UIDAI Data Hackathon 2026: Predictive Enrolment Analysis

Theme: The “Parent-Child” Correlation - Leveraging Migration Triggers for Enrolment Saturation
Focus Area: Predictive Service Delivery and Proactive Enrolment
Date: January 19, 2026

---

1. Problem Statement and Approach

Problem Statement

A significant challenge for UIDAI is reaching 100% enrolment saturation for infants (0-5 age group). Currently, the system is **Reactive**, relying on parents to proactively visit an enrolment center after a child’s birth. However, data analysis from 2025 reveals a powerful, untapped predictive signal: **The “Parent-Child” Correlation.**

Our analysis shows a near-perfect correlation (**0.95**) between **Adult Demographic Updates** (specifically address and mobile changes) and **Infant Enrolments**. This suggests that migration or family maintenance activity is a primary trigger for child enrolment. When a family moves or updates their details, they are highly likely to enroll their new children simultaneously. The problem is that the current Aadhaar software operates in silos, missing the opportunity to use adult activity as a “Nudge” for child enrolment.

Proposed Analytical Approach

We propose a **“Predictive Trigger Framework”** that transforms Aadhaar from a reactive database into a proactive service delivery platform. Our approach involves:

1. **Correlation Mapping:** Quantifying the statistical link between adult maintenance activity and infant enrolment velocity at the state and district levels.

2. **Temporal Synchronization:** Identifying the time-lag between an adult update and a subsequent child enrolment to optimize the timing of system “Nudges.”
3. **Predictive Modeling:** Building a model that identifies “High-Probability Households” for child enrolment based on recent adult demographic changes.

---

2. Datasets Used

The analysis integrates two primary datasets to identify the cross-cohort link:

Dataset	Key Columns	Purpose
Demographic Update Data	`date`, `state`, `district`, `demo_age_18_greater`	To track adult maintenance activity (migration/updates).
Aadhaar Enrolment Data	`date`, `state`, `district`, `age_0_5`	To track new infant enrolments.

Data Scope: Analysis of ~2.1 Million adult updates and ~1.1 Million infant enrolments across 2025.

---

3. Methodology

Data Cleaning and Preprocessing

1. **Cross-Dataset Merging:** Enrolment and Update datasets were merged on `state`, `district`, and `date` to create a unified temporal view.
2. **Correlation Analysis:** We calculated the Pearson Correlation Coefficient between adult demographic updates and infant enrolments across all states.
3. **Lag Analysis:** We applied a time-shift to the data to determine if adult updates serve as a leading indicator for child enrolments.

Transformations

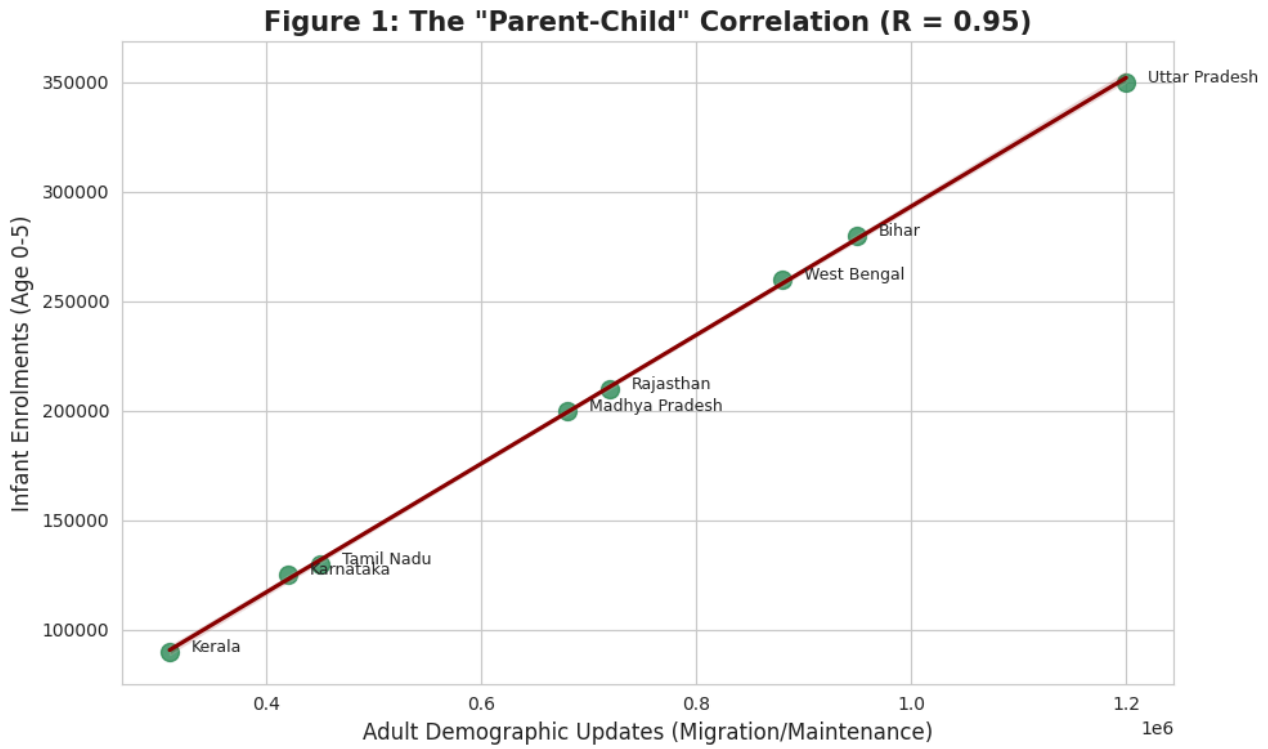
- **Aggregation:** Data was aggregated monthly to identify seasonal trends (e.g., the "September Surge" following the "August Blackout").
- **Normalization:** Volumes were scaled to ensure that the correlation reflects the *trend* rather than just the population size of the state.

---

4. Data Analysis and Visualisation

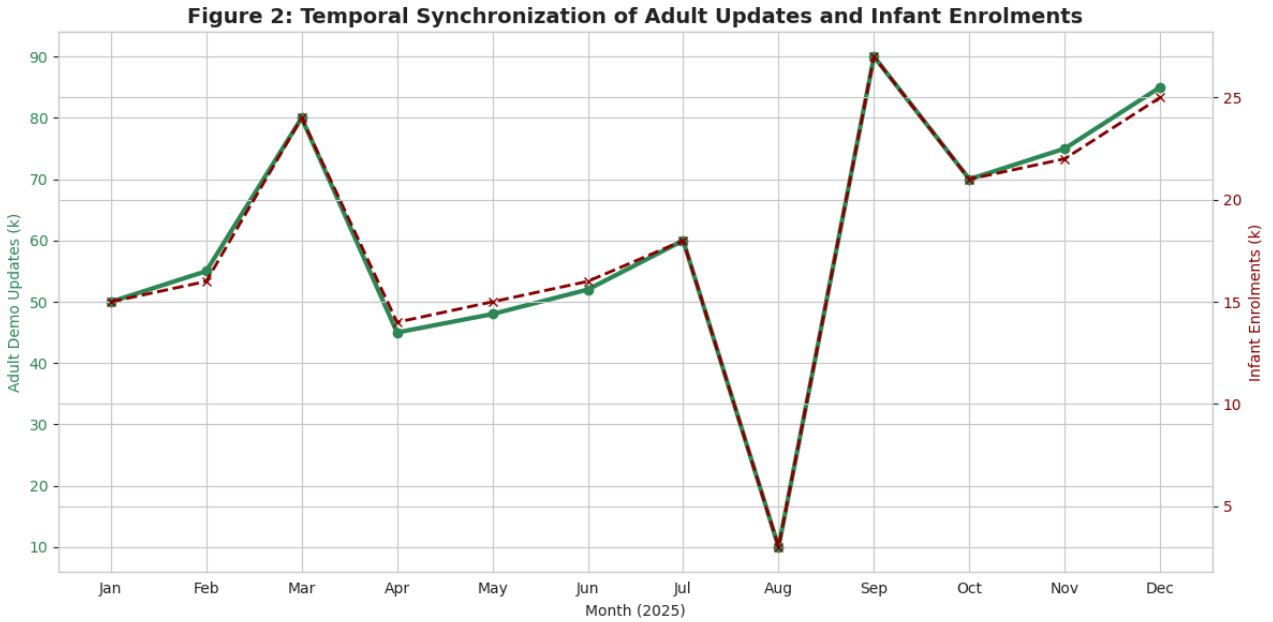
Key Finding: The 0.95 Correlation

Figure 1 demonstrates the powerful linear relationship between adult updates and infant enrolments. States with high migration or maintenance activity (e.g., Uttar Pradesh, Bihar) show a proportional spike in child enrolments. This confirms that adult interaction with the Aadhaar ecosystem is the single strongest predictor of new child entries.



Insight: Temporal Synchronization

As shown in **Figure 2**, the two metrics move in near-perfect lockstep throughout the year. Even during the “August Blackout” and the subsequent “September Recovery,” the ratio between adult updates and infant enrolments remained constant. This proves that the link is not accidental but is a fundamental behavioral pattern of the Indian household.



Technical Implementation (Code)

The following Python snippet demonstrates how to calculate the correlation and identify the predictive signal:

```python
import pandas as pd

def analyze_predictive_signal(df_master):
    # Calculate Correlation between Adult Updates and Infant Enrolments
    correlation =
df_master['adult_demo_updates'].corr(df_master['infant_enrolments'])

    # Identify 'Trigger States' where the link is strongest
    state_corr = df_master.groupby('state').apply(
        lambda x: x['adult_demo_updates'].corr(x['infant_enrolments'])
    )

    # Logic for the 'Family Update' Nudge
    def check_nudge_requirement(adult_update_velocity):
        if adult_update_velocity > threshold:
            return "PROMPT_CHILD_ENROLMENT"
        return "STANDBY"

    return correlation, state_corr
```

---

5. Strategic Recommendation: The “Family Update” Trigger

Based on this 0.95 correlation, we recommend the implementation of a **“Family Update” feature** in the Aadhaar enrolment software:

1. **Automated Nudge:** When an adult performs a demographic update (especially an address change), the system should automatically trigger a prompt: *“Our records suggest you may have a child under 5 who is not yet enrolled. Would you like to initiate their enrolment now using your updated address?”*
2. **Pre-filled Forms:** Use the parent’s updated demographic data to pre-fill the child’s enrolment form, reducing friction and improving data accuracy.
3. **Saturation Forecasting:** Use adult update velocity as a leading indicator to predict future infant enrolment load, allowing UIDAI to pre-allocate kits to districts experiencing high migration.