

AadhaarPulse – Unlocking Societal Trends in Aadhaar Enrolment and Updates

1. Problem Statement and Approach

Aadhaar enrolment and update datasets capture how citizens interact with India's digital identity infrastructure. Currently, this data is largely used for reporting counts rather than understanding societal behavior. This project aims to unlock hidden trends and anomalies from Aadhaar enrolment, demographic update, and biometric update data to support policy planning, infrastructure optimization, and proactive service delivery.

2. Datasets Used

The analysis uses Aadhaar datasets provided by UIDAI as part of the hackathon. All datasets were supplied in CSV format.

- 1 Aadhaar Enrolment Dataset: date, state, district, pincode, age-group enrolment counts
- 2 Aadhaar Demographic Update Dataset: date, state, district, pincode, age-group demographic updates
- 3 Aadhaar Biometric Update Dataset: date, state, district, pincode, age-group biometric updates

3. Methodology

- 1 Merged all dataset segments into consolidated datasets.
- 2 Removed duplicate records and standardized geographic fields.
- 3 Validated date formats and PIN codes.
- 4 Analyzed zero and missing value patterns.
- 5 Aggregated data by state, district, and PIN code.

4. Data Analysis and Visualisation

The analysis identified that Aadhaar enrolment is driven by semi-urban regions, while biometric and demographic updates are dominated by children aged 5–17. Sudden spikes and zero-activity days indicate operational stress points. Visualisations include heatmaps, time-series trends, and age-wise distribution charts.

Code Snippet (Excerpt)

```
import pandas as pd

df = pd.read_csv("aadhaar_enrolment.csv")
df = df.drop_duplicates()
df['date'] = pd.to_datetime(df['date'], dayfirst=True)

district_summary = df.groupby(['state', 'district']).sum()
```