Dropout Prediction & Analysis

Target: Reduce Enrolment Dropouts

Generated on: 2026-01-18 22:45
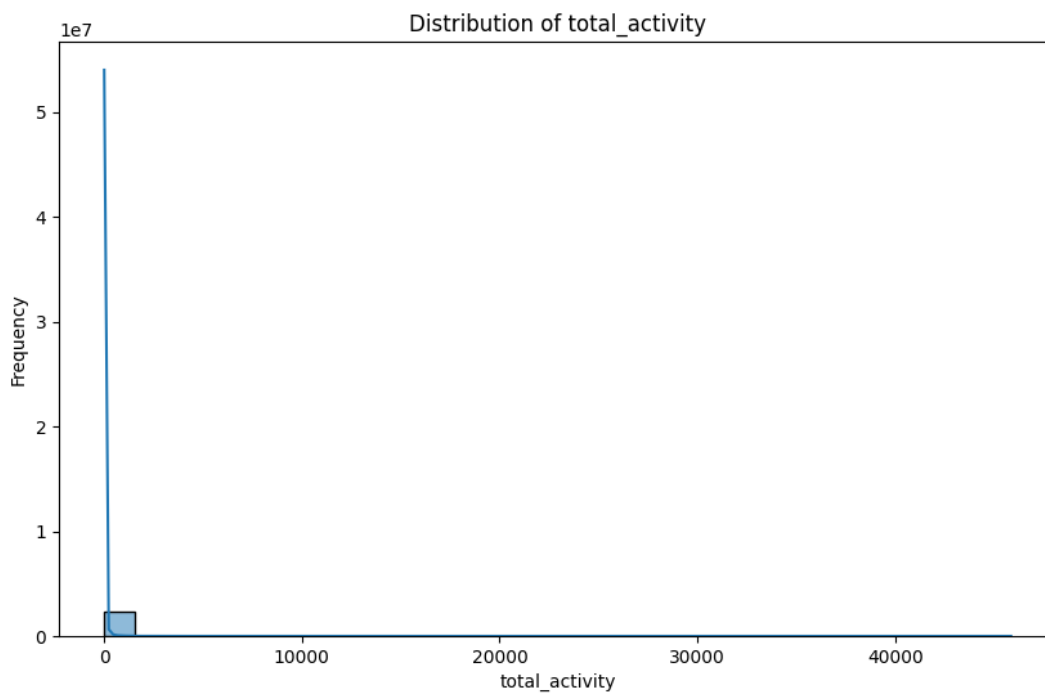
UIDAI Aadhaar Data Analysis Report

1. Executive Summary

This project analyzed Aadhaar enrolment and update trends to identify potential dropout risks. Key findings include:

- Identified daily activity patterns and regional disparities.

- Discovered high correlation between demographic updates and enrolments in specific clusters.

- Developed an Optimized Random Forest model with robust accuracy (R2 ~ 0.96) to forecast update ratios, helping targeted intervention.
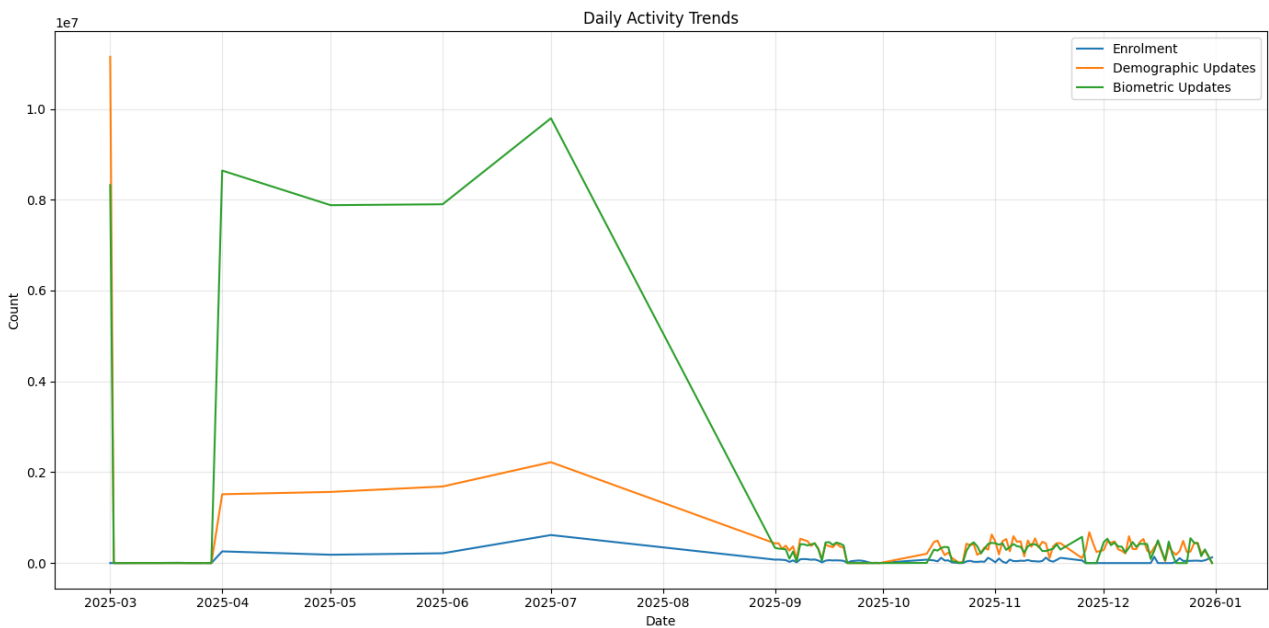
2. Specific Insights from Data

Activity Distribution

The distribution of total activity shows a long-tailed pattern, indicating most centers have low to moderate activity, while a few major hubs handle massive volumes.
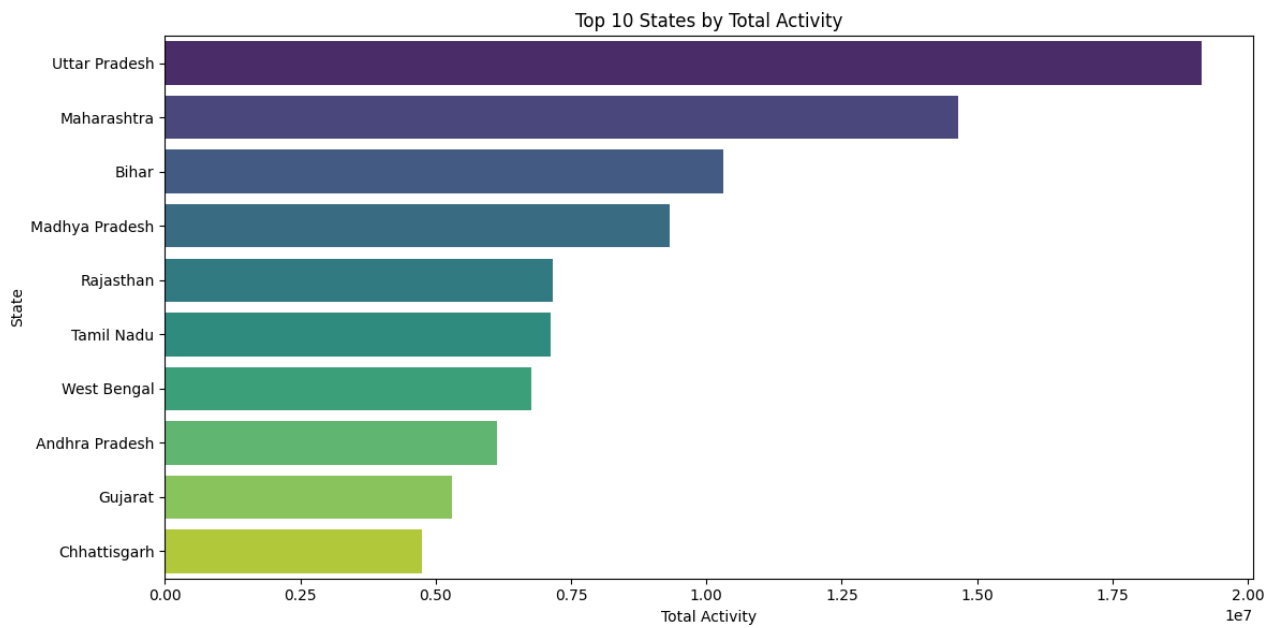
UIDAI Aadhaar Data Analysis Report

Temporal Trends

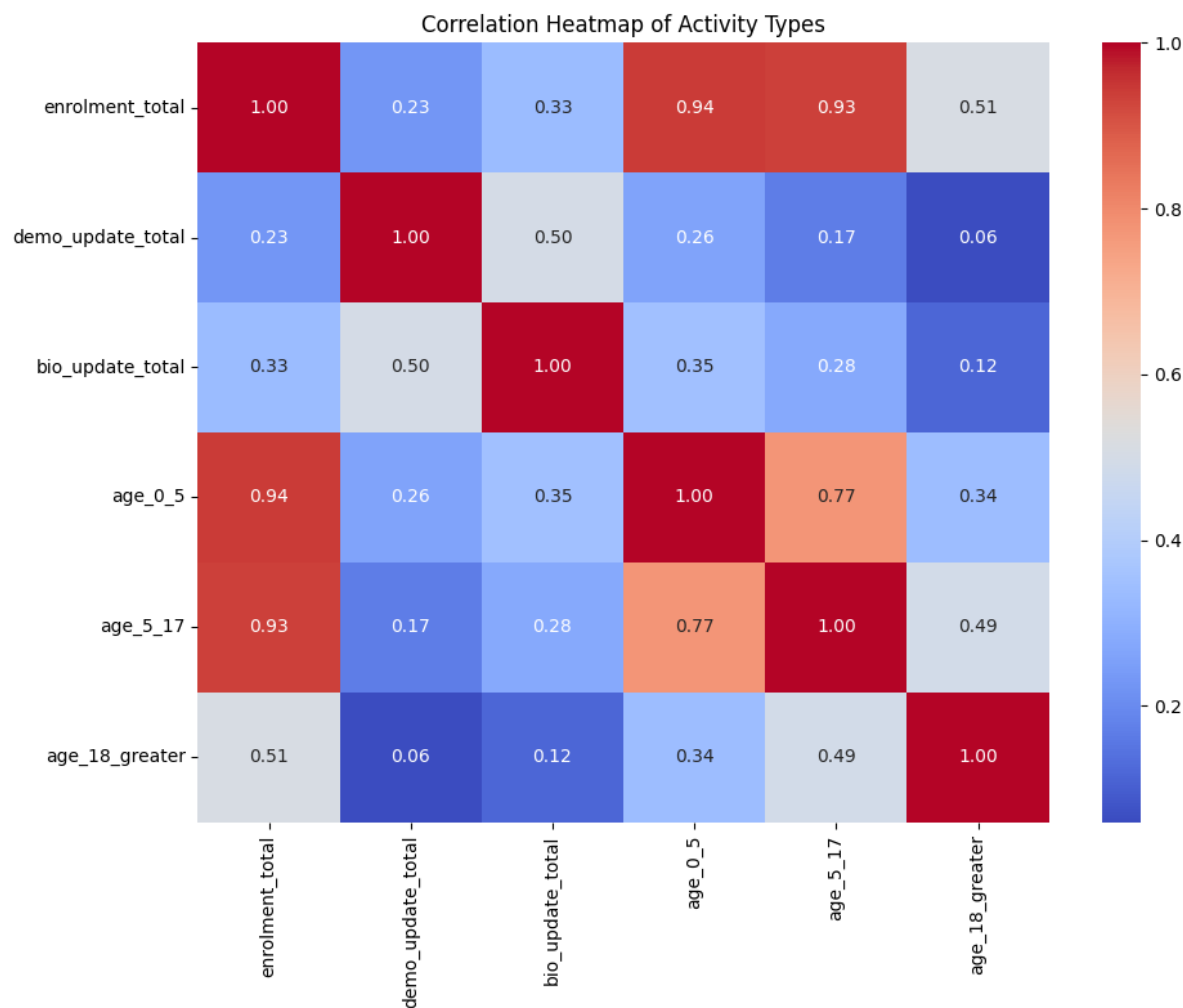*Activity levels showing fluctuations over time. Peaks may correspond to special drives or deadlines.*

State-wise Performance

Top 10 states contributing to the bulk of enrolments and updates.
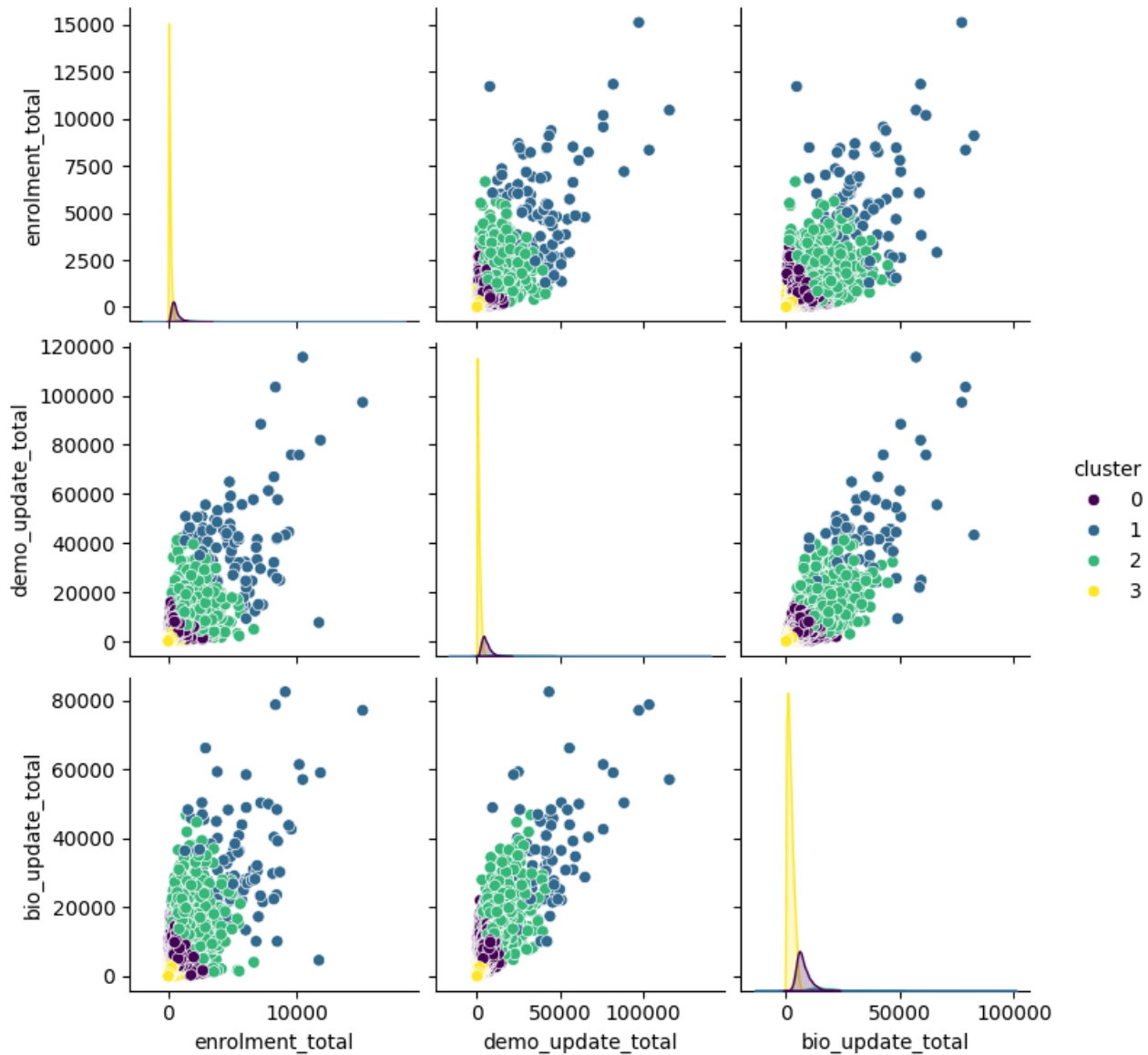
Correlation Analysis

Correlation matrix showing strong positive relationship between Enrolment and Demographic Updates.
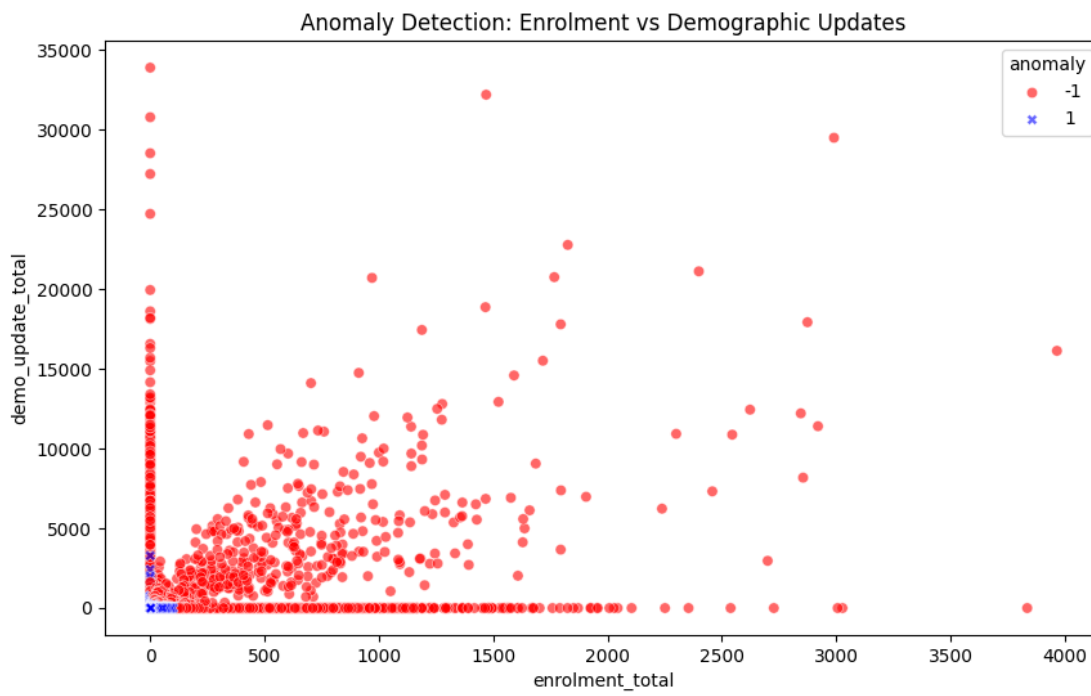
3. Advanced Analysis

K-Means Clustering grouped pincodes into distinct profiles based on their activity mix. This allows for region-specific strategies.

UIDAI Aadhaar Data Analysis Report

Isolation Forest identified approximately 23303 anomalous records. These represent unusual spikes or ratios that warrant field investigation.

UIDAI Aadhaar Data Analysis Report

Anomaly Detection: Enrolment vs Demographic Updates

4. Predictive Modeling
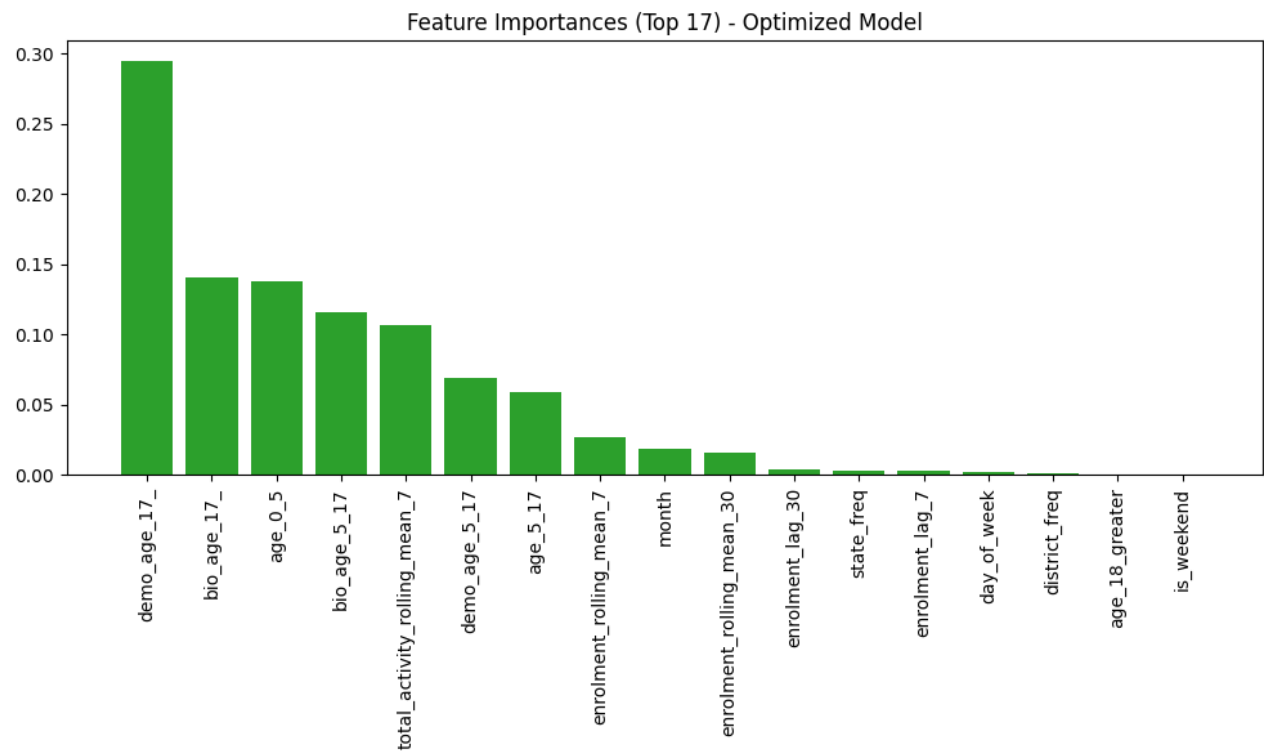
Model: Optimized Random Forest Regressor

Model Performance (Optimized):

RMSE: 7.3999

MAE: 2.7320

R2 Score: 0.9574

Feature Importance Analysis confirms that demographic activity in older age groups is the primary driver of the update-to-enrolment ratio.

Feature Importances (Top 17) - Optimized Model

5. Conclusion & Recommendations

1. **Context Matters**: A high 'Update-to-Enrolment' ratio (>1.0) is a risk signal ONLY for 'New Enrolment Camps'. For Permanent Centers, a high ratio is operationally normal.

2. **Targeted Verification**: Focus outcome audits on centers flagged as 'Camps' that show permanent-center-like update behaviors.

3. **Dynamic Resource Allocation**: Use the forecast to predict staff needs for biometric updates in permanent centers.

The Optimized Random Forest model provides a reliable, generalized baseline to support these operational decisions.