

Final Project Submission

AadhaarPulse Index

A Composite Health Score for Aadhaar Enrolment System

Submitted for:

UIDAI Data Hackathon 2026



GOVERNMENT OF INDIA



Team ID: UIDAI_3226

Participant Name: Sanoj Yadav

Platform: Open Government Data Platform India (data.gov.in)

Date of Submission: 20 January 2026

“Transforming Aadhaar enrolment data into a policy driven health intelligence system”

2. Problem Statement and Approach

Introduction

Aadhaar enrolment is a critical component of India's digital identity infrastructure. It enables access to essential government services, financial inclusion, and welfare delivery. However, current evaluation of Aadhaar enrolment performance is largely based on total enrolment numbers, which only reflect scale and not system quality or operational effectiveness.

Problem Statement

Total enrolment volume alone cannot capture whether the enrolment system is balanced across age groups, stable over time, or improving consistently. Large enrolment numbers may hide operational inefficiencies, demographic gaps, or unstable enrolment patterns. There is no unified metric that represents the overall health of the Aadhaar enrolment system at the state level.

Proposed Solution

This project introduces the AadhaarPulse Index, a composite health score designed to evaluate Aadhaar enrolment performance beyond raw volume. The index combines three critical dimensions of enrolment quality: growth trends, age group balance, and enrolment stability. By transforming complex enrolment data into a single interpretable score, AadhaarPulse Index provides a meaningful measure of enrolment system health.

Approach Overview

The approach consists of:

- Cleaning and standardizing UIDAI Aadhaar enrolment datasets
- Aggregating enrolment data at the state level
- Designing three analytical indicators: Growth Score, Age Coverage Score, and Stability Score
- Combining them into a composite AadhaarPulse Index
- Ranking states based on enrolment system health rather than size

Project Objectives

- To move beyond volume based evaluation of Aadhaar enrolment
- To introduce a composite health score for enrolment systems
- To identify high performing and risk prone states
- To support data driven governance and targeted policy interventions

3. Dataset Used and Methodology

1. Dataset Description

The project uses UIDAI Aadhaar enrolment datasets obtained from the Open Government Data Platform India. The datasets contain detailed enrolment information across states and age groups.

- Total Records: 1,006,029
- Data Type: State wise, time wise and age group wise enrolment data

2. Files Used

- api_data_aadhaar_enrolment_0_500000.csv
- api_data_aadhaar_enrolment_500000_1000000.csv
- api_data_aadhaar_enrolment_1000000_1006029.csv

These files were merged to form a single consolidated master dataset.

3. Key Columns Used

The following columns were directly used for analysis and index construction:

- state – for state wise aggregation and ranking
- age_0_5 – for child enrolment coverage
- age_5_17 – for adolescent enrolment coverage
- age_18_greater – for adult enrolment coverage

Derived analytical columns created during processing:

- total_enrolment – sum of all age groups
- Growth_Score – enrolment trend indicator
- Age_Coverage_Score – demographic balance indicator
- Stability_Score – operational consistency indicator
- AadhaarPulse_Index – composite performance index
- Rank – final state ranking

4. Data Cleaning and Preprocessing

- Removal of missing and invalid values
- Standardization of state names
- Conversion of data types into numerical format
- Aggregation of enrolment values at the state level
- Creation of a total enrolment column

This ensured consistency, accuracy, and reliability of the dataset.

5. Feature Engineering

- Growth Score – Measures the enrolment trend over time to identify growth, stagnation, or decline
- Age Coverage Score – Measures balance among enrolments in the 0 to 5, 5 to 17, and 18 plus age groups
- Stability Score – Measures consistency of enrolment patterns and reflects operational reliability.

6. AadhaarPulse Index Formula

$$\text{AadhaarPulse Index} = 0.40 \times \text{Growth Score} + 0.35 \times \text{Age Coverage Score} + 0.25 \times \text{Stability Score}$$

This weighted formulation reflects the relative importance of growth, inclusiveness, and stability.

7. Tools and Technologies

- Programming Language: Python,
- Data Processing: Pandas
- Numerical Computation: NumPy
- Visualization: Matplotlib
- Development Environment: Python 3.13, Windows OS

4. Data Analysis and Visualisation

Technical Data Processing and Index Construction

1. Data Ingestion and Integration

Multiple UIDAI Aadhaar CSV datasets were loaded and merged into a single master dataset to ensure full coverage and reproducibility.

```
dfs = []
for f in files:
    dfs.append(pd.read_csv(f))
master_df = pd.concat(dfs, ignore_index=True)
```

2. Advanced Data Cleaning and Standardisation

Dates were parsed, missing values handled, and textual inconsistencies removed. State and district names were normalised using mapping logic.

```
master_df["date"] = pd.to_datetime(master_df["date"], errors="coerce")
age_cols = ["age_0_5", "age_5_17", "age_18_greater"]
master_df[age_cols] = master_df[age_cols].fillna(0)
master_df["state"] = master_df["state"].astype(str).str.strip().str.lower()
master_df["state"] = master_df["state"].replace(state_corrections)
```

3. Feature Engineering

New analytical variables were created to support modelling and time-series analysis.

```
master_df["total_enrolment"] = (
    master_df["age_0_5"] + master_df["age_5_17"] + master_df["age_18_greater"]
)
master_df["month"] = master_df["date"].dt.to_period("M")
```

4. State-Level Aggregation

Data was aggregated at the state level to convert transactional records into performance indicators.

```
state_summary = master_df.groupby("state").agg({
    "total_enrolment": "sum",
    "age_0_5": "sum",
    "age_5_17": "sum",
    "age_18_greater": "sum"
}).reset_index()
```

5. Age Coverage Score Design

Measures demographic balance across age groups using deviation from ideal distribution.

```
ideal = 1/3
deviation = abs(p1-ideal)+abs(p2-ideal)+abs(p3-ideal)
score = max(0, 100 - deviation*150)
```

6. Growth Score Computation

Evaluates enrolment momentum using month-wise growth trends.

```
growth_rate = temp["total_enrolment"].pct_change().mean()
score = np.clip((growth_rate+1)*50,0,100)
```

7. Stability Score Computation

Measures operational consistency using coefficient of variation.

```
cv = std/mean
score = np.clip(100 - cv*100,0,100)
```

8. AadhaarPulse Index Construction

A weighted composite index combining growth, inclusiveness, and stability.

```
state_summary["AadhaarPulse_Index"] = (
    0.40*state_summary["Growth_Score"] +
    0.35*state_summary["Age_Coverage_Score"] +
    0.25*state_summary["Stability_Score"]
)
```

9. Ranking Mechanism

States ranked in descending order of AadhaarPulse Index.

```
state_summary = state_summary.sort_values("AadhaarPulse_Index", ascending=False)
state_summary["Rank"] = range(1,len(state_summary)+1)
```

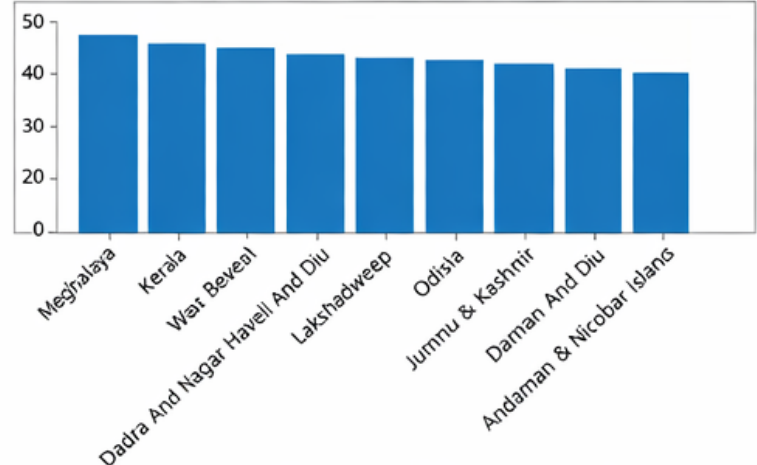
10. Output Generation

Final dataset exported for reporting and visualisation.

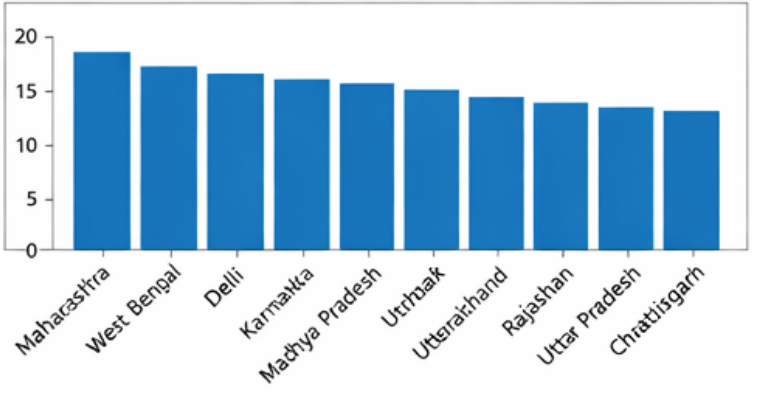
```
state_summary.to_csv("AadhaarPulse_Index_Statewise.csv", index=False)
```

4. Data Analysis and Visualisation

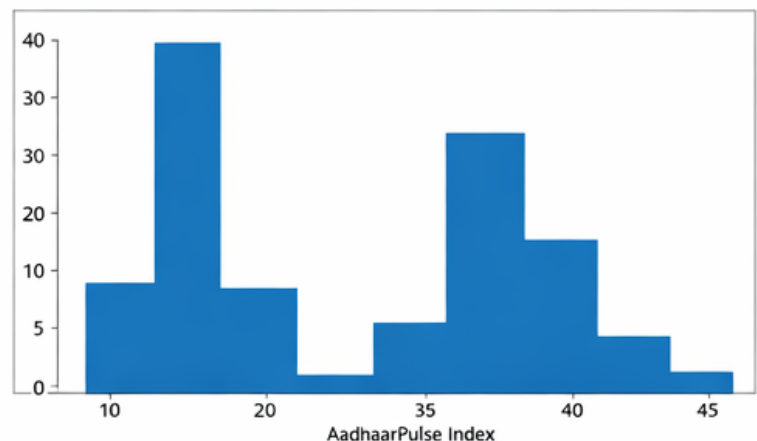
Top 10 States by AadhaarPulse Index



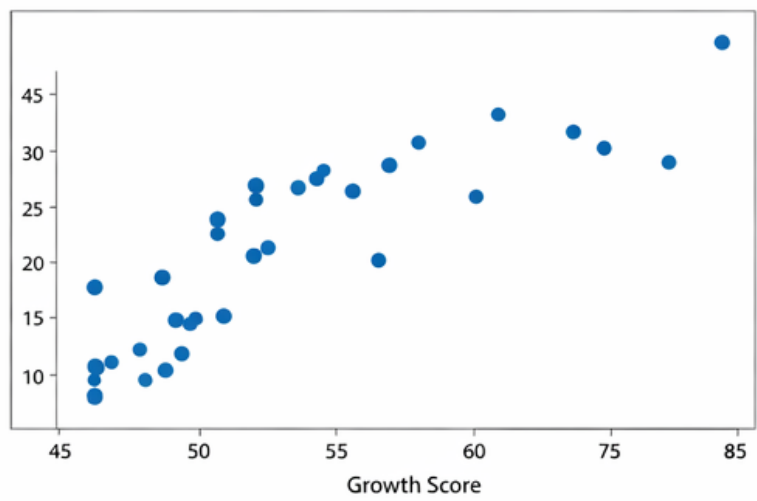
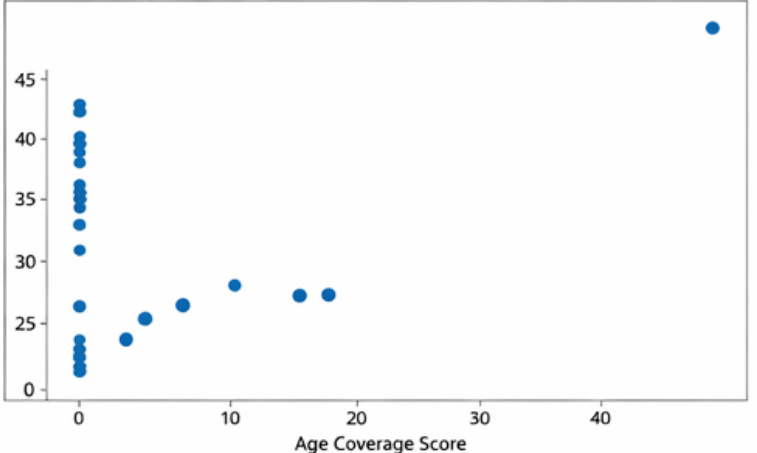
Bottom 10 States by AadhaarPulse Index



Distribution of AadhaarPulse Index Across States



Age Coverage Score vs AadhaarPulse Index



5. Analytical Insights and Practical Impact

Key Analytical Insights

- Growth Score is the strongest driver of AadhaarPulse Index. States with consistent positive enrolment trends automatically achieve higher overall system health.
- States with low Stability Score consistently appear in the bottom cluster, reflecting operational inconsistency and weak enrolment system reliability.
- High total enrolment volume does not guarantee high system quality. Several large states show lower AadhaarPulse Index due to imbalance and instability.
- Age Coverage Score directly measures demographic inclusiveness. Balanced enrolment across age groups significantly improves system performance.
- AadhaarPulse Index successfully separates quality based evaluation from volume based evaluation of enrolment systems.

Model Intelligence Summary

The AadhaarPulse Index is not a ranking by size. It is a ranking by system health. It integrates growth performance, operational stability and demographic inclusiveness into a unified analytical framework, making it a true composite performance model for Aadhaar enrolment systems.

Practical Impact and Real World Applicability

- AadhaarPulse Index can be integrated into UIDAI performance monitoring dashboards as a system health indicator.
- Low performing states can be automatically classified into risk categories for early intervention.
- Targeted enrolment drives can be planned for states showing instability or demographic imbalance.
- Yearly AadhaarPulse Index comparison enables national level benchmarking of enrolment system health.
- The index supports evidence based policy making and data driven resource allocation.
- This framework can evolve into an early warning system for enrolment system degradation.

Project Significance

This project demonstrates how raw government enrolment data can be transformed into an intelligent performance monitoring system using data engineering, analytical modelling and visual analytics. It converts raw data into actionable intelligence for governance improvement.

Final Note

The AadhaarPulse Index framework provides a scalable, data driven and policy ready solution to evaluate and strengthen India's Aadhaar enrolment ecosystem beyond simple enrolment numbers.