UIDAI Data Hackathon 2026: Project Report

Aadhaar Activity Insights 2025: Enrollment and Update Patterns Across India

Team Name: StatStrom
UIDAI Team ID: 3285
Team Members:

Member 1: Shravan Raut
Member 2: Viraj Shivale
Member 3: Mayur Kharche
Member 4: Prathmesh Tupkari

January 20, 2026

Project Repository:
https://github.com/Shravanraut2005/UIDAI-HACKATHON-2026-

Contents

1 Problem Statement and Approach

1.1 Problem Statement :

Unlocking Societal Trends in Aadhaar Enrolment and Updates and Identify meaningful patterns, trends, anomalies, or predictive indicators and translate them into clear insights or solution frameworks that can support informed decision-making and system improvements.

1.2 Strategic Approach :

To solve this problem, we followed a structured data analytics approach where the entire Aadhaar lifecycle from new registrations to demographic updates was treated as a continuous process.

Instead of only looking at overall totals, we focused on how the data behaves across:

- different age groups,
- different regions and states,
- and monthly timelines.

By doing this, we were able to understand where enrolment is happening the most, where updates are more frequent, and what these patterns suggest about real-world Aadhaar usage. This approach also helps to highlight early signs of uneven workloads or unusual patterns so that corrective action can be planned in advance.

2 Datasets Used

The analysis utilizes two primary datasets representing Aadhaar operations in the year 2025:

- **Aadhaar Enrolment Dataset (2025):** Nowhere near uniform, enrollment data splits by region first. Then it divides again - area codes shape how numbers show up. Time frames matter too; each slice covers a different stretch of years. Local rules add another level on top. Think preschool beginnings, then kids growing older. Teens come into view, followed by grown adults later on. Where people live shifts the totals noticeably. Cities differ from rural spots in clear ways. Layers stack up instead of sitting side by side. Patterns emerge when you look beyond surface groups.

- **Aadhaar Demographic Update Dataset (2025):** Starting young, the process sticks close from age five onward. Even when people grow past their teens, it does not fade away. Moving forward, each step gets matched with steady backing. Though most effort shows between five and seventeen, it never fully lets go. As years pass, what began simply keeps adjusting. When someone turns eighteen, the thread still holds. Without making noise about it, staying involved matters over time.

3 Methodology

The project was completed using a step-by-step data science pipeline:

1. **Data Cleaning:** The given data was large (close to one million records) and required proper cleaning before analysis. We ensured that:

   - State/region names were standardized,
   - Missing and inconsistent values were handled, and
   - The data structure was made uniform for smooth processing.

   For handling large-scale processing efficiently, we used Polars along with Python libraries, because it performs faster for bigger datasets.

2. **Dataset Integration:** After cleaning, we combined and aligned both datasets so that enrolment and update activities could be compared in a meaningful way. This integration made it possible to generate combined insights and observe the Aadhaar lifecycle as one complete process.

3. **Exploratory Data Analysis (EDA):**In this phase, we explored the dataset to understand:

   - **Top contributing states:** Identifying regions with the highest volume of records.
   - **Category-wise patterns:** Analyzing how data is distributed across different classifications.
   - **Overall trends:** Observing fluctuations and patterns over the relevant time period.

   EDA helped us identify the states where enrolments and updates were heavily concentrated and also helped in deciding what Key Performance Indicators (KPIs) and visuals would be most useful for the final dashboard.

4. **KPI Calculation:**To convert the data into clear measurable insights, we calculated several critical KPIs:

   - **Total Enrolments:** The cumulative count of new entries across India.
   - **Total Demographic Updates:** The total volume of modifications made to existing records.
   - **Highest Enrolment State:** Identification of the region with the maximum new registrations.
   - **Highest Update State:** Identification of the region with the maximum maintenance activity.
   - **Update-to-Enrolment Ratio:** A metric used to determine the primary nature of system activity.

   These indicators facilitate a comparative analysis of the workload across different states. Specifically, they help determine whether the system is in a *growth phase* (characterized by higher enrolments) or a *maintenance phase* (characterized by a higher volume of updates).

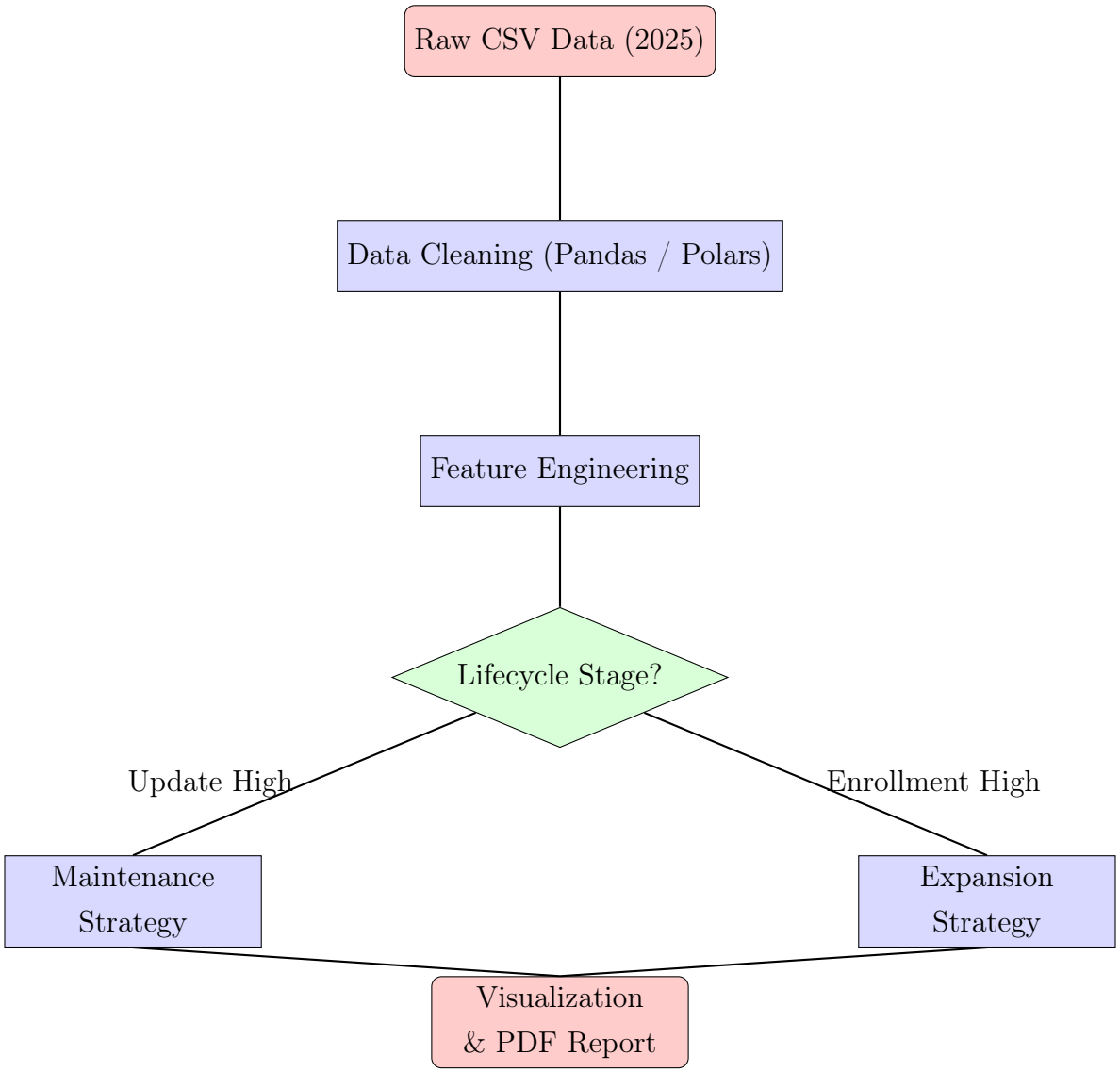
   KPI DASHBOARD (2025)

   Total Enrolments (India): 2057960
   Total Demographic Updates (India): 36597559
   Top Enrolment State: Uttar Pradesh -> 338863
   Top Updates State: Uttar Pradesh -> 6460511
   Enrollment Share Age 0–5
   Update Share Age 17+

   The above KPI values are the final outcomes we got after analyzing the dataset. After cleaning and studying the enrolment and demographic update records for 2025, these numbers helped us clearly understand the overall activity in India. They also show which state had the highest workload and which age groups contributed the most. Overall, these KPIs give a simple and clear picture of whether the system is seeing more fresh enrolments or more update-related activity.

3.1 System Architecture

Our system architecture follows a modular and well-organized flow where raw data is transformed into insights:



This architecture ensures that the project output is not only correct, but also scalable and easy to extend for future datasets.

4 Data Analysis and Visualization

4.1 Key Findings and Insight

From our combined analysis of both datasets, we observed the following important insights:

- **Child-Centric Enrollment:**A major portion of new Aadhaar enrolments is observed in the age group below 5 years, which indicates that Aadhaar registration is strongly linked with early-life identity creation and essential documentation needs.
- **Adult-Centric Maintenance:**Most demographic updates are performed by people aged 18 years and above. This suggests that Aadhaar maintenance activities mainly occur due to changes like relocation, corrections, or updates required in adulthood.
- **Geographic Concentration:** New enrolments and updates are not evenly distributed across India. A limited set of states contribute a major percentage of activity, showing workload concentration in specific regions.
- **Operational Shift:**Overall, the Aadhaar system shows a trend where the focus is moving from rapid expansion to long-term maintenance. This means the system requires more attention towards improving update processes, quality control, and workload balancing rather than only increasing enrolment numbers.

4.2 Top 10 States with Highest Enrollment in India (2025)

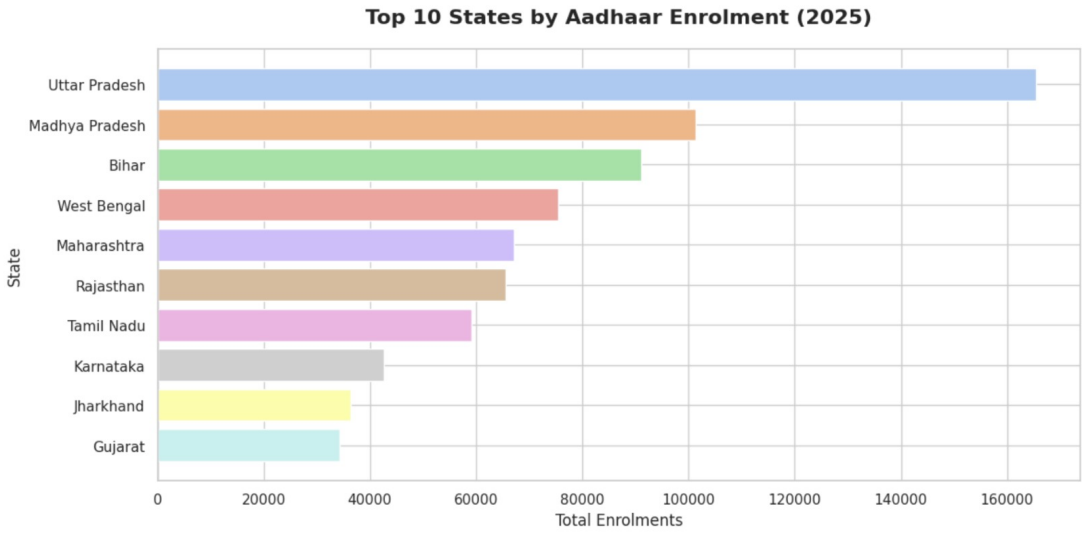


Figure 1: Top 10 States with Highest Enrollment India (2025)

4.3 Top 10 States with Total Demographic Updates

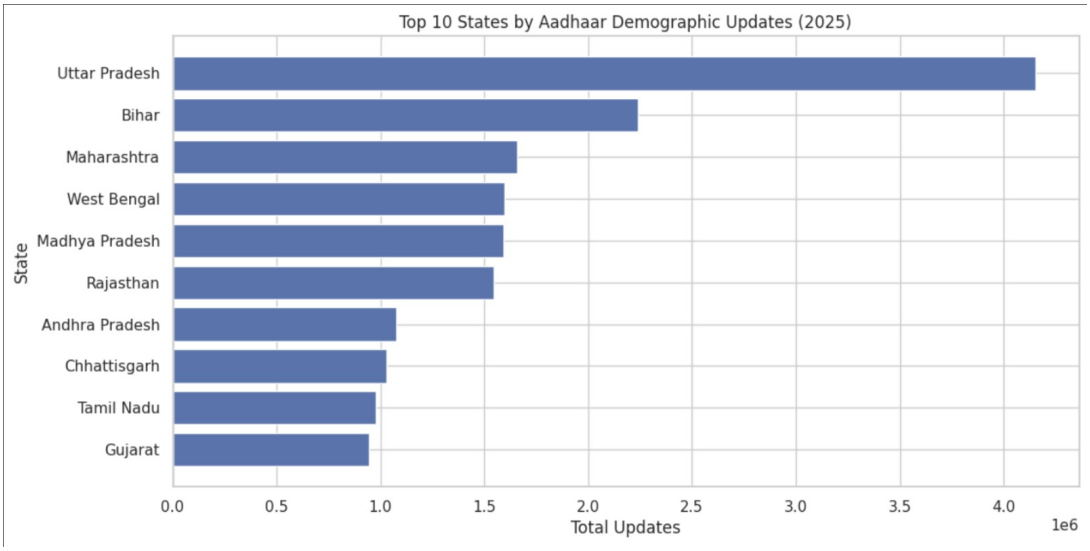


Figure 2: Top 10 States with total demographic updates

4.4 State-wise Enrollments and Update Summary

Table 1: State-wise Enrollment and Updates Summary

Index	State	Total Enrollment	Total Updates	Update/Enrollment Ratio
0	Uttar Pradesh	165,501	4,155,348	25.11
1	Madhya Pradesh	101,324	1,593,785	15.73
2	Bihar	91,172	2,239,927	24.57
3	West Bengal	75,519	1,597,004	21.15
4	Maharashtra	67,107	1,658,750	24.72

4.5 Monthly Update and Enrollment Analysis

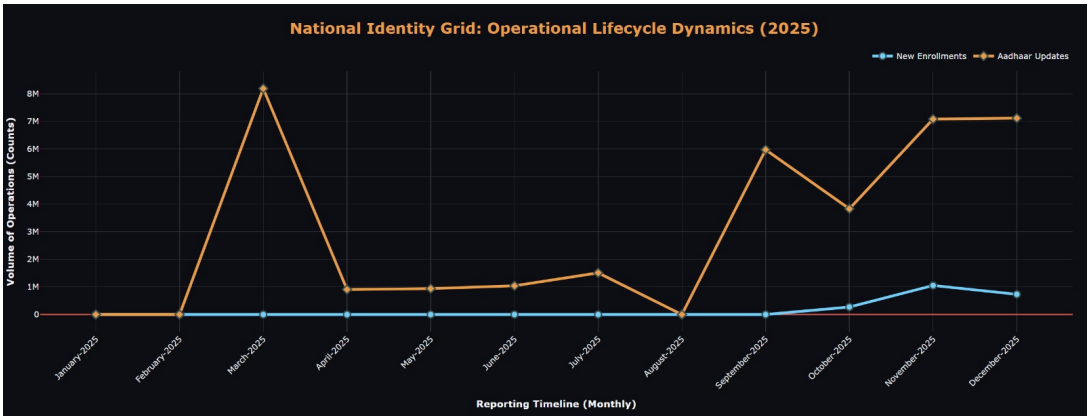


Figure 3: Monthly Update and Enrollment Analysis

4.6 Project Dashboard

We developed a centralized dashboard that integrates major **KPIs**, state-based insights, and lifecycle trends into a single interface. This implementation facilitates a faster understanding of:

- **High-Enrolment Regions:** Identifying geographical areas with the most significant user acquisition.
- **High-Update States:** Highlighting regions requiring frequent data maintenance and modifications.
- **Key Operational Indicators:** Monitoring the overall health and efficiency of the data processing pipeline.

The dashboard serves as a *quick reference tool* designed for strategic planning and automated reporting, reducing the time required to derive actionable insights from the raw data.

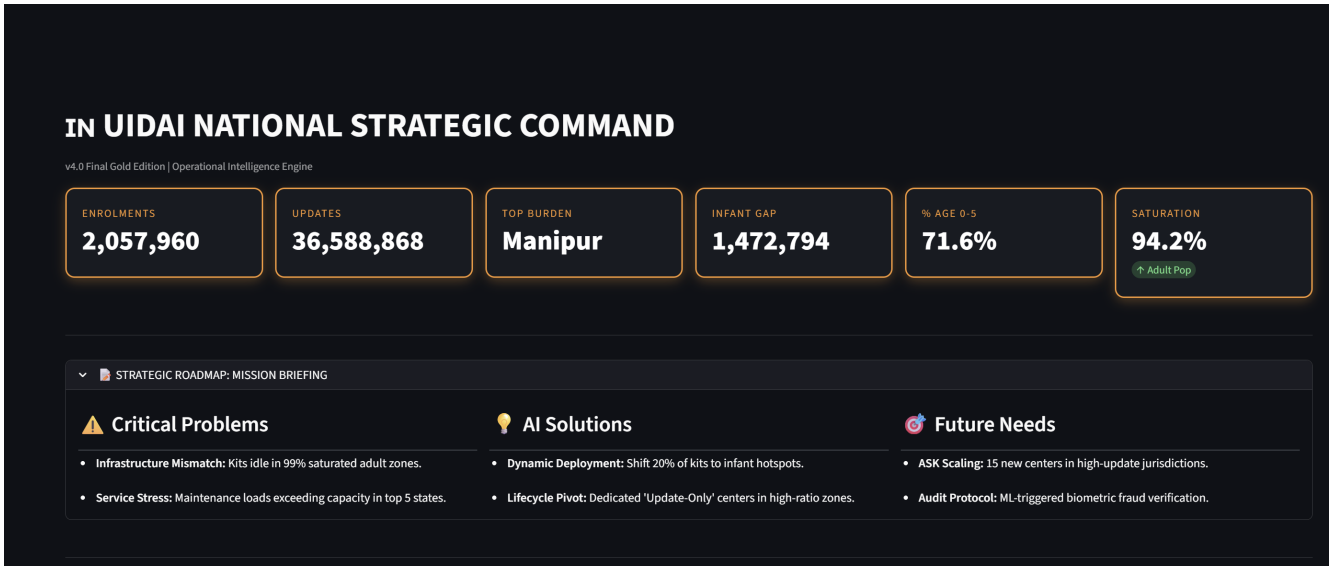


Figure 4: Main Dashboard

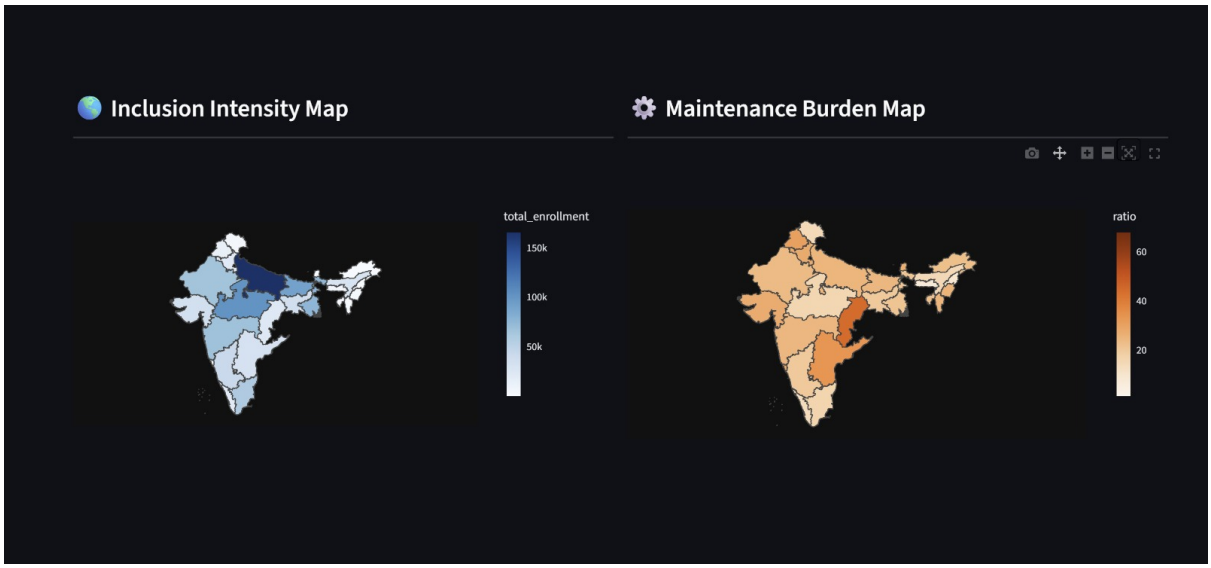


Figure 5: Inclusion Intensity and Maintainance Burden Map

Figure 6: Top 10 Priority States

5 Conclusion

From this project, it is clear that Aadhaar enrolment and demographic updates follow different patterns across age groups and regions. While new enrolments are mostly driven by young children, demographic updates are largely driven by adults, meaning the Aadhaar ecosystem is now more focused on long-term maintenance.

Also, since workload is concentrated in certain states, UIDAI can benefit from state-level targeted planning. By using such data-driven insights, service delivery can be improved through better resource allocation, stronger monitoring, and more efficient update processes.