UIDAI Data Hackathon 2026

Unlocking Societal Trends in Aadhaar Enrolment and Updates

Team ID:- UIDAI_7198

Team Members:-

1. Aaron Rao
2. Abhinand Meethele Valappil
3. Shreyansh Arora
4. Aditi Karn
5. Garv Randhar

---

1. Problem Statement and Approach

1.1 Problem Statement

Aadhaar enrolment and update behaviour varies significantly across regions and age groups due to population mobility, biometric stability, urbanisation, and administrative reach. Understanding these variations is critical for improving service delivery, infrastructure planning, and targeted interventions.

This project analyses anonymised Aadhaar enrolment, biometric update, and demographic update data to identify patterns, anomalies, and service pressure across Indian states and districts. The goal is to translate these insights into an explainable decision-support framework to help UIDAI identify high service-stress regions and prioritise administrative action.

1.2 Approach Overview

The project follows a structured data-driven pipeline:

1. Integration of anonymised enrolment and update datasets
2. Data cleaning, validation, and standardisation
3. Multi-level exploratory analysis
4. Development of an explainable Service Stress Indicator
5. Deployment through an interactive decision-support dashboard

2. Datasets Used

Only anonymised and aggregated UIDAI datasets were used, strictly following hackathon guidelines.

2.1 Dataset Description

Aadhaar Enrolment Dataset

- date, state, district, pincode
- age_0_5, age_5_17, age_18_greater

Aadhaar Biometric Update Dataset

- date, state, district, pincode
- bio_age_5_17, bio_age_17_

Aadhaar Demographic Update Dataset

- date, state, district, pincode
- demo_age_5_17, demo_age_17_

Multiple CSV files per dataset were programmatically merged for analysis.

3. Methodology

3.1 Data Cleaning and Preprocessing

Key preprocessing steps included:

- Dataset consolidation
- Date standardisation
- Removal of invalid state entries
- State name standardisation
- District- and state-level aggregation
- Creation of derived metrics (total enrolments, updates)

These steps ensured analytical consistency and reliability.

3.2 Feature Engineering

To enable fair cross-district comparison:

- **Biometric Update Ratio** = Biometric Updates / Enrolments
- **Demographic Update Ratio** = Demographic Updates / Enrolments

These ratios form the foundation of the Service Stress Indicator.

---

4. Data Analysis and Visualisation

4.1 Univariate Analysis

4.1.1 Age-wise Enrolment Distribution

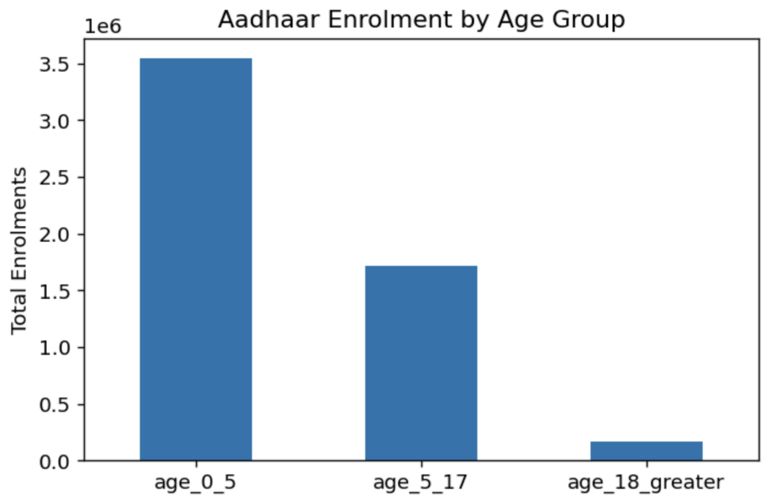

Figure 1: Age-wise distribution of Aadhaar enrolments

Key observations:

- Enrolment activity is concentrated among younger age groups
- Adult enrolment shows near-saturation
- Ongoing enrolments largely reflect births and children reaching enrolment age

4.1.2 Update Type Comparison

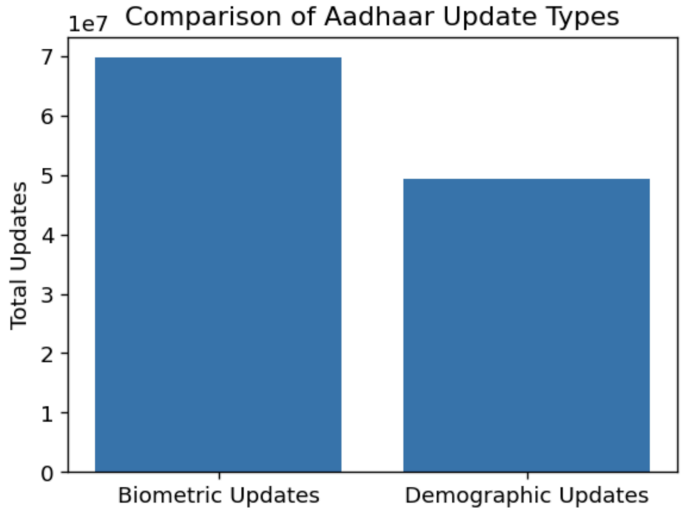

Figure 2: Biometric vs demographic update volumes

Key observations:

- Biometric updates exceed demographic updates by ~40%
- Indicates higher biometric maintenance demand, especially among children
- Highlights importance of biometric update infrastructure

4.2 Bivariate Analysis

4.2.1 Enrolment vs Biometric Updates

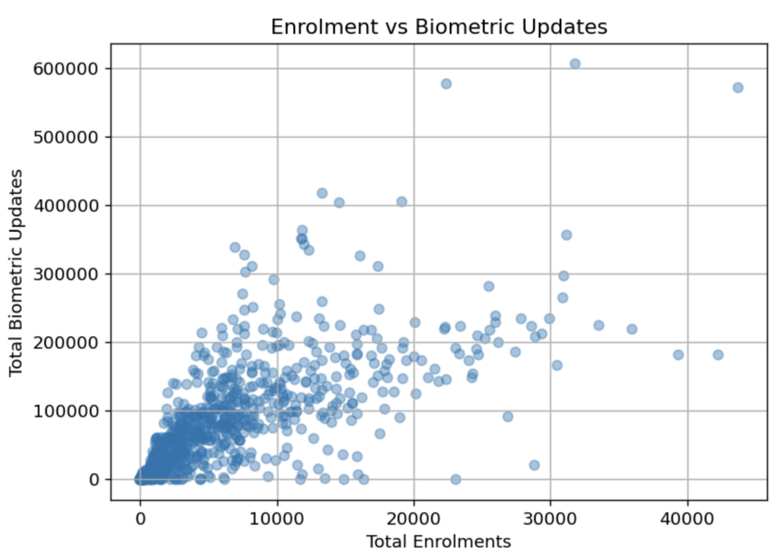

Figure 3: Enrolment vs biometric updates by district

Key insights:

- Positive correlation exists, but with significant outliers
- Several districts show disproportionately high update intensity
- Indicates service demand not explained by enrolment size alone

4.2.2 Enrolment Saturation vs Service Stress

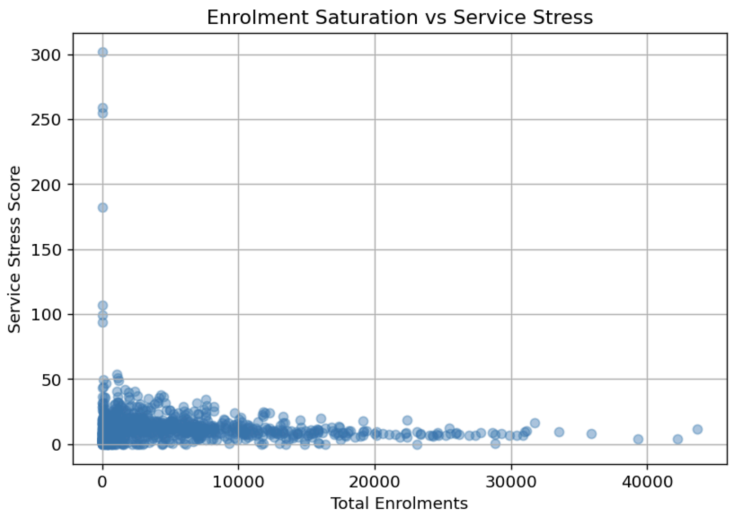

Figure 4: Enrolment vs service stress scores

Key insights:

- Inverse relationship observed
- Low-enrolment districts often experience highest stress
- Validates need for normalised metrics

4.2.3 Biometric vs Demographic Update Intensity

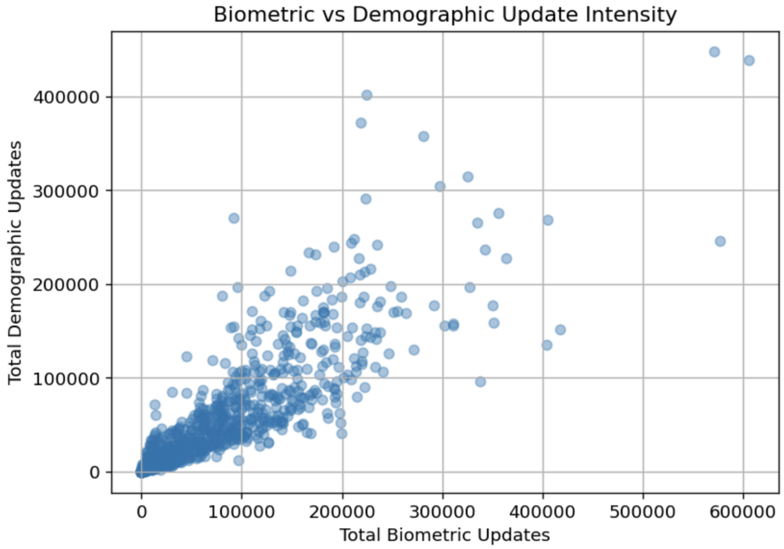

Figure 5: Biometric vs demographic updates

Key insights:

- Strong linear correlation
- Districts face simultaneous pressure across both update types
- Infrastructure must support both update categories

4.3 State-Level Analysis

4.3.1 Within-State Stress Variability

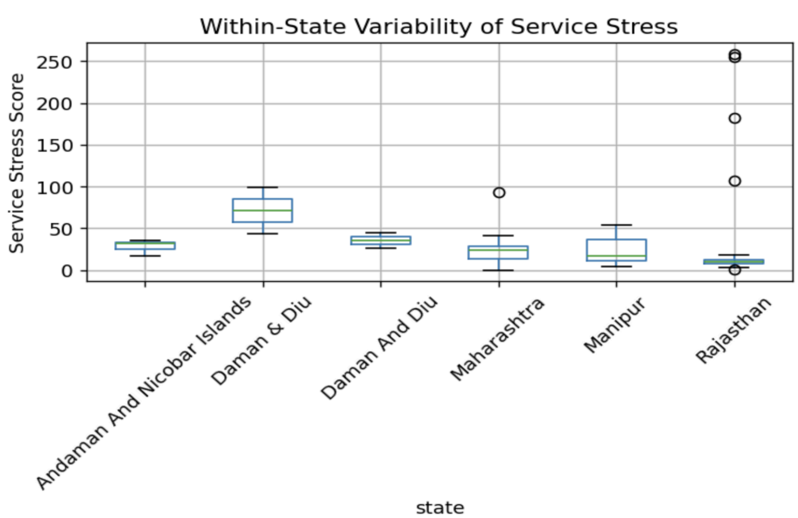

Figure 6: Stress score distribution within states

Key insights:

- Union territories show higher variability and extreme outliers
- Larger states show generally low stress with isolated hotspots
- Suggests targeted, not blanket, interventions

4.3.2 Top States by Average Stress

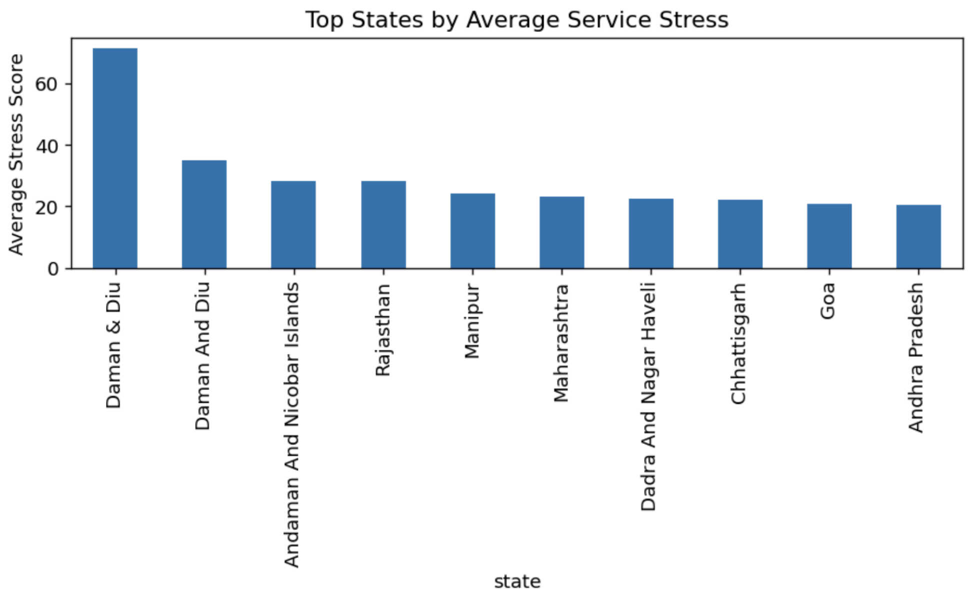

Figure 7: Top 10 states by average stress

Key insight:

- Union territories and smaller states dominate high-stress rankings
- Tourism, migration, and geographic constraints contribute to stress

4.4 District-Level Analysis

4.4.1 Stress Score Distribution

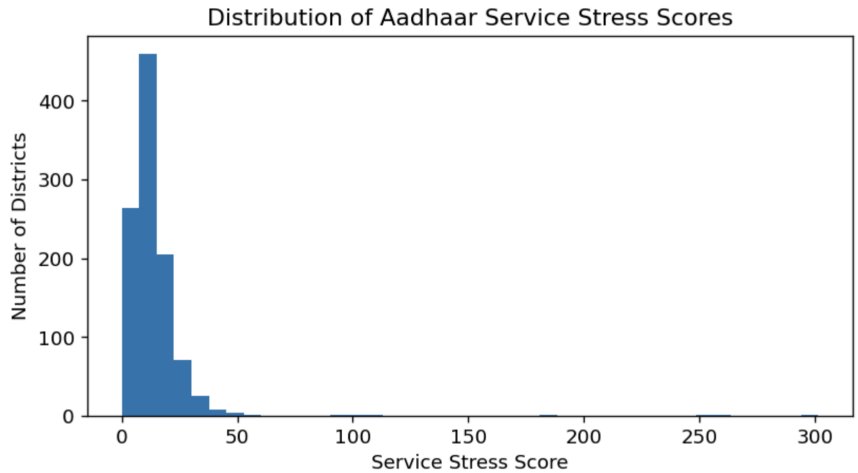

Figure 8: Stress score histogram

Key insights:

- Right-skewed distribution
- ~85% districts operate under low stress
- Small fraction experience extreme stress (>100)

4.4.2 Top High-Stress Districts

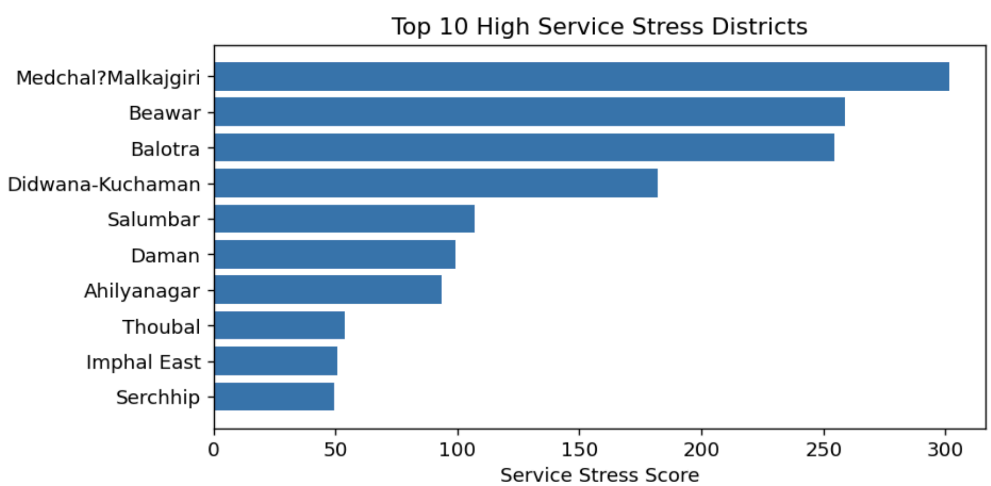

Figure 9: Highest stress districts

Key insights:

- Urbanisation-driven stress (e.g., Medchal-Malkajgiri)
- Rajasthan dominates high-stress districts
- Northeastern and UT regions show persistent challenges

4.5 Key Analytical Insights Summary

- Adult enrolment near-saturation
- Biometric updates dominate service demand
- Urban and peri-urban districts show elevated stress
- Population mobility is a major stress driver
- Stress concentrated in limited but critical districts

---

5. Proposed Working Model: Aadhaar Service Stress Indicator

5.1 Model Objective

Identify districts experiencing disproportionate Aadhaar service demand.

5.2 Model Design

Service Stress Score = 0.5 × Biometric Update Ratio + 0.5 × Demographic Update Ratio

The model is transparent, normalised, and action-oriented.

5.3 Classification

Score Range	Category	Recommended Action
0–25	Stable	Routine monitoring
25–50	Moderate Mobility	Enhanced monitoring
50–100	Service Stress	Infrastructure assessment
100+	Priority Intervention	Urgent action

6. Demonstration and User Interface

The model is deployed via an interactive Streamlit dashboard.

Key Capabilities

- State/district selection
- Real-time stress computation
- Visual stress indicators
- Downloadable PDF reports
- Comparative benchmarking

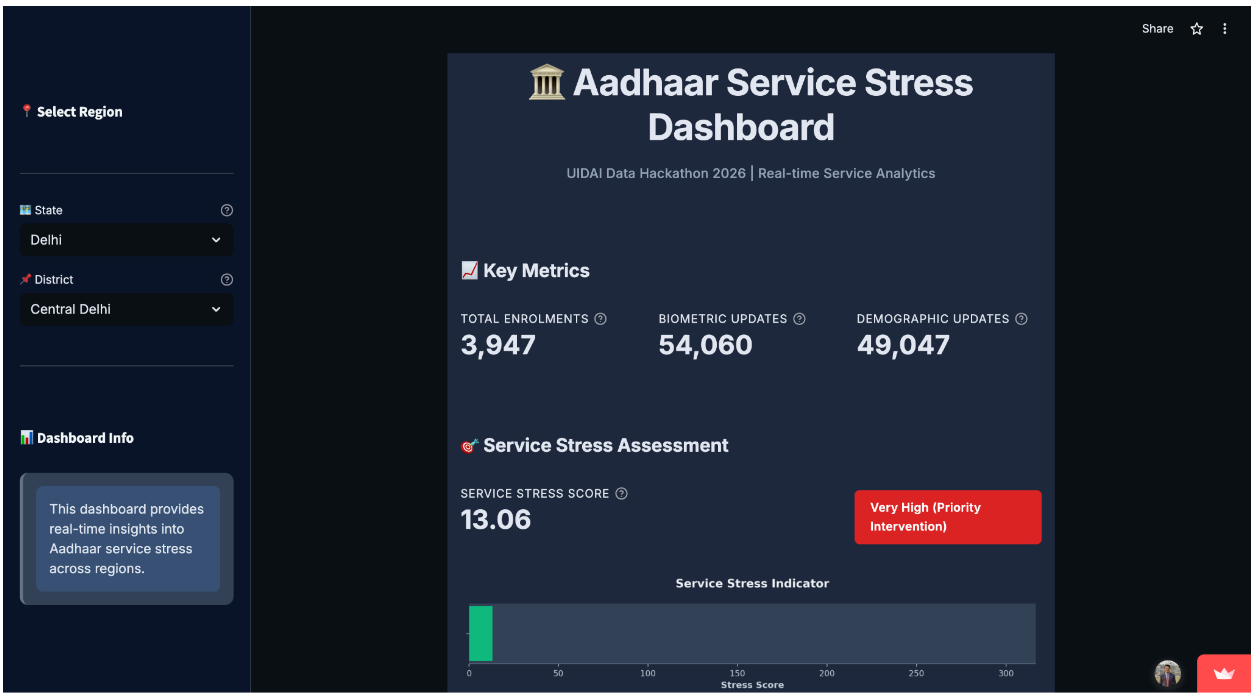

Figure 10: Dashboard showing district analytics and stress indicator

---

7. Impact and Applicability

Administrative Benefits

- Proactive infrastructure planning
- Objective resource prioritisation
- Early-warning detection of service stress

Scalability

- Time-series extension
- Predictive analytics
- Multi-indicator integration
- API and system integration

---

8. Conclusion

This project demonstrates how anonymised Aadhaar data can be transformed into actionable administrative intelligence. The Service Stress Indicator provides a transparent, scalable framework for identifying service pressure, enabling proactive planning, equitable resource allocation, and improved citizen experience.

8.1 Technology Stack

The solution was developed entirely using **open-source technologies**, ensuring transparency, reproducibility, and ease of future enhancement. The technology stack was selected to balance analytical rigor, explainability, and practical deployability.

Programming Language

- **Python**
  Used as the primary language for data processing, analysis, modelling, and application development due to its extensive ecosystem of data science libraries and strong community support.

Data Processing & Analysis

- **Pandas**
  Used for data ingestion, cleaning, transformation, aggregation, and feature engineering across large Aadhaar enrolment and update datasets.
- **NumPy**
  Used for numerical operations and efficient computation of derived indicators and ratios.

Data Visualisation

- **Matplotlib**
  Used to generate publication-quality visualisations, including trend plots, distributions, comparisons, and stress indicators, with explicit axis scaling for administrative readability.

Modeling & Analytics

- **Explainable Analytical Model**
  A transparent, rule-based service stress indicator was implemented using normalised ratios instead of black-box machine learning models to ensure interpretability and policy relevance.

User Interface & Interaction

- **Streamlit**
  Used to build an interactive, web-based dashboard allowing users to:
    - Select states and districts dynamically
    - View computed service stress scores
    - Interpret results through visual indicators
    - Generate and download region-specific reports

Reporting & Export

- **ReportLab**
  Used to generate structured, timestamped PDF reports programmatically for district-level and state-level summaries, enabling offline review and administrative sharing.

Deployment & Version Control

- **Git & GitHub**
  Used for version control, collaborative development, and deployment integration.
- **Streamlit Community Cloud**
  Used to deploy the interactive dashboard for live demonstration and evaluation.

Open-Source Compliance

- The entire solution is built using **open-source libraries only**, in compliance with hackathon guidelines.
- No proprietary tools, APIs, or external datasets were used.

Platform Compatibility

- The solution was developed and tested on **macOS (Apple Silicon)** and deployed on a **Linux-based cloud environment**, ensuring cross-platform compatibility.

Project Links

GitHub Repository: https://github.com/Aaronrao989/uidai-aadhaar-service-stress

Live Dashboard: https://uidai-aadhaar-service-stress.streamlit.app/
