UIDAI DATA HACKATHON 2026

1. INTRODUCTION

India's Aadhaar program is one of the largest digital identity systems in the world, playing a crucial role in enabling access to government services, financial inclusion, and digital governance. Aadhaar enrolment and update data provide valuable insights into population coverage, service reach, and regional disparities.

With increasing reliance on Aadhaar for welfare delivery, banking, healthcare, and education, it becomes essential to understand enrolment patterns and update trends across states, genders, and age groups. Data-driven analysis of Aadhaar statistics can help policymakers identify gaps, improve service infrastructure, and ensure inclusive access.

This project focuses on analysing UIDAI's open datasets to uncover meaningful trends, patterns, and insights related to Aadhaar enrolment and updates across India.

2. PROBLEM STATEMENT

Unlocking Societal Trends in Aadhaar Enrolment and Updates

Despite widespread Aadhaar adoption, variations exist across regions and demographic groups in terms of enrolment and update activity. Understanding these variations is essential for optimizing enrolment centers, awareness programs, and policy interventions.

Problem Statement:

To analyse Aadhaar enrolment and update data to identify regional, gender-wise, and temporal trends, and translate these findings into actionable insights that can support informed decision-making and system improvements.

3. OBJECTIVES OF THE STUDY

The key objectives of this analysis are:

- To study state-wise Aadhaar enrolment distribution across India.
- To analyse age-wise enrolment trends and identify inclusion patterns.
- To examine year-wise enrolment growth and identify long-term trends.
- To identify regions with comparatively low enrolment that may require focused intervention.
- To derive insights that can help improve Aadhaar service delivery and outreach programs.

4. DATASET USED

The analysis uses Aadhaar Monthly Enrolment datasets published by the **Unique Identification Authority of India (UIDAI)** and made available through the **Open Government Data (OGD) Platform – data.gov.in.**

Dataset Details:

- **Dataset Name**: Aadhaar Enrolment Statistics
- **Source**: UIDAI, Government of India
- **Format**: CSV
- **Coverage:** State-wise, age-wise, and year-wise enrolment data

Key Columns Used:

The dataset is publicly available and ensure data authenticity and reliability.

- Date – Monthly enrolment reference date
- State – State or Union Territory name
- District – District-level enrolment information
- Pincode – Pincode-level enrolment area
- Age_0_5 – Enrolments for age group 0-5 years
- Age_5_17 – Enrolments for age group 5-17 years
- Age_18_greater – Enrolments for age group 18 years and above

Total enrolment was derived by aggregating enrolments across all age groups.

5. Methodology:

A systematic data analysis approach was followed to extract meaningful insights from the dataset.

5.1 Data Collection

The Aadhaar enrolment dataset was downloaded in CSV format from the **data.gov.in** portal.

5.2 Data Cleaning and Preprocessing

- Missing and null values were identified and removed.
- Column names were standardized for consistency.
- Data types were verified to ensure numerical accuracy.

5.3 Data Transformation

- Aggregation operations were performed using group-by functions.
- State-wise, age-wise, and year-wise totals were calculated.
- Age-wise enrolment value were aggregated to compute total enrolments for analysis.

5.4 Tools and Technologies Used

- **Google Colab / Jupyter Notebook**
- **Python Programming Language**
- **Libraries:** Pandas, Matplotlib

6. DATA ANALYSIS AND VISUALIZATION

All visualizations were generated using python libraries in Google Colab.

Dataset Preview

```
<class 'pandas.core.frame.DataFrame'>
Index: 620149 entries, 0 to 206710
Data columns (total 7 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   date            620149 non-null  object
 1   state           620149 non-null  object
 2   district        620149 non-null  object
 3   pincode         620149 non-null  int64
 4   age_0_5         620148 non-null  float64
 5   age_5_17        620148 non-null  float64
 6   age_18_greater  620146 non-null  float64
dtypes: float64(3), int64(1), object(3)
memory usage: 37.9+ MB
```

	pincode	age_0_5	age_5_17	age_18_greater
count	620149.000000	620148.000000	620148.000000	620146.000000
mean	519908.601928	3.868362	2.056333	0.221095
std	206142.037265	22.041287	18.162051	4.008257
min	679.000000	0.000000	0.000000	0.000000
25%	363530.000000	1.000000	0.000000	0.000000
50%	518345.000000	2.000000	0.000000	0.000000
75%	700138.000000	3.000000	1.000000	0.000000
max	855456.000000	2688.000000	1812.000000	855.000000

Fig1: Preview of UIDAI Aadhar enrolment dataset

6.1 State-wise Aadhaar Enrolment Analysis

State-wise analysis reveals significant variation in Aadhaar enrolment across India. States with larger populations and better digital infrastructure show higher enrolment figures, while certain regions reflect comparatively lower enrolment.

Insight:

High enrolment concentration indicates strong outreach, whereas lower enrolment may suggest the need for additional awareness and infrastructure support.

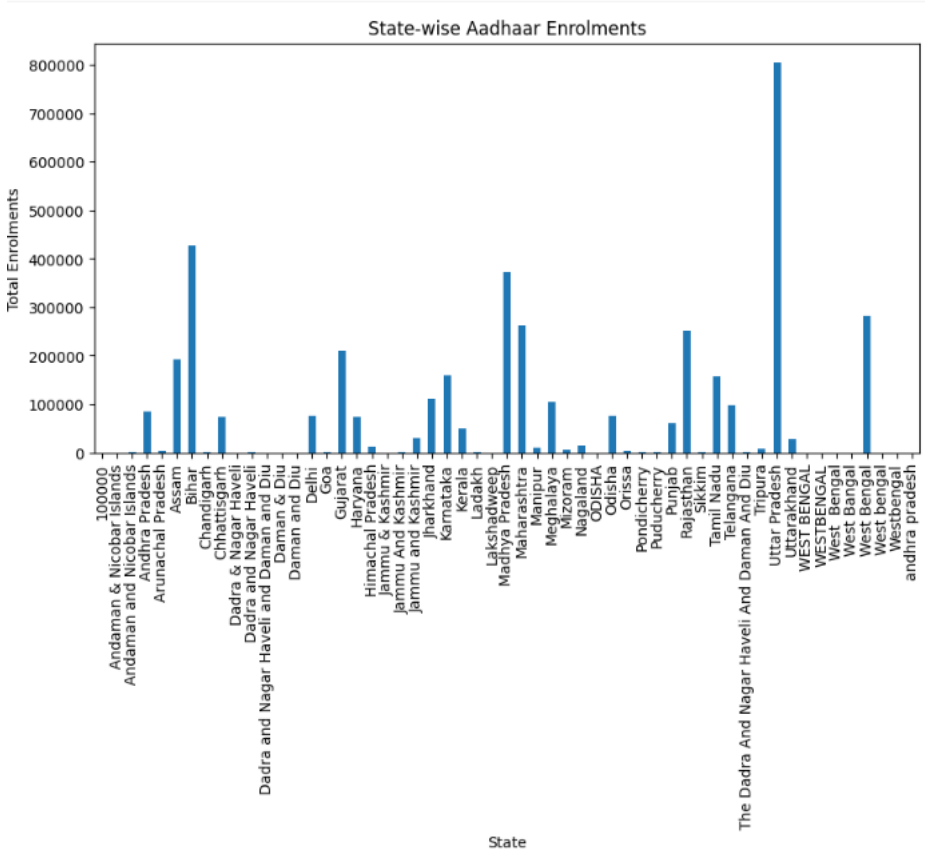

Fig2: State-wise Aadhaar enrolment distribution

6.2 Age-wise Aadhaar Enrolment Analysis

The dataset categorizes enrolments into three age groups: 0–5 years, 5–17 years, and 18 years and above. Analysis shows that the 18+ age group accounts for the highest proportion of enrolments, reflecting adult participation in Aadhaar updates and enrolment services.

Insight:

High enrolment in the adult age group indicates the role of Aadhaar in financial services, welfare access, and identity verification.

6.3 Year-wise Aadhaar Enrolment Trend

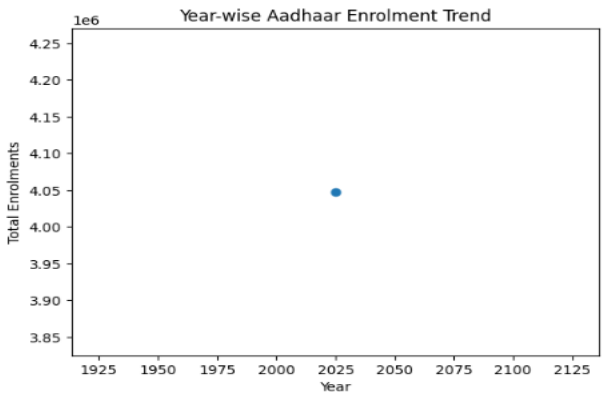

Fig3: Year-wise Aadhaar enrolment trend(2025)

The dataset corresponds to a single year (2025). Therefore, year-wise aggregation results in a single data point representing total enrolments for that year. This confirms the correctness of temporal aggregation while highlighting the dataset's time scope limitation. Future analysis with multi-year data can reveal long-term enrolment trends.

7. KEY FINDINGS AND INSIGHTS

- Aadhaar enrolment varies significantly across states and regions.
- Year-wise trends indicate sustained growth in Aadhaar adoption.
- Regions with lower enrolment may benefit from targeted awareness campaigns and improved enrolment infrastructure.

8. SOCIETAL AND POLICY IMPACT

The insights derived from this analysis can assist UIDAI and policymakers in:

- Optimizing the allocation of enrolment centers.
- Designing region-specific awareness programs.
- Improving update services for demographic changes.
- Strengthening inclusive governance and digital access.

This data-driven approach supports evidence-based policymaking and helps enhance the effectiveness of Aadhaar-related initiatives.

9. LIMITATIONS OF THE STUDY

- The analysis is limited to available open datasets.
- Real-time or district-level data was not included.
- Predictive modeling was limited due to dataset granularity.

10. CONCLUSION

This study demonstrates the importance of analysing Aadhaar enrolment and update data to uncover meaningful trends and patterns. By leveraging open government data and simple analytical techniques, the project provides valuable insights that can support system improvements and inclusive governance.

Data-driven analysis such as this can play a critical role in strengthening India's digital identity ecosystem and ensuring that Aadhaar services reach all sections of society effectively.

11. FUTURE SCOPE

- District-level and central-level analysis
- Advanced predictive modeling for enrolment forecasting
- Integration with socio-economic indicators

“This project reflects the potential of open government data in enabling inclusive, data-driven governance and informed decision-making.”