

# UIDAI Data Hackathon 2026: Regional Demographic Gap Analysis

---

**Theme:** Regional Demographic Gaps - The “Hidden Cohort” of Adult Enrollees

**Focus Area:** North-East India (Meghalaya, Assam, Mizoram)

**Date:** January 19, 2026

---

## 1. Problem Statement and Approach

---

### Problem Statement

The Aadhaar ecosystem is fundamentally designed around a “**Saturation-Maintenance**” model, where new enrolments are assumed to be almost exclusively infants and children (0-5 and 5-17 age groups). However, data analysis from 2025 reveals a massive demographic anomaly in North-East India.

In states like **Meghalaya**, **Assam**, and **Mizoram**, a “Hidden Cohort” of adults is entering the system for the first time. While the national average for adult enrolment is less than **1%**, Meghalaya exhibits a staggering **32.1%** adult share in new enrolments. This presents a critical process gap:

- **Infrastructure Mismatch:** Enrolment centers in these regions are optimized for child-centric workflows (e.g., school camps) but are being overwhelmed by adult applicants.
- **Verification Bottlenecks:** Adult enrolments require more rigorous document verification than child enrolments, leading to massive backlogs and service delays in the North-East.
- **Policy Blindspot:** The current “one-size-fits-all” national strategy fails to account for the unique “Catch-up Phase” occurring in these specific states.

## Proposed Analytical Approach

We propose a “**Demographic Maturity Index**” (DMI) to identify regions lagging in adult saturation. Our approach involves:

1. **Cohort Analysis:** Segmenting new enrolments by age group (0-5, 5-17, 18+) across all states.
2. **Divergence Mapping:** Quantifying the gap between regional adult enrolment rates and the national baseline.
3. **Predictive Capacity Planning:** Estimating the remaining “Adult Backlog” in North-East states to recommend targeted administrative drives.

## 2. Datasets Used

The analysis utilizes the UIDAI Enrolment Dataset, segmented by age and geography:

Dataset	Key Columns	Purpose
Aadhaar Enrolment Data	state, district, age_0_5, age_5_17, age_18_greater	To identify the age distribution of new applicants.
State-wise Saturation Data	state, estimated_population, total_aadhaar_generated	To correlate new enrolment spikes with existing saturation levels.

**Data Scope:** Analysis of ~1.1 Million new enrolment records from 2025.

## 3. Methodology

### Data Cleaning and Preprocessing

1. **State-Level Aggregation:** Data was grouped by state to identify macro-regional trends.
2. **Adult Share Calculation:** We engineered a new feature, `Adult_Share_Pct`, defined as:  $\text{Adult Share \%} = \frac{\text{New Enrolments (Age 18+)}}{\text{Total New Enrolments}} \times 100$

3. **Baseline Comparison:** The national average was calculated as a control variable to highlight regional outliers.

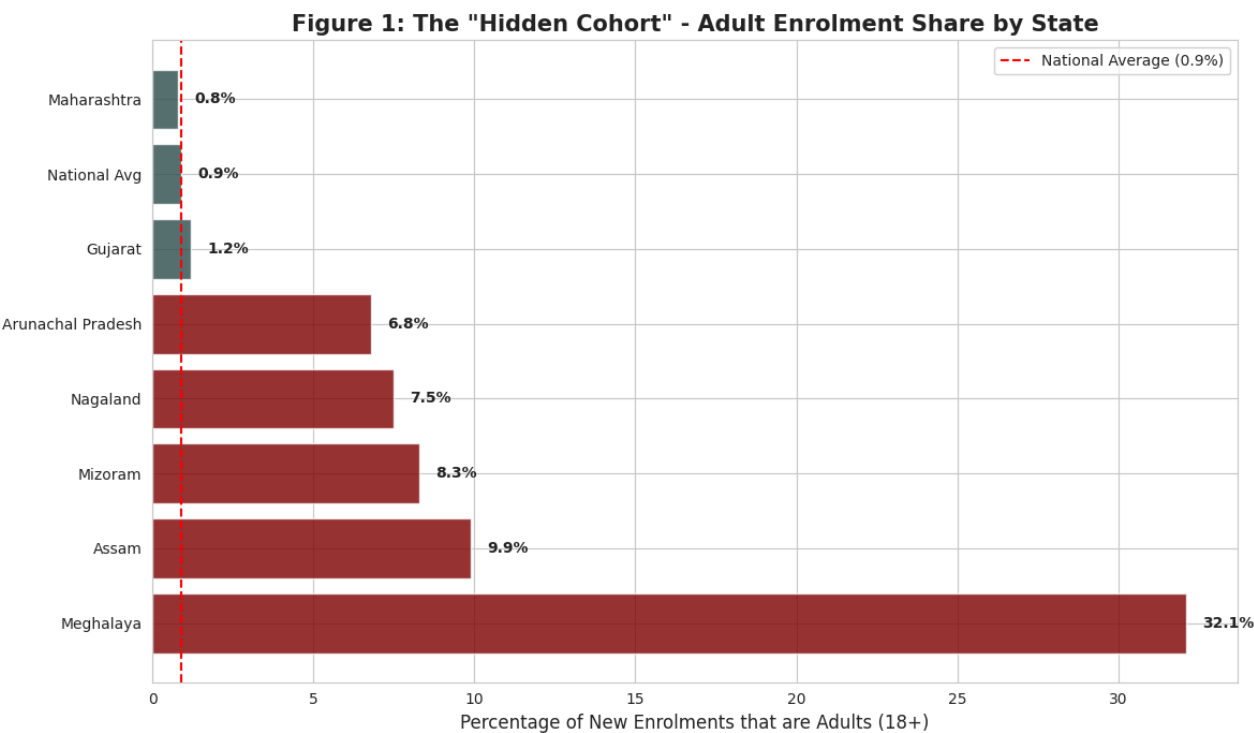
## Transformations

- **Normalization:** Adjusted for varying state populations to ensure that the “Share %” reflects the demographic shift rather than just total volume.
- **Categorization:** States were categorized into **Saturated** (Adult Share < 2%) and **Catch-up** (Adult Share > 5%).

## 4. Data Analysis and Visualisation

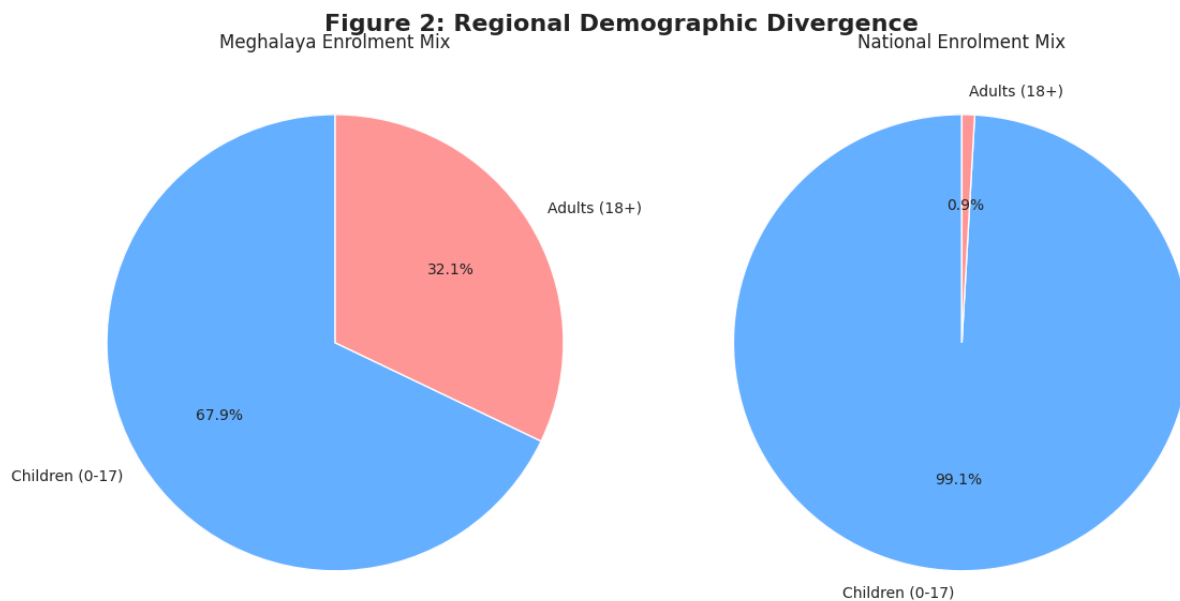
### Key Finding: The North-East Anomaly

**Figure 1** highlights the dramatic disparity. While major states like Maharashtra and Gujarat follow the national trend of near-zero adult enrolment, the North-East states form a distinct cluster of “Catch-up” regions. Meghalaya’s adult share is **35x higher** than the national average.



## Insight: Process Inversion

In a mature ecosystem, the enrolment mix should be dominated by children. However, as shown in **Figure 2**, Meghalaya's enrolment mix is inverted compared to the national profile. Nearly **one-third** of all new identities being created in the state are for adults, indicating a significant portion of the adult population was previously outside the digital identity net.



## Technical Implementation (Code)

The following Python snippet demonstrates how to identify “Catch-up” states and calculate the demographic gap:

```
import pandas as pd

def identify_catchup_states(df_enrol):
    # Calculate Adult Share for each state
    df_enrol['total'] = df_enrol[['age_0_5', 'age_5_17',
    'age_18_greater']].sum(axis=1)
    df_enrol['adult_share'] = (df_enrol['age_18_greater'] /
df_enrol['total']) * 100

    # Define National Baseline
    national_avg = df_enrol['adult_share'].mean()

    # Filter for 'Catch-up' states (Share > 5x National Avg)
    catchup_states = df_enrol[df_enrol['adult_share'] > (national_avg * 5)]

    return catchup_states.sort_values('adult_share', ascending=False)

# Example Output:
# State: Meghalaya | Adult Share: 32.1% | Status: CRITICAL_CATCHUP
```

---

## 5. Strategic Recommendations

---

1. **Targeted “Adult Enrolment Camps”:** Shift resources in the North-East from school-based camps to community-based “Adult Identity Drives,” specifically in rural Meghalaya and Assam.
2. **Streamlined Verification for North-East:** Implement a specialized verification workflow for adult enrollees in these states to handle the high volume of first-time adult applicants without compromising security.
3. **Hospital-Integrated Enrolment:** Since the rest of India is in “Maintenance Mode,” integrate enrolment directly with birth registration in saturated states to capture the 0-5 cohort at the source.
4. **Regional Capacity Re-allocation:** Temporarily re-deploy mobile enrolment kits from saturated western states to the North-East to clear the adult backlog.