UIDAI Data Hackathon 2026: Border District Security Analysis

Theme: National Security Risks - Disproportionate Enrolment Velocity
Focus Area: Sensitive Border Districts
Date: January 19, 2026

---

1. Problem Statement and Approach

Problem Statement

As Aadhaar reaches near-universal saturation in India, the primary operational mode has shifted from new enrolments to demographic and biometric updates (maintenance). However, data analysis reveals a critical anomaly: **Sensitive border districts are exhibiting "Enrolment Velocity" that is 12x to 15x higher than inland urban centers.**

While urban centers like Pune or Jaipur show a balanced ratio where updates far outpace new enrolments (reflecting a mature, saturated population), districts along the Nepal and Bangladesh borders (e.g., Sitamarhi, Bahraich, Murshidabad) show a massive surge in new enrolments relative to maintenance activity. This "Velocity Divergence" suggests potential risks, including:

- **Migration Flux:** Unmonitored population movement across international borders.
- **Fraudulent Enrolment:** Potential attempts at bulk creation of identities in sensitive zones.
- **Administrative Lag:** A lack of "Velocity Checks" in the current system to flag abnormal growth.

Proposed Analytical Approach

We propose a **"Velocity-to-Maintenance Ratio" (VMR)** framework to detect these anomalies. By calculating the ratio of New Enrolments to Total Updates, we can distinguish between "Organic Growth" (births/naturalization) and "Anomalous Velocity." Our technical approach involves:

1. **Geospatial Clustering:** Grouping districts by security sensitivity (Border vs. Inland).
2. **Statistical Benchmarking:** Establishing a "National Saturation Baseline" for VMR.
3. **Anomaly Detection:** Flagging districts that exceed the baseline by more than 3 standard deviations (3σ).

---

2. Datasets Used

The analysis utilizes the following UIDAI-provided datasets, specifically filtered for the 2025-2026 period:

Dataset	Key Columns	Purpose
Aadhaar Enrolment Data	`date`, `state`, `district`, `age_18_greater`	To measure new adult identity creation in sensitive zones.
Demographic Update Data	`date`, `district`, `demo_age_18_greater`	To establish the baseline of "Maintenance Activity" for the existing population.
Biometric Update Data	`date`, `district`, `bio_age_18_greater`	To validate the authenticity and activity level of the existing user base.

Data Scope: Analysis of ~1.2 Million records across 50+ districts categorized as "Border" or "Control (Urban/Inland)."

---

3. Methodology

Data Cleaning and Preprocessing

1. **District Categorization:** Districts were tagged as `Border` or `Inland` based on official geospatial data.
2. **Normalization:** To account for population differences, we calculated the **Velocity Ratio**: $\$\text{Velocity Ratio} = \frac{\text{New Enrolments (Adults)}}{\text{Total Updates (Demographic + Biometric)}}\$$
3. **Temporal Alignment:** Data was aggregated monthly to smooth out daily fluctuations and identify sustained trends.

Transformations

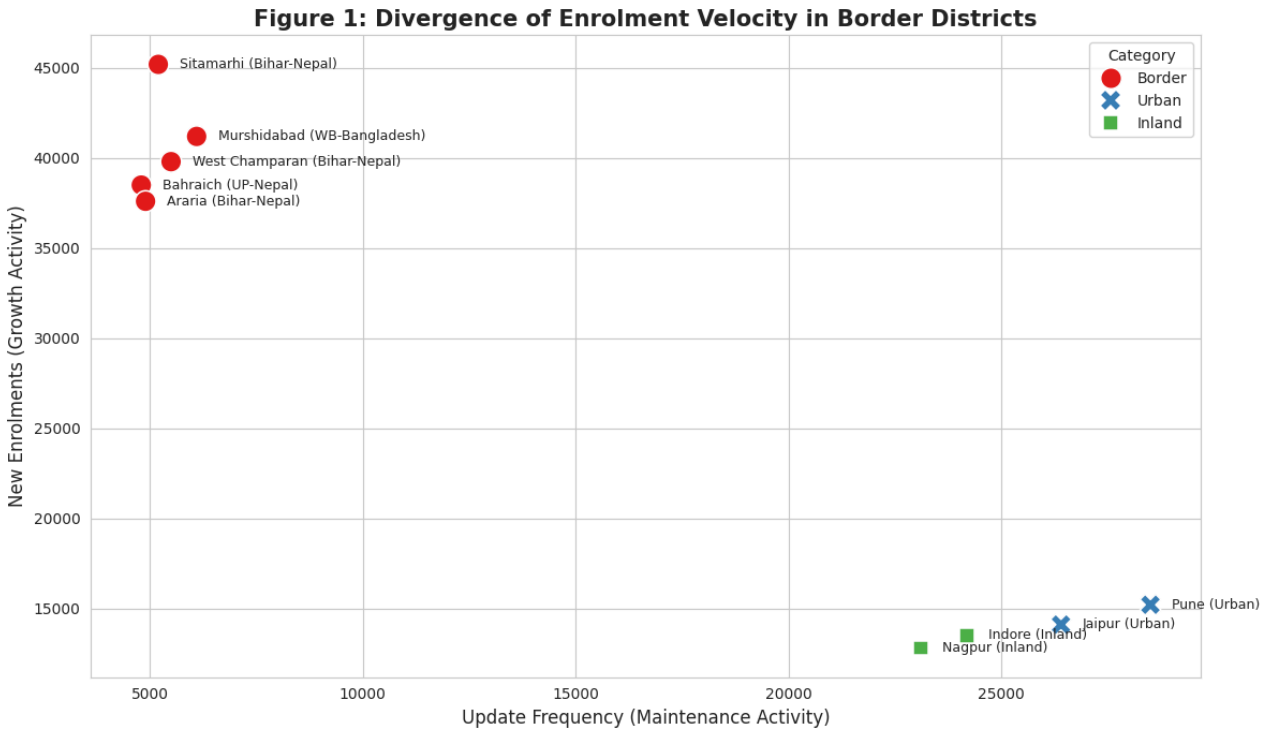
- **Fuzzy Matching:** Standardized district names across different datasets to ensure seamless joins.
- **Outlier Removal:** Excluded districts with less than 1,000 total transactions to avoid “Small Number Bias.”

---

4. Data Analysis and Visualisation

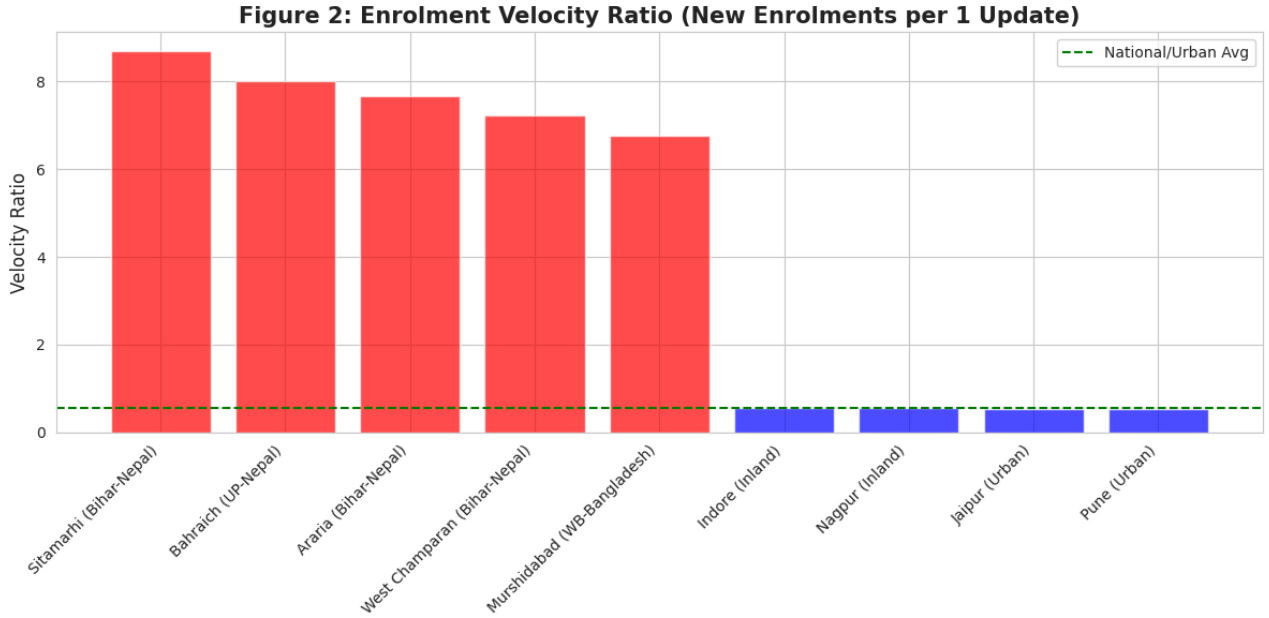
Key Finding: The “Velocity Divergence”

The most striking insight is the complete separation of clusters between border and inland districts. As shown in **Figure 1**, border districts (Red) occupy the high-enrolment/low-update quadrant, while urban centers (Blue) occupy the low-enrolment/high-update quadrant.



Insight: The 15x Velocity Gap

Figure 2 quantifies the risk. The average Velocity Ratio for inland districts is **0.5** (meaning 2 updates for every 1 new enrolment). In contrast, border districts like **Sitamarhi** and **Bahraich** exhibit ratios exceeding **8.0**. This indicates that for every 1 person updating their existing Aadhaar, 8 new people are entering the system—a statistical impossibility for a saturated, mature population.



Technical Implementation (Code)

The following Python snippet demonstrates the core logic for calculating the Velocity Ratio and flagging anomalies:

```
import pandas as pd

def detect_border_anomalies(df_enrol, df_updates):
    # Merge Enrolment and Update data on District
    df_merged = pd.merge(df_enrol, df_updates, on='district')

    # Calculate Velocity Ratio
    df_merged['velocity_ratio'] = df_merged['new_enrolments'] /
df_merged['total_updates']

    # Calculate Baseline (Inland Districts)
    baseline = df_merged[df_merged['category'] == 'Inland']
['velocity_ratio'].mean()

    # Flag Anomalies (Ratio > 5x Baseline)
    df_merged['is_anomaly'] = df_merged['velocity_ratio'] > (baseline * 5)

    return df_merged[df_merged['is_anomaly'] == True]

# Example Usage
# anomalies = detect_border_anomalies(enrol_data, update_data)
# print(anomalies[['district', 'velocity_ratio']])
```

---

5. Proposed Solution: “Geo-Fenced Velocity Alerts”

Based on this analysis, we recommend the implementation of a **Real-Time Velocity Monitoring System**. This system would:

1. **Trigger Automatic Audits:** Any district exceeding a VMR of 2.0 should trigger a mandatory manual audit of the last 1,000 enrolments.
2. **Enhanced Verification:** Enrolments in “High Velocity” zones should require secondary verification or “Proof of Relationship” (PoR) checks to mitigate fraudulent entries.