UIDAI DATA HACKATHON-2026

Project Title:
Aadhaar Service Stress Index (ASSI):
A Data-Driven Framework to Identify and Predict Aadhaar Enrolment Pressure Across Indian States.

Team _ID: UIDAI_9177

- ## Abstract:

The Aadhaar ecosystem serves as a foundational digital identity infrastructure for residents across India. Managing Aadhaar enrolment services efficiently requires timely understanding of enrolment demand and operational pressure across regions. However, current approaches largely rely on retrospective analysis of enrolment volumes, limiting proactive planning of enrolment infrastructure.

This project proposes the **Aadhaar Service Stress Index (ASSI)**, a composite, data-driven index designed to quantify and forecast enrolment-related service stress across Indian states and districts. Using a large-scale anonymised Aadhaar enrolment dataset, the study incorporates enrolment volume, demographic composition, growth trends, temporal volatility, and anomaly detection to assess regional service pressure. The index enables identification of high-risk regions and supports predictive planning of enrolment resources.

The proposed framework demonstrates how administrative data can be transformed into actionable intelligence, enabling UIDAI to shift from reactive service management to proactive, evidence-based decision-making.

- ## Problem Statement:

  Aadhaar enrolment demand varies significantly across regions and time due to demographic composition and growth patterns. Certain states and districts experience enrolment overload, leading to operational stress and inefficient service delivery. Currently, there is no unified quantitative measure to assess or predict enrolment-related service pressure across regions. This limits proactive planning of enrolment infrastructure and resource allocation.

  There is a need for a data-driven framework to quantify, compare, and forecast Aadhaar enrolment service stress to support informed administrative decision-making.

- ## Objectives:

  The primary objectives of this project are:

1. To analyze Aadhaar enrolment patterns across Indian states and districts.
2. To quantify enrolment-related service pressure using demographic and temporal factors.
3. To design a composite index (ASSI) that represents enrolment service stress on a standardized scale.
4. To identify regions with high current and emerging enrolment pressure.
5. To forecast future enrolment stress to support proactive administrative planning.
6. To derive actionable policy recommendations for improving enrolment service delivery.

- ## DATASET DESCRIPTION

The study uses an anonymised Aadhaar enrolment dataset provided for the UIDAI Data Hackathon. After consolidating the data provided across multiple files, the final analytical dataset consists of 212,595 enrolment records.

Key Characteristics:

- Geographic Granularity: State, District, Pincode
- Temporal Granularity: Date of enrolment activity
- Demographic Attributes:
    - Enrolments for age group 0–5 years
    - Enrolments for age group 5–17 years
    - Enrolments for age group 18 years and above

Column Name	Description
date	Date of enrolment activity
state	State name
district	District name
pincode	Enrolment location pincode
age_0_5	Enrolments for ages 0–5
age_5_17	Enrolments for ages 5–17
age_18_greater	Enrolments for ages 18+

- ## MOTIVATION FOR THE STUDY

  Aadhaar enrolment demand is not solely driven by population size but also by demographic composition, temporal fluctuations, and regional growth patterns. Regions with moderate enrolment volumes may experience higher operational stress due to rapid growth, adult-heavy enrolments, or seasonal spikes.

  By introducing a structured index to capture these dynamics, this project aims to bridge the gap between raw enrolment data and administrative decision-making.

- # Methodology

1. Data Collection and Consolidation

The Aadhaar enrolment dataset was provided across multiple files representing enrolment activity at different geographic and temporal granularities. All files were programmatically loaded and consolidated into a single dataset to enable unified analysis. Column names were standardized and verified for schema consistency across files.

The consolidated dataset contains enrolment counts categorized by age groups (0–5, 5–17, and 18+), along with state, district, pincode, and date attributes. No personally identifiable information was present in the dataset, ensuring privacy compliance.

2. Data Preprocessing

Data preprocessing steps were applied to ensure analytical reliability and temporal consistency:

Date fields were parsed using explicit day-first formatting to handle regional date representations.

Missing or invalid numeric values in enrolment fields were handled through appropriate imputation or exclusion.

Enrolment counts were converted to numeric format to ensure accurate aggregation and computation.

Records with insufficient historical context (e.g., first observation periods causing undefined growth metrics) were excluded from predictive modeling.

3. Feature Engineering

To transform raw enrolment data into operational stress indicators, several derived features were constructed:

3.1 Total Enrolments

Total enrolments were calculated as the sum of enrolments across all age groups. This represents the baseline enrolment demand at a given location and time period.

3.2 Age-Weighted Enrolment Load

Different age groups impose varying levels of operational effort during Aadhaar enrolment. To capture this, an age-weighted enrolment load was computed by assigning higher weights to adult enrolments (18+) and moderate weights to child enrolments. This feature reflects demographic-adjusted service pressure rather than raw volume alone.

3.3 Temporal Aggregation

Enrolment data was aggregated at a monthly district level to analyze trends over time and reduce daily noise. This aggregation enabled meaningful computation of growth rates, volatility, and long-term stress patterns.

4. Stress Component Construction

Four key stress components were derived to capture different dimensions of enrolment pressure:

- Growth Rate: Measures month-to-month change in enrolment demand, indicating emerging stress even when absolute volumes are moderate.
- Volatility: Represents the standard deviation of enrolment volumes over time, capturing operational uncertainty and planning difficulty.
- Anomaly Indicator: Isolation Forest–based anomaly detection was used to identify sudden enrolment surges, such as those caused by migration, special enrolment drives, or administrative campaigns.
- Age-Weighted Load: Captures direct operational effort based on demographic composition.

5. Aadhaar Service Stress Index (ASSI) Construction

The Aadhaar Service Stress Index (ASSI) was formulated as a composite index combining all stress components into a single interpretable score.

All stress components were normalized using Min-Max scaling to ensure comparability.

Weights were assigned based on operational relevance rather than purely statistical optimization, preserving interpretability for administrative decision-makers.

The final ASSI score ranges from 0 to 100, where higher values indicate greater enrolment service stress.

Based on ASSI values, regions were categorized into Low, Medium, and High stress levels to support actionable decision-making.

6. Stress Prediction and Forecasting

To enable proactive planning, a Random Forest regression model was used to predict future ASSI values based on historical stress components.

A time-aware train–test split was applied to ensure realistic forecasting.

Model performance was evaluated using Mean Absolute Error (MAE) and R^2 metrics.

Feature importance analysis was conducted to maintain model explainability.

Predicted ASSI values were used to identify emerging high-risk regions requiring early administrative intervention.

- **Results & Analysis**

1. Overview of ASSI Distribution
   The Aadhaar Service Stress Index (ASSI) was computed for all districts across monthly time periods. The resulting ASSI values span a wide range, indicating substantial variation in enrolment service stress across regions.
   The distribution of ASSI scores shows that while a majority of regions operate under low to moderate stress, a smaller but significant subset consistently experiences high service stress. This confirms that enrolment pressure is not uniformly distributed and cannot be addressed through uniform resource allocation.

2. Regional Variation in Service Stress
   Analysis of average ASSI scores across districts reveals strong geographic disparities in enrolment-related service pressure. Certain districts exhibit persistently high ASSI values over time, suggesting sustained operational overload rather than temporary fluctuations.
   These high-stress districts are not always those with the highest enrolment volumes, highlighting that service stress is influenced by growth dynamics, demographic composition, and volatility, in addition to absolute enrolment counts.

3. Demographic Contribution to Stress
   Age-weighted enrolment load emerged as a dominant contributor to service stress. Districts with a higher proportion of adult (18+) enrolments tend to exhibit higher ASSI values, reflecting increased operational effort associated with biometric capture and data correction.
   This finding demonstrates that demographic composition plays a critical role in determining service stress, and that raw enrolment numbers alone are insufficient for effective planning.

4. Identification of High-Stress Regions
   Based on the defined ASSI thresholds, districts were classified into Low, Medium, and High stress categories. A focused subset of districts consistently falls into the High Stress category, marking them as priority areas for administrative intervention.
   These regions represent critical points where additional enrolment centers, staff redistribution, or temporary facilities may be required to maintain service quality.

5. Predictive Insights and Emerging Risk Zones

Using historical stress components, future ASSI values were predicted to identify emerging high-risk regions. The predictive analysis indicates that several districts currently operating under medium stress are likely to transition into high stress in subsequent periods if current trends persist.

This predictive capability enables UIDAI to move from reactive responses to proactive enrolment infrastructure planning, reducing service disruptions and citizen inconvenience.

6. Key Findings Summary

The following key findings emerge from the analysis:

- Aadhaar enrolment service stress varies significantly across regions and time.
- Demographic composition, particularly adult enrolments, strongly influences operational pressure.
- Growth rate and volatility are critical indicators of emerging stress.
- High service stress is concentrated in a limited number of districts, enabling targeted intervention.
- Predictive modeling provides early warning signals for future service overload.

- # Conclusion & Future Scope

1. Conclusion

This project introduced the **Aadhaar Service Stress Index (ASSI)** as a unified, data-driven framework to quantify and forecast enrolment-related service pressure across Indian states and districts. By moving beyond raw enrolment counts and incorporating demographic composition, growth trends, volatility, and anomaly detection, the proposed index provides a more accurate representation of operational stress on Aadhaar enrolment services.

The analysis demonstrates that enrolment service stress is unevenly distributed across regions and time, with a limited number of districts experiencing consistently high pressure. The predictive component further highlights emerging risk zones, enabling a shift from reactive service management to proactive planning. Overall, ASSI serves as an interpretable and scalable decision-support tool that can assist UIDAI in improving enrolment efficiency and citizen experience.

2. Future Scope

While the current study focuses on enrolment data, the framework can be extended in several ways:

- Integration of Aadhaar update and correction datasets to capture full service load.
- Incorporation of population and migration indicators for enhanced forecasting accuracy.
- Development of real-time dashboards for continuous monitoring of service stress.
- Expansion of the index to include infrastructure and staffing capacity metrics.
- These extensions can further strengthen the applicability of ASSI as a comprehensive governance analytics tool.