

UIDAI Aadhaar Data Analysis

Executive Summary Report

Generated by: Shuvam Banerji Seal, Alok Mishra, Aheli Poddar

Date: January 20, 2026

UIDAI Aadhaar Data Analysis Project

EXECUTIVE SUMMARY

DATASET OVERVIEW

- Total Records: 6.1+ Million across 3 datasets
- Biometric: 3.5M records (biometric authentication and updates)
- Demographic: 1.6M records (demographic corrections and updates)
- Enrollment: 982K records (new enrollments)
- Geographic Coverage: 36 States/UTs, ~960 Districts, 8,000+ Pincodes
- Attributes: 26 columns including temporal, geographic, and socioeconomic data

KEY FINDINGS

1. TEMPORAL PATTERNS
 - Clear weekday-weekend distinction in enrollment patterns
 - Tuesday shows peak enrollment activity
 - Weekends show 40-60% lower enrollment than weekdays
2. GEOGRAPHIC INSIGHTS
 - Significant regional imbalance with North and West dominating
 - Northeast region shows concerning underrepresentation
 - Urban-rural divide evident in enrollment rates
3. SOCIOECONOMIC CORRELATIONS
 - Strong positive correlation between HDI and enrollment ($r=0.68$)
 - Literacy rate significantly impacts adult enrollment ($r=0.72$)
 - Per capita income influences update frequency
4. MACHINE LEARNING RESULTS
 - Classification: Random Forest achieved 100% accuracy
 - Regression: R^2 values ~ 0.13 - 0.16 (indicating need for more features)
 - Clustering: Optimal $k=5$ clusters identified
 - Anomaly Detection: 5% contamination rate identified

RECOMMENDATIONS

- ✓ Increase weekend availability in underserved areas
- ✓ Prioritize enrollment drives in Northeast and Central regions
- ✓ Improve enrollment center density in low-literacy areas
- ✓ Use cluster profiles for customized outreach strategies