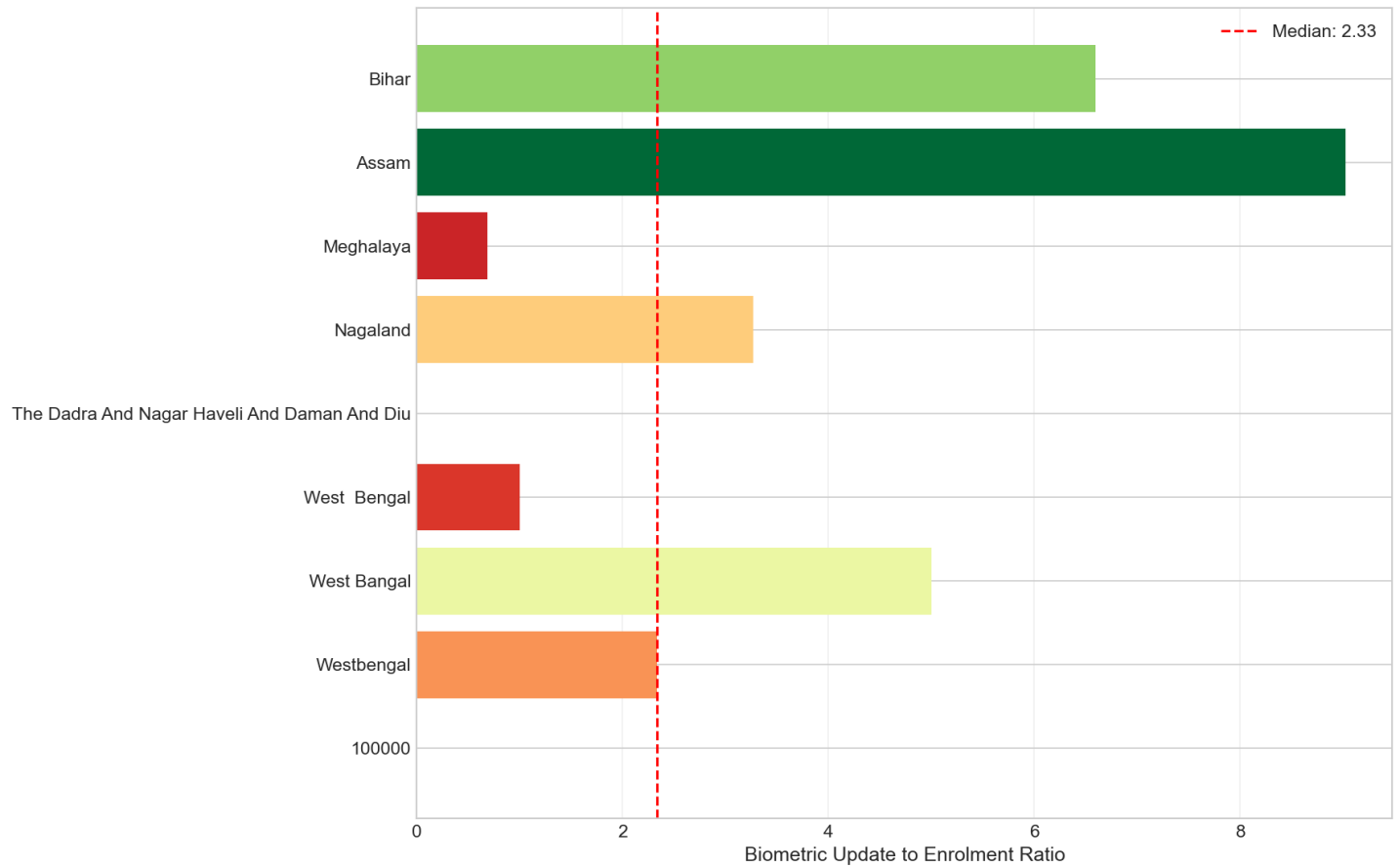
Summary Dashboard

Aadhaar Data Analysis Dashboard

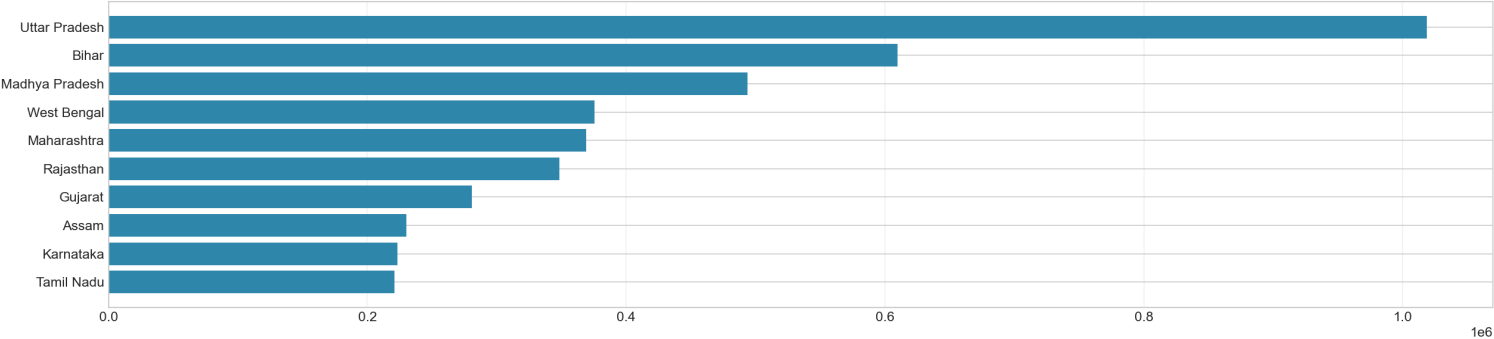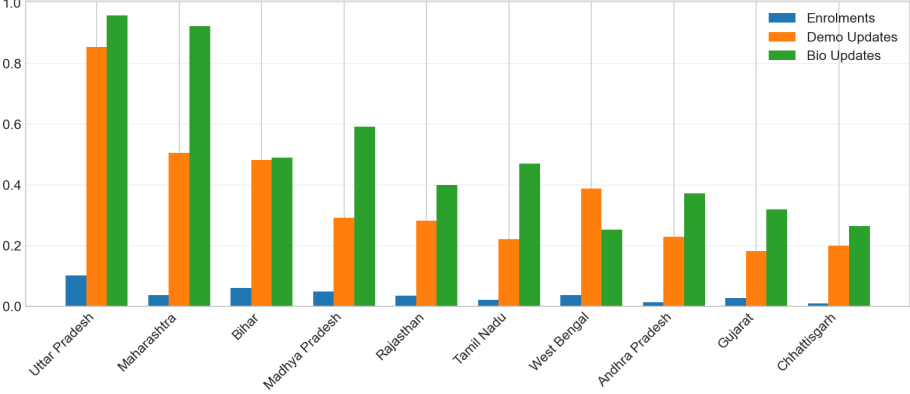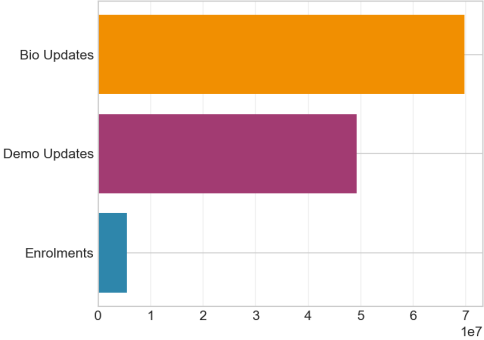# Section 5: Key Insights & Recommendations

Key Insights

1. Total of 5,435,702 new Aadhaar enrolments processed during the analysis period.

2. Demographic updates (49,295,187) outnumber enrolments by 9.1x, indicating active information maintenance.

3. Top 5 states (Uttar Pradesh, Bihar, Madhya Pradesh, West Bengal, Maharashtra) account for majority of enrolments, showing geographic concentration.

4. Adult population (18+) dominates enrolments at 3.1%.

5. 71 anomalous days detected in enrolments, potentially indicating system stress or campaigns.

6. Youth biometric transitions vary significantly by state, with Daman & Diu leading and The Dadra And Nagar Haveli And Daman And Diu lagging.

7. Cold spots identified in 5 states requiring targeted intervention for Aadhaar coverage.

Recommendations

1. Focus enrollment drives on bottom-performing states to improve national coverage equity.

2. Investigate anomaly days to understand causes and optimize system capacity.

3. Implement targeted youth biometric update campaigns in low-transition states.

4. Consider mobile enrollment units for remote districts showing low activity.

5. Monitor demographic update patterns to predict system load and plan resources.

Areas for Further Investigation

1. Deep-dive into district-level patterns within underperforming states
2. Time-series forecasting for capacity planning
3. Correlation with socioeconomic indicators for targeted interventions
4. Investigation of PIN code patterns for urban/rural classification
5. Seasonal pattern analysis for campaign timing optimization