

# Unlocking Societal Trends in Aadhaar Enrolment, Demographic and Updates

## A Data-Driven Framework for Insights, Anomaly Detection, and System Optimisation

**Participants:** Rajdeep Ramola (rajdeepramolascr.cse@shivalikcollege.edu.in)

Vipul Bijalwan

Gopal Datt

**Date:**20 January 2026

## 1. Problem Statement and Approach

Aadhaar is the world's largest identity ecosystem, and its enrolment and update processes generate high-volume data that reflect societal behaviour, mobility patterns, demographic shifts, and administrative performance. The datasets on **Enrollments**, **Biometric Updates**, and **Demographic Updates** provide a unique opportunity to extract meaningful trends that can support policy decisions and strengthen operational efficiency.

### **Objective:**

To identify meaningful patterns, trends, anomalies, and predictive indicators using three Aadhaar datasets—**Enrolment**, **Biometric Update**, and **Demographic Update**—and convert them into actionable insights that can help optimise service delivery and inform evidence-based administrative decision-making.

### **Analytical Approach:**

Conduct dataset-wise analysis

Identify geographic, temporal, and demographic trends.

Detect anomalies that may signal operational, infrastructural, or behavioural issues.

Develop predictive indicators for workforce planning, enrolment surges, and update requirements.

Build a solution framework to support real-time monitoring and proactive intervention.

## 2. Datasets Used

This study uses three UIDAI-provided datasets:

### 2.1 Aadhaar Enrolment Dataset

**Columns include:**

State, District, Date, Pincode, age\_0\_5, age\_5\_17, age\_18\_greater, Latitude, Longitude

**Purpose:**

Assess enrolment penetration, state-wise enrollments, and evaluate seasonal and geographic enrolment trends.

---

### 2.2 Biometric Update Dataset

**Columns include:**

State, District, Date, bio\_age\_5\_17, bio\_age\_17\_, Pincode, Latitude, Longitude

**Purpose:**

Understand biometric update cycles, identifying low penetration areas, and infrastructural challenges (e.g., device failures, low capture quality).

---

### 2.3 Demographic Update Dataset

**Columns include:**

State, District, Date, demo\_age\_5\_17, demo\_age\_17\_, Longitude, Latitude

**Purpose:**

Identify population mobility, concentration of population district-wise, district-level activity, and behavioural patterns in demographic corrections.

## 3. Methodology

### 3.1 Data Ingestion and Standardisation

Merged smaller datasets into 3 large dataset.

---

## 3.2 Data Cleaning

Standardized states names using state mapping and fixing old names like Uttaranchal and Orissa.

Fixed district names using an external dataset by performing a join on pincode column.

Detected duplicate records and eliminated redundant entries.

Added longitude and latitude column to all the datasets using pincode mapping for better plotting on an Indian map.

---

## 3.3 Feature Engineering

Created new analytical variables for grouped analysis:

**State grouped** = to group updates and enrollments according to states

**District grouped** = to group updates according to districts

**Updates grouped** = combine all the updates of different age groups

---

## 3.4 Analytical Techniques

Univariate, bivariate, and multivariate analysis

Time-series decomposition for each dataset

Geospatial analysis using Plotly/Folium

---

## 3.5 Tools Used

Python, Pandas, NumPy

Matplotlib, Plotly, Seaborn

Folium/Kepler.gl for geographic charts

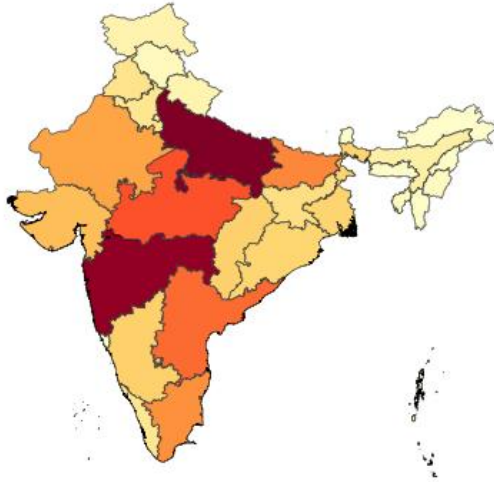
Excel and Power BI for cross-verification

GitHub for version control

## 4. Data Analysis and Visualisation

### 4.1 Enrolment Trends

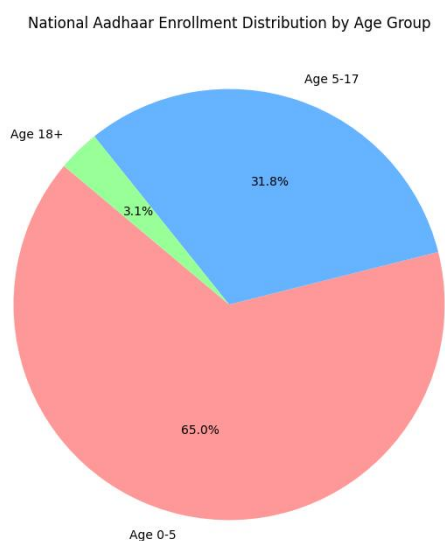
- Density Heatmap of Enrollments based on states



Abnormal rates of enrollments observed in bordering areas of Gujrat and West Bengal

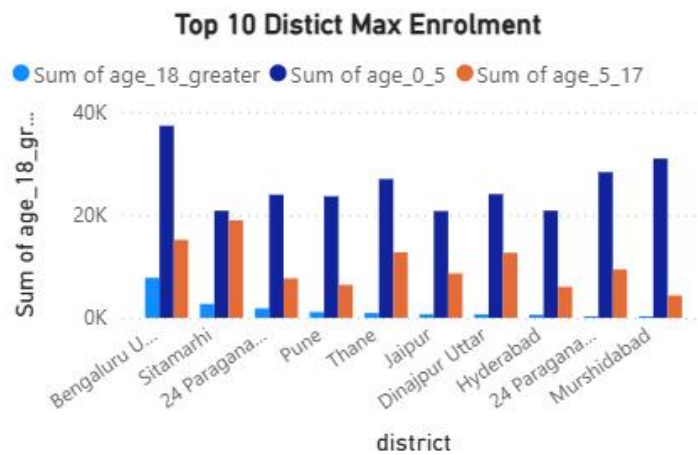
This might be due to illegal forging of documents for illegal immigrants.

- Enrollment stats of different age groups

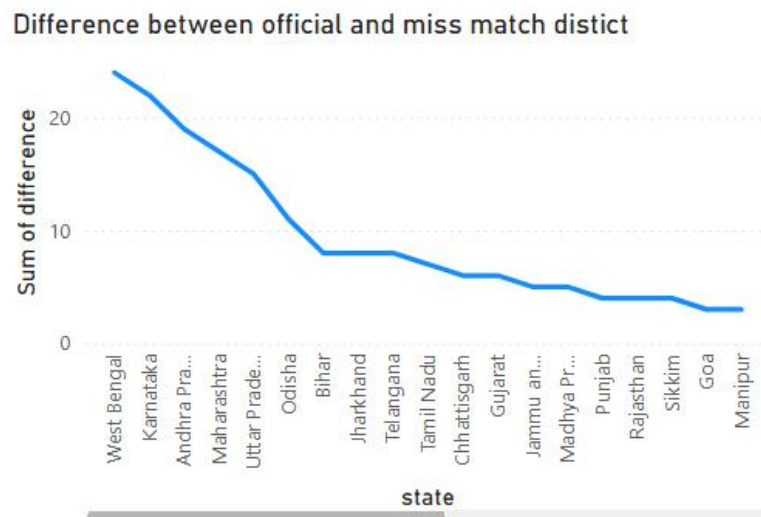


Lower rate of enrollment in adults which means most of the adults have been covered in case of enrollments.

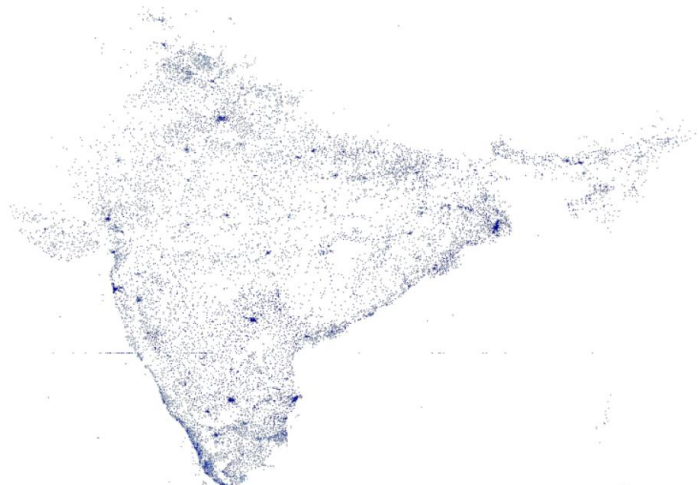
- Top 10 districts with max enrollments

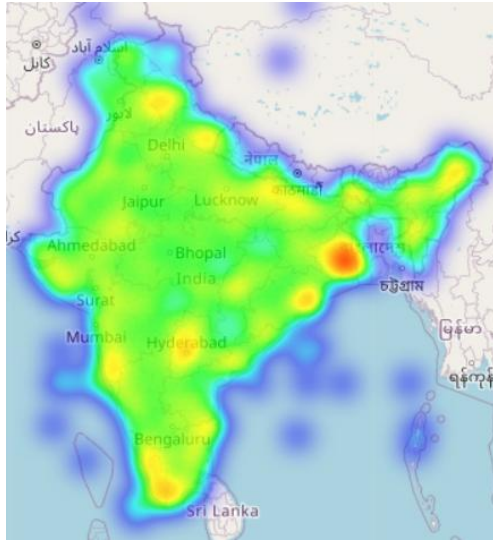


- States with highest errors in district names



- Plottings of aadhar enrollments

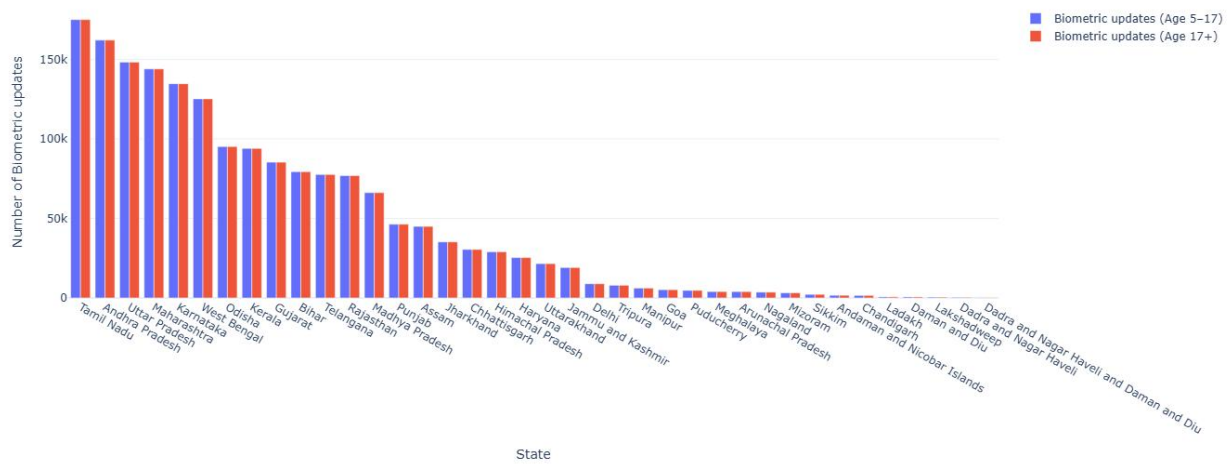




Abnormal patterns of enrollments in lower borders of West Bengal

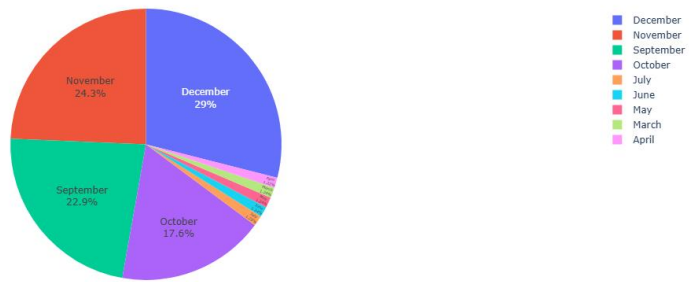
## 4.2 Biometric Update Patterns

- Statewise updates statistics



- Monthwise update distribution

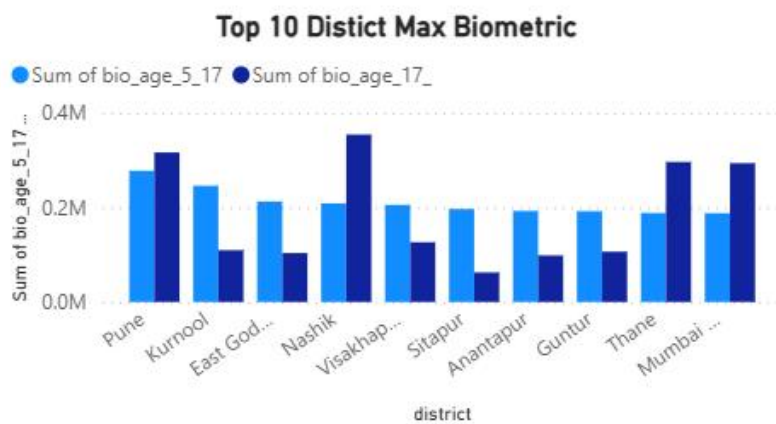
Month-wise Biometric updates Distribution (Age 5-17)



- Population-wise distribution of updates



- Top 10 districts with max updates

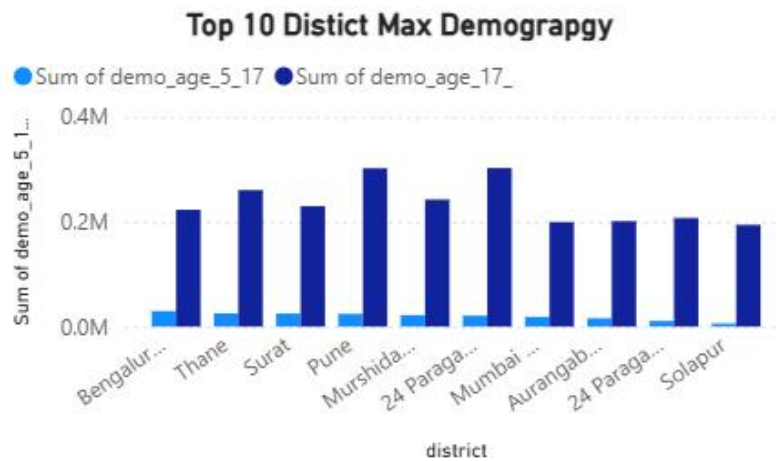


## 4.3 Demographic Update Behaviour

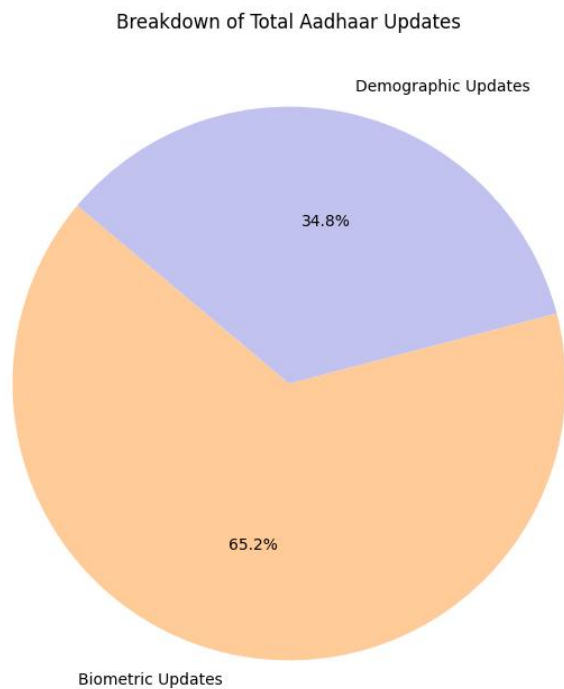
- Total people applying for demographic change



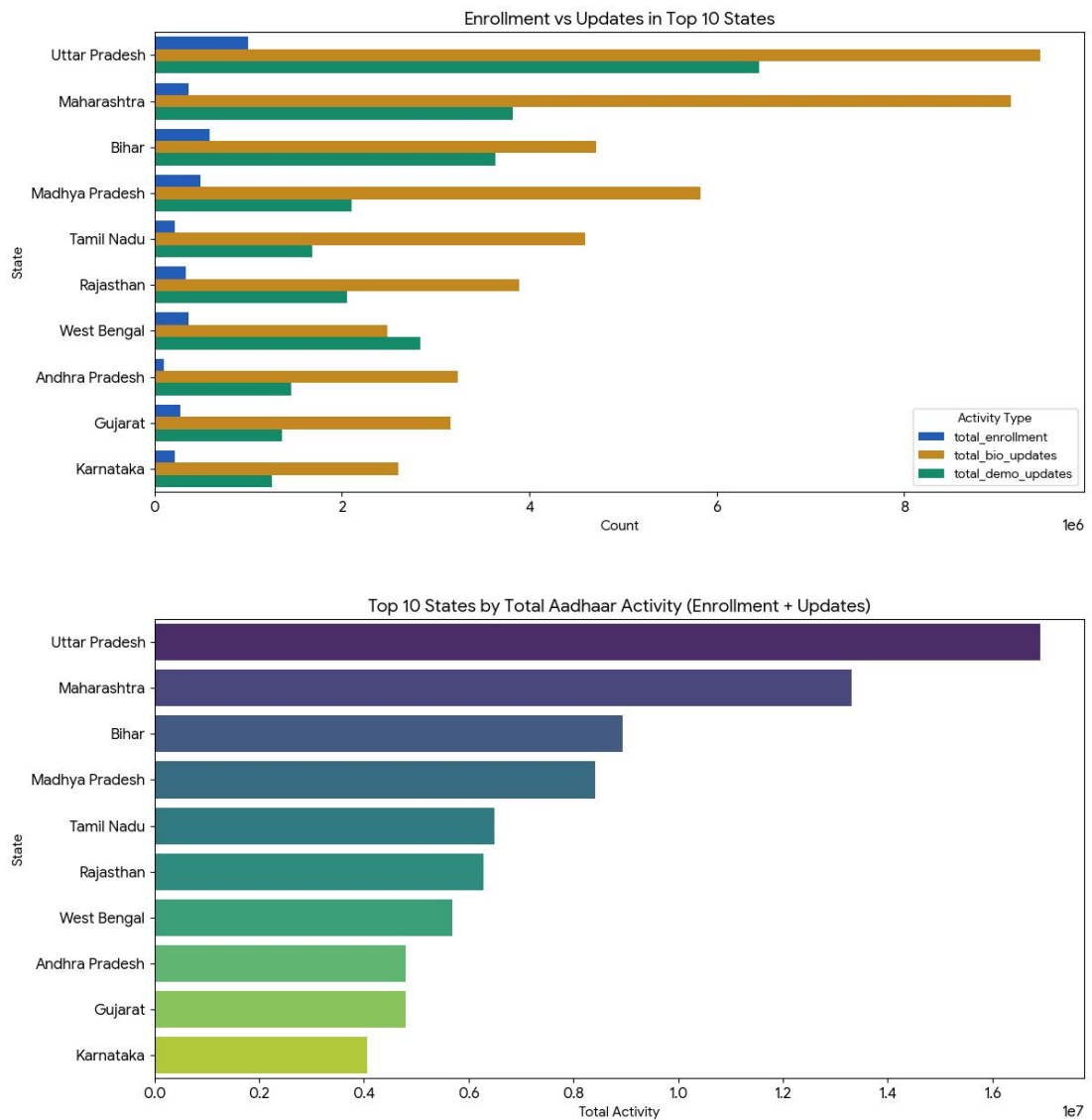
- Top 10 districts with max demographic updates



## 4.4 Cross-Dataset Correlation







## 4.5 Anomaly Detection Findings

Examples of detected anomalies:

Records with invalid state names, district names and invalid pin codes.

Unexpected spikes in demographic updates due to local administrative drives or migration surges.

Redundant records

Higher enrollments in bordering areas of India especially Gujrat and West Bengal.

Lower enrollments and updates in tribal belt of Chattisgarh, Jharkhand and Andhra Pradesh.

---

## 4.6 Forecasting Results

Predictive models show strong seasonal patterns across all three datasets.

Expected workload surges identified for **March**, **May**, and **October**.

---

## 5. Key Insights

### 5.1 Behavioural Insights

Migration-heavy districts consistently lead in demographic updates.

Mandatory biometric updates for children are uneven across states, signalling awareness gaps.

Enrolment saturation is nearly achieved in many southern and western states.

### 5.2 Operational Insights

Infrastructure weaknesses detected in districts with long-term low biometric update counts. ( Odisha, Chattisgarh)

### 5.3 Social Insights

Increased demographic updates suggest rising digital adoption.

Higher enrollments of 0–5 age group indicate natural population growth and tracking.

---

## 6. Recommendations and Solution Framework

### 6.1 Integrated Dropdown menu for data validation

Selecting state, districts and pincodes through dropdown menu:

For strict data validation

For maintaining clean data for Data analysis in future

State: Uttar Pradesh

Uttar Pradesh  
Maharashtra  
Karnataka  
Tamil Nadu  
Uttar Pradesh  
West Bengal

District: Lucknow

Lucknow  
Kanpur Nagar  
Varanasi  
Allahabad  
Agra  
Meerut

Tehsil: Sarojini Nagar

Sarojini Nagar  
Lucknow Sadar  
Bakshi Ka Talab  
Mohanlalganj  
Maliabad  
Mal

Pincode: 226008

226008  
226401  
226301  
226002  
226009

---

## 6.2 Early-Warning System

Automated alerts triggered by:

Zero-activity months

40% drop in any update/enrolment category

Device failure patterns

Abnormal spikes (potential misuse or sudden mass-update drives)

---

## 6.3 Targeted Outreach Programmes

Campaigns focused on biometric updates for children.

Special service windows in migration-heavy districts.

Public awareness programmes on demographic updates.

---

## 6.4 Predictive Capacity Planning

Allocate additional devices and operators before expected surge months.

Increase staffing in districts with rising migration and update loads.

Use forecasts to balance load between enrolment centres.

---

# 7. Impact & Applicability

## Administrative Impact

Faster detection of operational failures

Optimised resource allocation across districts

Improved service availability and efficiency

Evidence-based administrative planning

## Social Impact

Better service accessibility for children, migrants, and vulnerable populations

Reduced queues and waiting times

Enhanced reliability of Aadhaar as a foundational ID

## Scalability

The framework can scale to:

Panchayat-level monitoring

National real-time dashboards

Integration with mobile enrolment units and CSC networks

---

# Appendix

<https://github.com/Vipulbijalwan/UIDAI-Data-Hackathon.git>

