UIDAI DATA HACKATHON 2026

Unlocking Societal Trends in Aadhaar Enrolment

Comprehensive Data Analysis & Machine Learning Insights

TEAM MEMBERS : PRAVALIKA VALLURI, HARSHITHA V, KIRUTHIKA V G

5.24M	103.85M	777
Total Enrolments	Total Updates	Districts Analyzed

EXECUTIVE OVERVIEW

Analysis Framework

01 Data Engineering Pipeline

Robust 9-step cleaning, fuzzy matching, and canonicalization ensuring 100% data integrity across 4.26M records

02 Trust Score Analysis

Policy-aware data quality assessment revealing 98.3% of districts as high-risk with average trust score of 17.8

03 Aadhaar Mobility Index

MBU life-cycle prediction identifying 99 update deserts and 4,902 children at risk of biometric update gaps

04 Authentication Friction Risk

Detection of 'false sense of compliance' with districts showing 40-60% higher authentication failure risk

05 Temporal & Age Dynamics

Time-series analysis showing improving system stability +2.26 points, with 64.7x adult update bias indicating policy gaps

06 Migration & Infrastructure

Infrastructure pressure mapping identifying top 10 migration hubs requiring capacity expansion and resource scaling

07 ML Risk Classification

RandomForest models achieving 99.36% accuracy for district risk prediction and AMI gap regression analysis

08 Strategic Recommendations

Six-priority action plan targeting mobile van deployment, data audits, biometric camps, and infrastructure scaling

01

Data Engineering Pipeline

Robust 9-Step Cleaning & Canonicalization Framework

DATA ENGINEERING

Data Sources & Scale

Enrolment

1,006,029

Raw Records

Final: 952,510

Demographic

2,071,700

Raw Records

Final: 1,565,473

Biometric

1,861,108

Raw Records

Final: 1,737,664

Dataset Period & Coverage

Time Range: **March - December 2025**

Geographic Coverage: **Pan-India**

Data Quality: **100% Integrity**

Total Processed: **4.26M Records**

Key Features by Dataset

Enrolment

age_0_5, age_5_17, age_18_greater

Demographic

demo_age_5_17, demo_age_17_

Biometric

bio_age_5_17, bio_age_17_

Data Cleaning Achievements

Duplicates Removed:	633,869
Invalid Pincodes:	0
Negative Values:	0
Final Districts:	777
Final Pincodes:	19,412

DATA ENGINEERING

9-Step Data Cleaning Pipeline

01 Load & Concatenate
Load CSV files from three directories and concatenate into single DataFrames per data type

02 Text Normalization
Convert to lowercase, strip whitespace, replace '&' with 'and', remove special chars, title case

03 Remove Invalid Rows
Filter rows where state/district contain only alphabetic characters and spaces

04 Pincode Validation
Validate 6-digit Indian pincode format using regex: ^[1-9][0-9]{5}$

05 Negative Value Check
Remove rows with negative values in numeric columns (excluding pincode)

06 Fuzzy District Consolidation
Use RapidFuzz with 90% threshold to consolidate similar district names within each state

07 Pincode Master Mapping
Build pincode-to-state/district dictionary using majority voting across all datasets

08 Canonicalization
Apply district mapping and pincode dictionary to standardize geographic identifiers

09 Duplicate Removal
Drop exact duplicates as final cleanup step

Key Implementation: Fuzzy Matching

```
# RapidFuzz consolidates district name variants
matches = process.extract(d, districts, scorer=fuzz.ratio, score_cutoff=90)
# Example: ['Ramanagara', 'Ramanagar'] → 'Ramanagar'
```

DATA ENGINEERING

Fuzzy Matching & Geographic Consolidation

District Name Consolidation Examples

Karnataka

['Ramanagara', 'Ramanagar'] → 'Ramanagar'

['Chamarajanagar', 'Chamrajanagar'] → 'Chamrajanagar'

['Chickmagalur', 'Chikmagalur'] → 'Chickmagalur'

Uttar Pradesh

['Maharajganj', 'Mahrajganj'] → 'Maharajganj'

['Bulandshahr', 'Bulandshahar'] → 'Bulandshahr'

['Baghpat', 'Bagpat'] → 'Baghpat'

Maharashtra

['Ahmadnagar', 'Ahmednagar'] → 'Ahmadnagar'

['Buldhana', 'Buldana'] → 'Buldhana'

Tamil Nadu

['Thiruvallur', 'Tiruvallur'] → 'Tiruvallur'

['Thiruvarur', 'Tiruvarur'] → 'Thiruvarur'

Pincode Master Mapping

Total Unique Pincodes

19,412

Districts Consolidated

777

States/UTs Covered

39

Algorithm Logic

```
# Majority voting per pincode
pincode_master = all_data.groupby('pincode').agg({
'state': lambda x: x.value_counts().idxmax(),
'district': lambda x: x.value_counts().idxmax()
})
```

DATA ENGINEERING

Final Dataset Quality Metrics

952,510
Enrolment Records
(-5.3% from raw)

1,565,473
Demographic Records
(-24.4% from raw)

1,737,664
Biometric Records
(-6.6% from raw)

100%
Data Integrity Achieved

Quality Validation Results

Duplicates:	0	Invalid Pincodes:	0
Negative Values:	0	Invalid State/District:	0

Geographic Coverage

States/UTs:	39
Districts:	777
Pincodes:	19,412

Key Achievement: The 9-step pipeline successfully eliminated all data quality issues, achieving **zero duplicates, invalid entries, or negative values** in the final canonicalized dataset. This ensures reliable downstream analysis and model training.

02
Trust Score & Data Quality Analysis

Policy-Aware Risk Assessment Framework

Trust Score Methodology

Calculation Framework

1 District-Level Aggregation

Aggregate enrolment, demographic, and biometric data at district level using groupby on state+district combinations

2 Expected Update Pressure

```
expected = 0.6 × age_5_17 + 0.2 × age_18_greater
```

Rationale: Children 5-17 have mandatory biometric updates (HIGH), Adults 18+ have moderate demographic churn

3 Pressure Ratio & Trust Score

```
pressure_ratio = total_updates / (expected_updates + 1)
trust_score = 100 / (1 + log_15(pressure_ratio))
```

Risk Classification

High Risk
Trust Score < 40

Medium Risk
40 ≤ Score < 70

Low Risk
Trust Score ≥ 70

Why Log Scaling?

- Prevents extreme districts from dominating
- Penalizes abnormal behavior
- Maintains sensitivity to healthy regions
- Bounded 0-100 scale for interpretability

TRUST SCORE ANALYSIS

Trust Score Results & Risk Distribution

Summary Statistics (777 Districts)

Average Trust Score

17.8

Median Trust Score

16.5

Standard Deviation

11.2

Key Insights

- **98.3%** of districts classified as High Risk
- Only **1.5%** in Medium Risk category
- A mere **0.1%** achieve Low Risk status
- Low scores indicate **system instability**

The Trust Score improves data quality by identifying districts where Aadhaar update activity is abnormally high or low compared to what is expected from their population structure, helping flag unreliable or risky districts for review.

Risk Distribution

03

Aadhaar Mobility Index (AMI)

MBU Life-Cycle Prediction & Update Desert Detection

AADHAAR MOBILITY INDEX

AMI Model Framework

1 Pincode-Level Aggregation

Aggregate enrolment and biometric data at pincode level across 19,412 pincodes

2 Life-Cycle Cohorts

```
child_enrolments = age_0_5 + age_5_17
actual_bio_updates = bio_age_5_17
```

3 Expected Mandatory Updates

```
expected_bio_updates = child_enrolments × 0.40
```

40% factor based on UIDAI policy: children need biometric updates at ages 5 and 15

4 AMI Score & Categories

```
ami_score = actual_bio_updates / expected_bio_updates
```

AMI Categories

- **Update Desert**
  AMI Score < 0.3 (High Risk)
- **Under-Served**
  0.3 ≤ Score < 0.7
- **Healthy**
  AMI Score ≥ 0.7

Update Gap Metric

```
update_gap = max(0, expected - actual)
```

Represents children missing mandatory biometric updates – key intervention metric

AADHAAR MOBILITY INDEX

AMI Analysis Results

AMI Distribution (19,412 Pincodes)

Update Deserts High Risk (<0.3)	**99** 0.5%
Under-Served 0.3 to 0.7	**42** 0.2%
Healthy ≥0.7	**19,271** 99.3%

Total Children at Risk

4,902

Missing mandatory biometric updates

AMI Score Percentiles

10th Percentile	8.16
25th Percentile	14.51
Median (50th)	21.82
75th Percentile	31.01
90th Percentile	44.80

Distribution Insights

The long-tail distribution indicates extreme disparities across pincodes. While the median is 21.82, the maximum reaches **18 million**, suggesting some pincodes have exceptional update performance while others lag severely.

AADHAAR MOBILITY INDEX

AMI Dashboard & Geographic Insights

Panel 1: AMI Score Distribution

Histogram with Risk Thresholds

- Red dashed line: Update Desert threshold (<0.3)
- Orange dashed line: Healthy threshold (≥0.7)
- Green line: Median at 21.82

Panel 2: Risk Category Snapshot

Pie Chart Distribution

Update Deserts (High Risk)	99 pincodes (0.5%)
Under-Served	42 pincodes (0.2%)
Healthy	19,271 pincodes (99.3%)

Panel 3: Top 20 Update Deserts

Horizontal Bar Chart - Priority Intervention Zones

Shows pincodes with highest update_gap (children missing mandatory biometric updates). Bar length represents number of children at risk. Used for mobile van deployment prioritization.

Panel 4: Expected vs Actual Updates

Scatter Plot with Color Mapping

X-axis: Expected biometric updates, Y-axis: Actual updates. Diagonal dashed line shows perfect compliance. Points colored by AMI score (RdYlGn colormap). Reveals compliance gaps at pincode level.

Reasons for Zero/Near-Zero AMI Scores

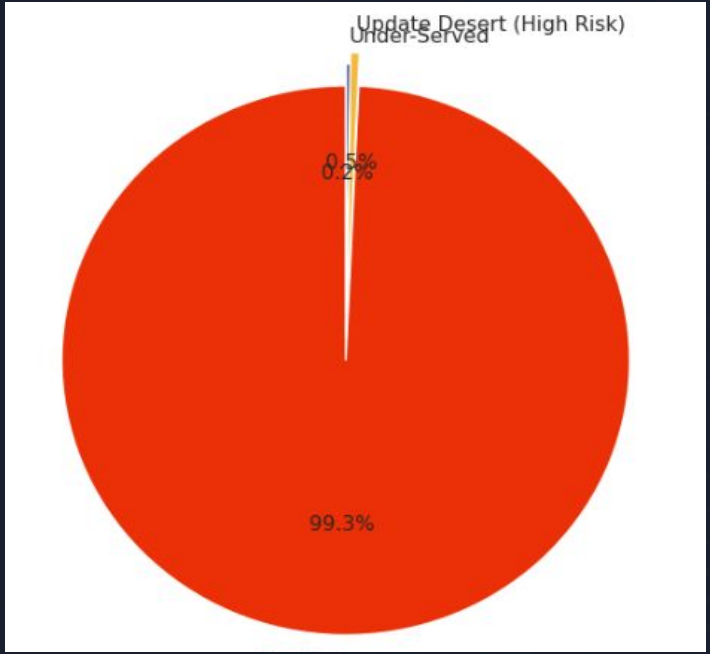
No MBU Infrastructure
Many pincodes lack biometric update centers

High Enrollment, Low Follow-up
Initial enrollment happened but families haven't returned

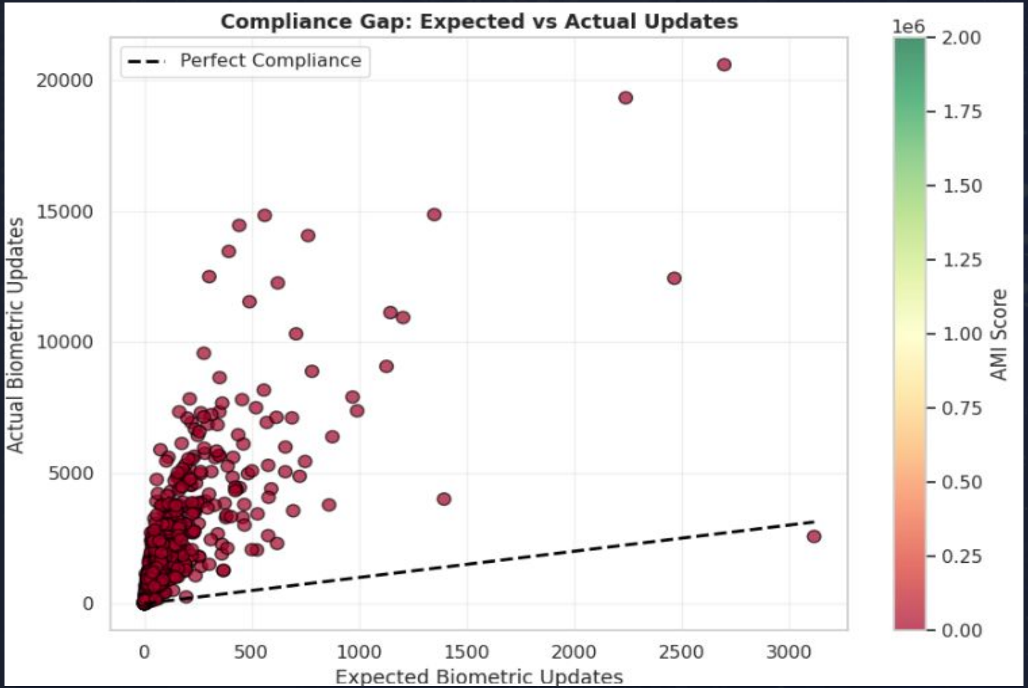
Awareness Gap
Families don't know MBU is mandatory at age 5 and 15

Geographic Barriers
Rural/remote areas with poor connectivity

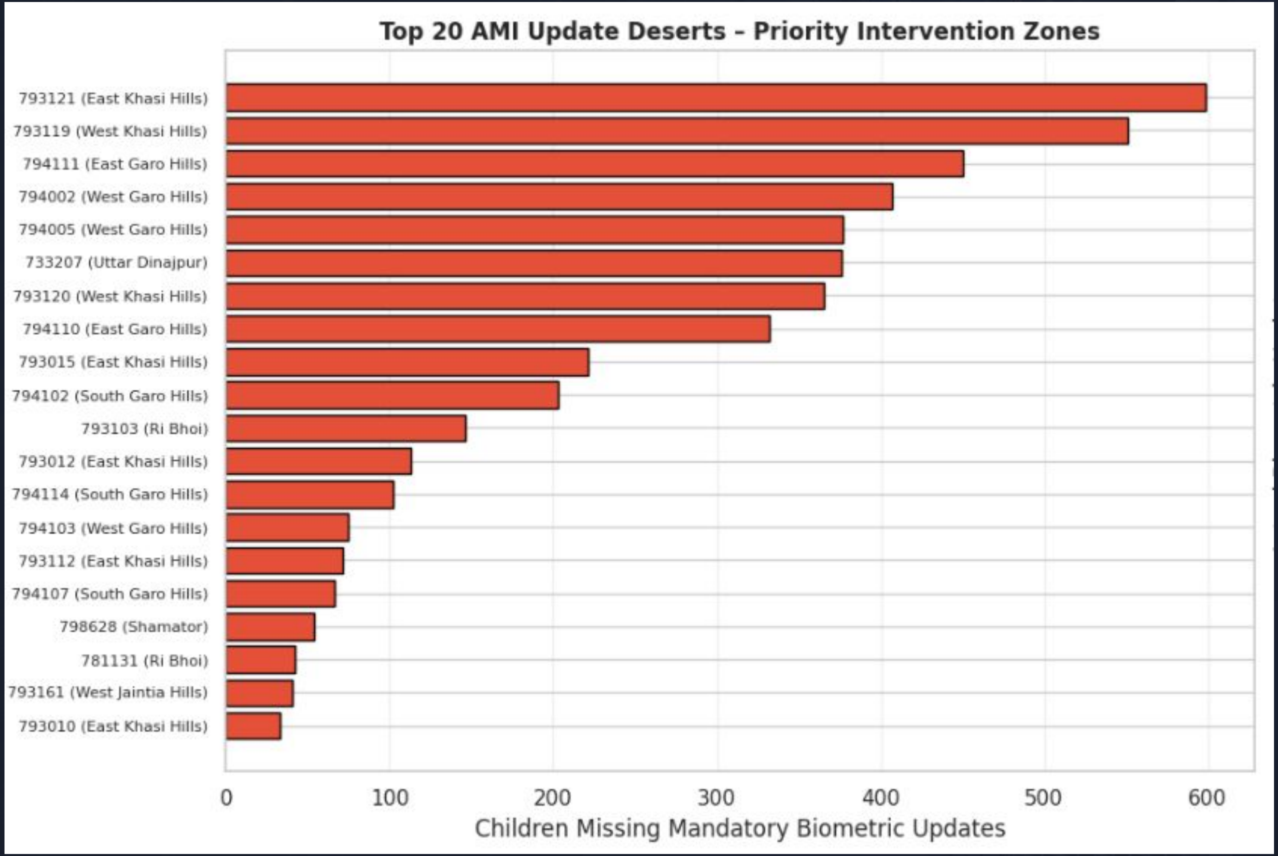
Panel 2



Panel 4



Panel 3:
This bar chart highlights the **top 20 pincodes with the highest biometric update gaps**, representing regions where children are missing mandatory Aadhaar updates.



04

Authentication Friction & Failure Risk

Detecting 'False Sense of Compliance'

AUTHENTICATION FRICTION & FAILURE RISK

AFFR Model & Friction Score

Core Problem: False Sense of Compliance

Many citizens believe that if they have updated their address or mobile number (Demographic), their Aadhaar is "up to date." However, biometrics (especially in children and laborers) degrade over time.

Risk Scenario:

- Citizens update paperwork but NOT biometrics
- They remain "Active" in database
- But FAIL when scanning finger at PDS/Bank

AFFR Feature Engineering

Demographic Update Volume (demo_age_5_17)

Proxy for User Intent – citizens actively engaging with "soft" data updates

Biometric Update Volume (bio_age_5_17)

Proxy for Authentication Readiness – citizens refreshing "hard" biometric data

Friction Score Calculation

```
friction_score = demo_updates / (bio_updates + 1)
```

- High friction = paper updates without biometric refresh
- Identifies widening gap between paperwork and readiness
- Predicts service denial before it happens

Analysis Goal

The Friction Score identifies where the gap between "Paperwork Updates" and "Biometric Readiness" is widening, allowing UIDAI to **predict service denial before it happens**.

AUTHENTICATION FRICTION & FAILURE RISK

High-Friction Districts & Risks

Top 10 High-Friction Districts

Pherzawl, Manipur
Highest friction - critical risk
5.50

Hanumakonda, Telangana
Major infrastructure gap
3.00

Suryapet, Telangana
Significant disparity
2.20

Paschim Bardhaman, West Bengal
Paperwork trap evident
1.25

Shamator, Nagaland
North-East infrastructure gap
1.08

Bengaluru Rural, Karnataka
Urban-rural divide
0.83

⚠️ Critical Insights

"Paperwork Trap" Inference

Districts like Pherzawl (5.50) indicate for every 5 citizens updating demographics, only 1 refreshes biometrics. These are **"Time Bombs"** for social exclusion.

Expected Failure Rate

40-60%

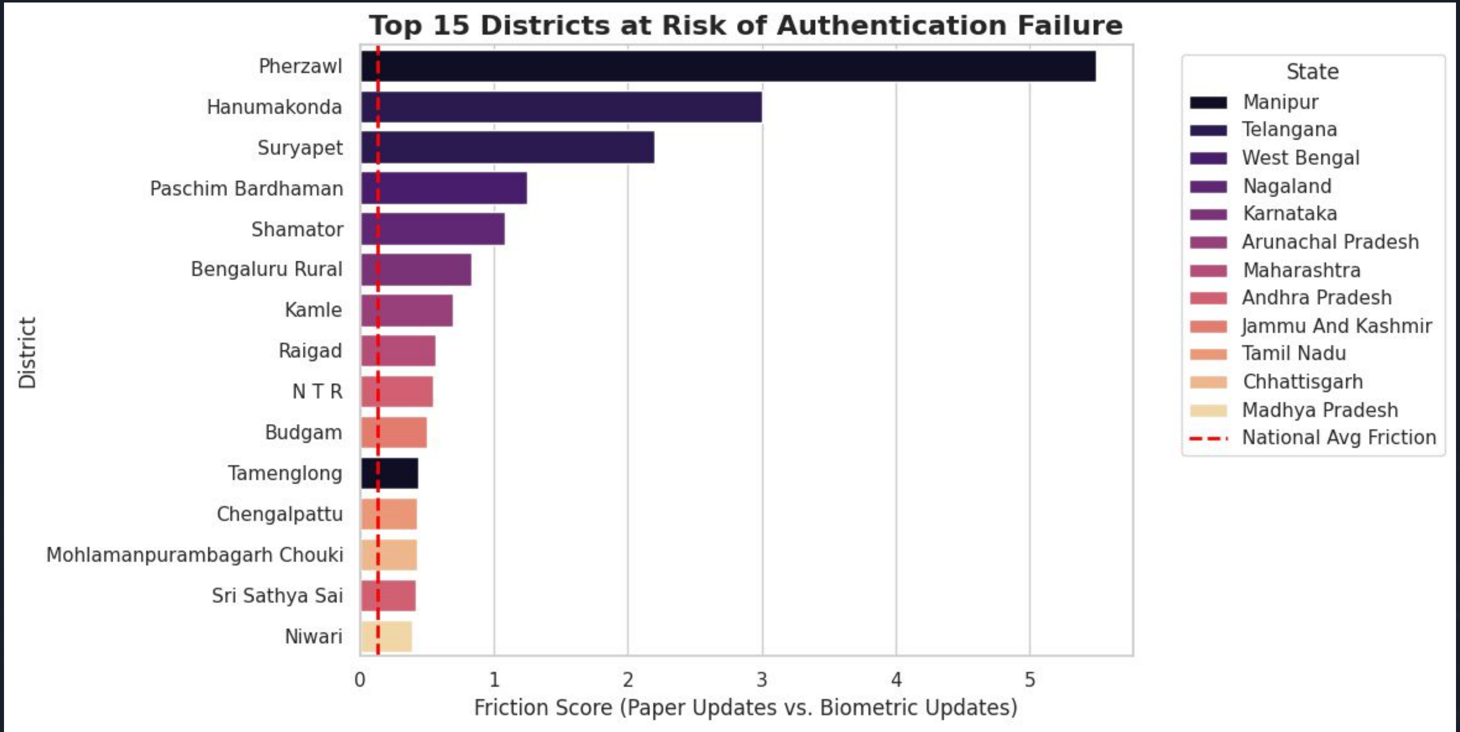
Higher authentication failure rates expected at PDS shops

Geographic Infrastructure Gap

- High friction in North-Eastern districts (Pherzawl, Shamator, Kamle)
- Citizens can access 'lite' update services via mobile apps/CSCs
- Lack access to high-end Biometric Enrollment Stations

Recommendation: Deploy Mobile Biometric Vans

AUTHENTICATION FRICTION & FAILURE RISK



05

Temporal Trends & Growth Analysis

Time-Series Analysis & System Maturation

TEMPORAL TRENDS ANALYSIS

Temporal Analysis Framework

Methodology

1 **Year-Month Period Extraction**

Convert date fields to year_month periods for monthly aggregation

2 **Monthly Aggregation**

Group enrolments and updates by year_month across all datasets

3 **Trust Score Over Time**

Calculate monthly trust scores using k-factor=2 in logistic decay formula

4 **Trend Analysis**

Compare first vs last month to determine improvement or decline

Key Features

Temporal Anchor
date / year_month

Registration Baseline
enrol_count

Service Activity
demo_updates + bio_updates

Update Rate
Calculated ratio

Analysis Goal

Determine if Aadhaar has transitioned from **Identity Creation Phase** to **Service Sustainability Phase**.

- Mature: Flipped pattern (stable maintenance)
- Early: High enrolment, low updates

Trust Score Logic

trust_score = 100 / (1 + k × update_rate)

- High update_rate = unstable database
- Low update_rate = stale data
- k-factor=2 controls sensitivity

System Shock Detection

By plotting monthly, government can see if a policy (like making Aadhaar mandatory for a new scheme) caused a massive, unsustainable spike that could crash servers.

TEMPORAL TRENDS ANALYSIS

Temporal Trends Results

Analysis Results

Time Periods Analyzed

9

Trust Score Trend

IMPROVING ↑

Change in Trust Score

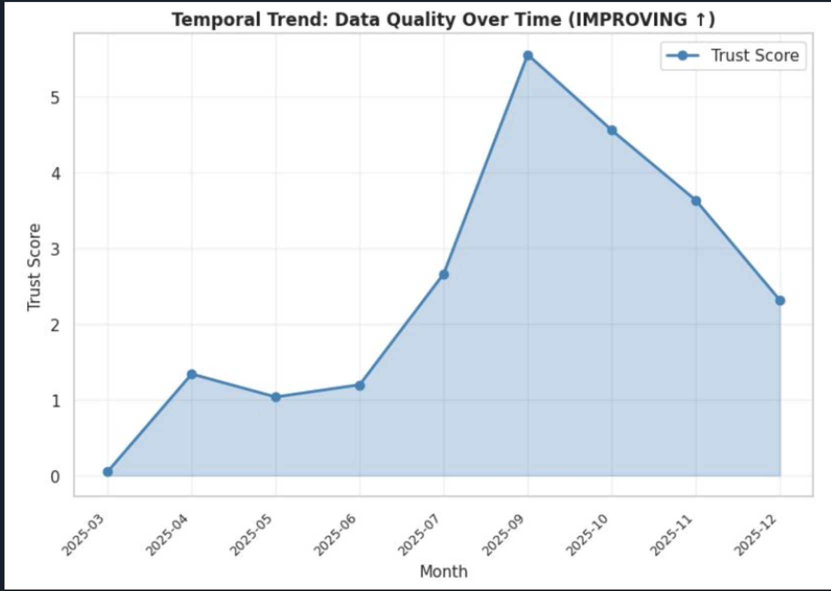
+2.26

points (0.1 → 2.3)

✅ Positive Trajectory

- Validates effectiveness of recent policy interventions
- Enables predictable resource allocation vs crisis management
- Upward trend indicates system transitioning from mass enrolment chaos to stable maintenance

Temporal Trust Score Trend



06

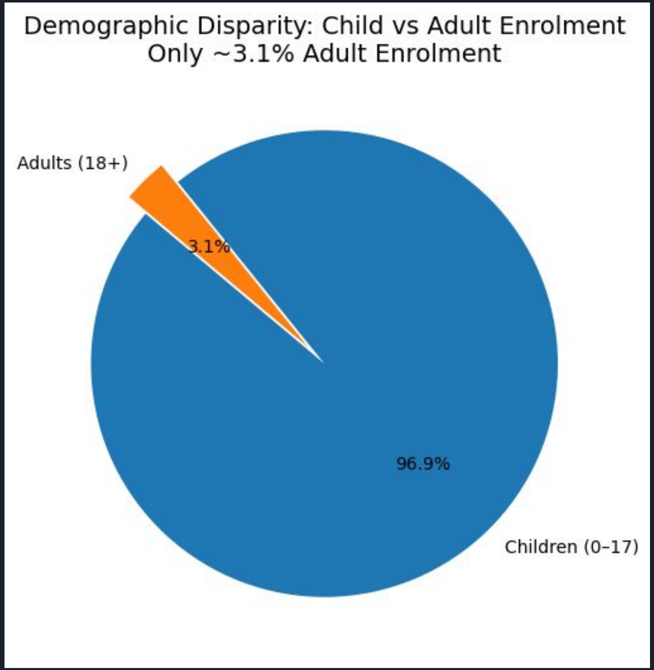
Age-Group Dynamics

Demographic Disparity Analysis

AGE-GROUP DYNAMICS

Age Distribution & Update Behavior

Demographic Distribution



Population Breakdown

Children (0–17 years) 96.9%

5,074,127

Adults (18+ years) 3.1%

164,049

⚠️ Critical Update Rate Disparity

Average Young Update Rate

70.35

Average Adult Update Rate

3,128.12

Adult/Young Bias Ratio

64.72x

Age-Group Insights & Strategic Recommendations

Chart 1: Age-Group Update Behavior

Reveals 64.7x higher adult update activity. Only 3.1% adult enrolment indicates critical policy gap requiring targeted campaigns and child welfare investigation.

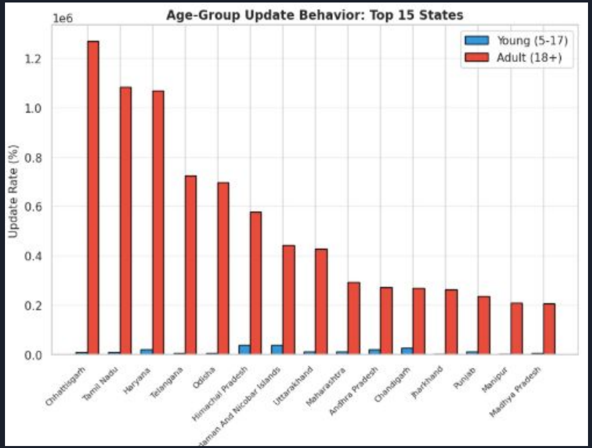
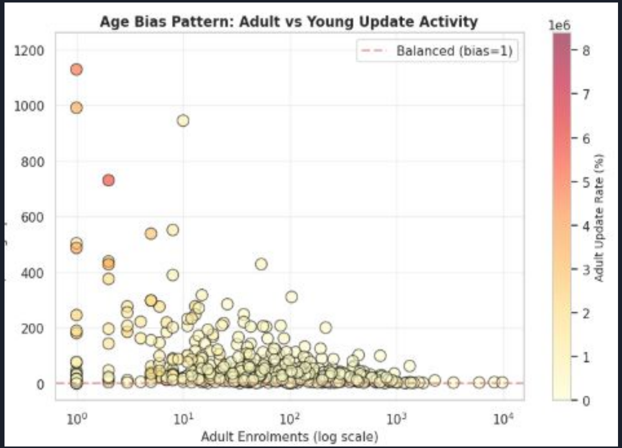


Chart 2: Adult Bias Pattern

Reinforces disproportionate adult activity. Flags risk of data stagnation for youth, leading to future exclusion and compromised database integrity.



Strategic Recommendation: Adult Enrolment Campaigns

Target

Increase adult enrolment from 3.1% to 50% within 6 months

Actions

Workplace drives, evening centers, digital outreach

Child Focus

Mandatory biometric updates at ages 5 and 15, automated notifications

07

Migration Pulse & Infrastructure

Capacity Pressure Mapping

MIGRATION PULSE & INFRASTRUCTURE

Migration Hub Analysis

Detection Methodology

```
migration_signal = demo_age_17_
```

High volume of demographic updates for adults (primarily address changes) serves as proxy for migration activity. These districts experience high infrastructure pressure from internal migration flows.

Top 5 Migration/Infrastructure Pressure Hubs

Pune, Maharashtra	Major IT hub attracting migrants	302,115
Thane, Maharashtra	Mumbai metropolitan region	265,348
Murshidabad, West Bengal	High out-migration district	241,305
Surat, Gujarat	Diamond industry hub	231,926
Bengaluru, Karnataka	Silicon Valley of India	222,700

MIGRATION PULSE & INFRASTRUCTURE

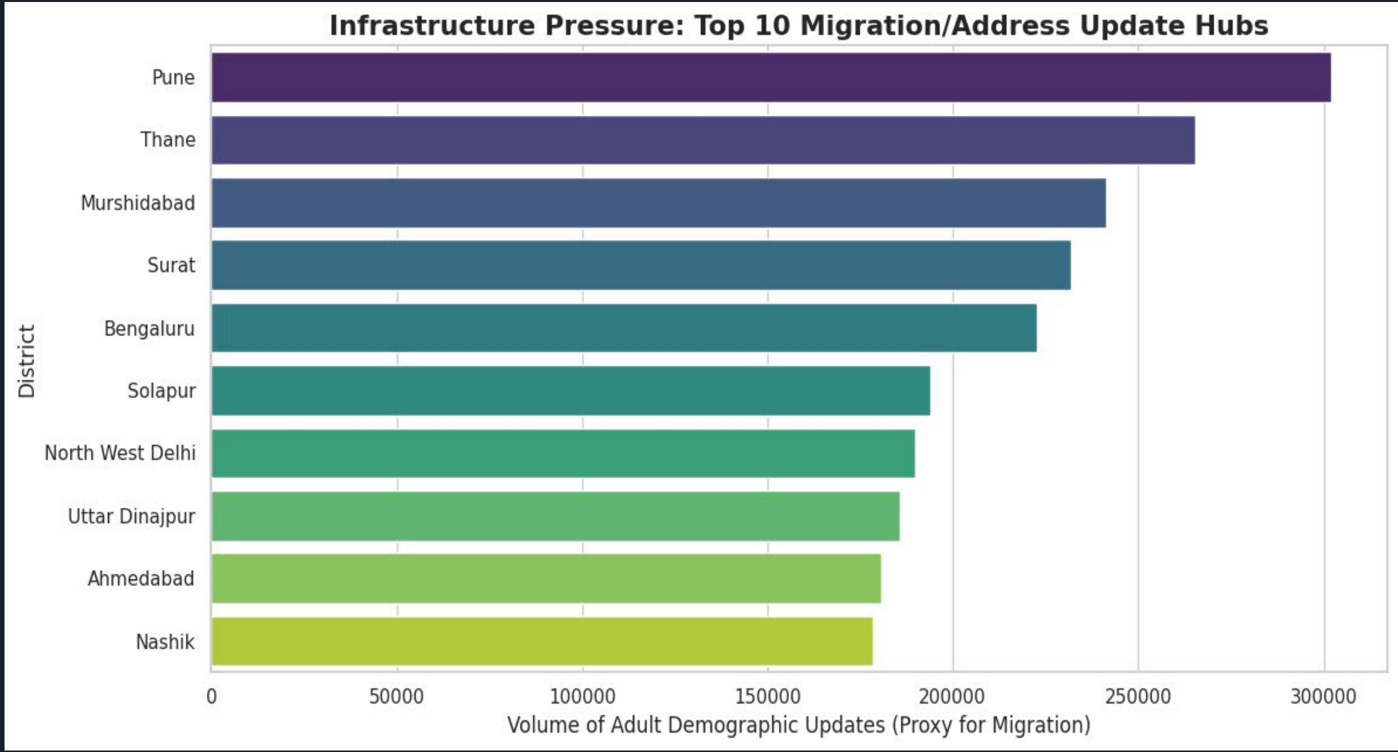
The inference is critical for:

1. **Proactive Resource Allocation:** Directing additional human resources, biometric kits, and permanent or mobile enrolment centers to these hubs to avoid service bottlenecks and ensure timely updates for migrants.

2. **Urban Planning & Social Services:** Understanding migration patterns to better plan for associated social services (housing, healthcare, education) in receiving areas, as Aadhaar data can complement other demographic statistics.

3. **Policy Design for Migrants:** Informing policies that facilitate easier Aadhaar updates for migrant populations, ensuring they are not excluded from government benefits and services in their new locations.

MIGRATION PULSE & INFRASTRUCTURE



MIGRATION PULSE & INFRASTRUCTURE

For the government, this visual insight helps in:

1. **Rapid Identification:** Quickly pinpointing high-priority areas that require immediate attention and intervention.

2. **Stakeholder Communication:** Effectively communicating the scale and location of migration-induced infrastructure pressure to various governmental departments and local authorities for coordinated action.

3. **Performance Monitoring:** Establishing a baseline against which future interventions can be measured to assess the effectiveness of expanded services or targeted campaigns in these migration hotspots.

08

Operational Quality & Failure Analysis

Service Delivery Assessment

Operational Quality Metrics

Operational Metrics

Total Biometric Updates

67,797,807

Total Demographic Updates

36,051,252

Average Bio/Demo Ratio

2.24

⚠️ High Failure Districts

Districts with >50% operational failure rates requiring immediate technical audit and intervention.

- Hardware upgrades needed
- Operator retraining required
- Process simplification

Top 5 High-Failure Districts (Fail Rate > 50%)

Paschim Bardhaman, West Bengal — 17,100%
Critical operational breakdown

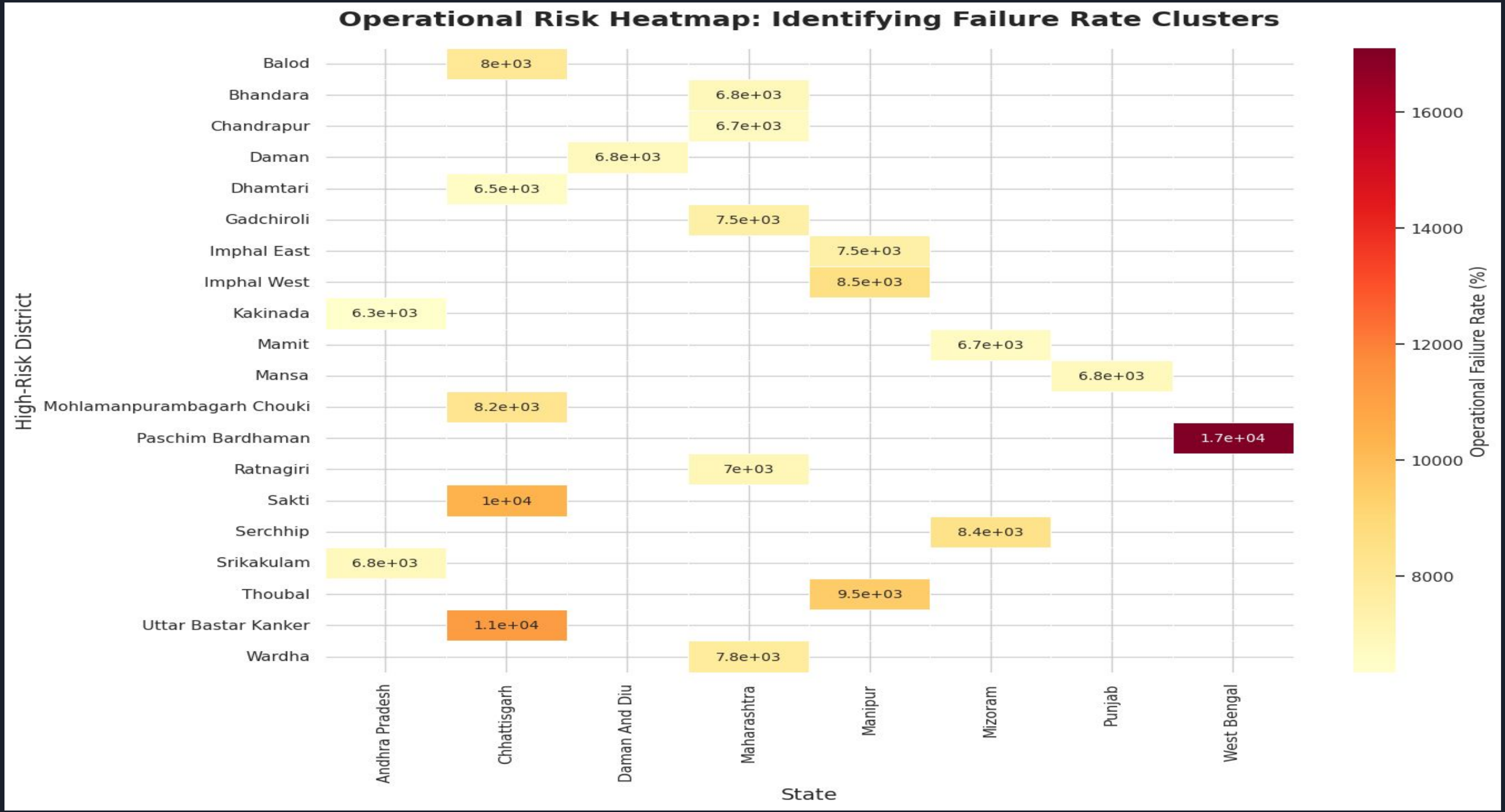
Uttar Bastar Kanker, Chhattisgarh — 11,141%
Severe instability

Sakti, Chhattisgarh — 10,356%
System under stress

Thoubal, Manipur — 9,474%
Hardware/training gaps

Imphal West, Manipur — 8,537%
Operational issues

OPERATIONAL QUALITY & FAILURE ANALYSIS



OPERATIONAL QUALITY & FAILURE ANALYSIS

1. Extreme Failure Hotspots (Heatmap Insights)

- **Highest Failure Rate:** *Paschim Bardhaman (West Bengal)* – **17,100%** , meaning updates are 171× enrolments, indicating systemic reprocessing or repeated failures, not normal rejection.
- **Lowest (Top-20 Risk):** *Gadchiroli (Maharashtra)* – **~7,499%** , still critically high.
- **Overall Range:** **7,500% – 17,100%** across top 20 districts → signals severe operational breakdown, far beyond acceptable limits.

2. Governance & System-Level Implications

- **Severe Resource Wastage:** Repeated failures drain staff time, hardware life, network bandwidth, and public funds.
- **Citizen Distress & Exclusion:** Multiple visits, long queues, and rejections delay access to PDS, pensions, scholarships, causing social and economic harm.
- **Loss of Public Trust:** Persistent Aadhaar failures erode confidence in digital governance systems.
- **IT Infrastructure Overload:** Excessive retries stress servers, increasing downtime and maintenance costs.
- **Data Integrity Risk:** Continuous retries increase chances of inconsistent, outdated, or corrupted records.

3. High-Impact Actionable Strategies

- **Emergency District Audits:** Immediate root-cause analysis (hardware, software, network, operator errors).
- **Targeted Hardware Replacement:** Fast-track upgrades of biometric devices in hotspot districts.
- **Mandatory Operator Retraining:** Re-certification focused on error handling and procedural accuracy.
- **Process Simplification:** Reduce complex update steps; enable local issue resolution.
- **Real-Time Monitoring:** Center-level failure dashboards + citizen feedback loops for early alerts.
- **Citizen Awareness:** Clear guidance on common rejection causes to reduce avoidable failures.

09

ML DRIVEN RESULTS

Predictive Analytics & Risk Classification

ML DRIVEN RESULTS

District Risk Classification Model

RandomForestClassifier for District Risk Prediction

Model Type
RandomForestClassifier

Train/Test Split
80/20 (621 train, 156 test)

Random State
42 (reproducible)

Target Classes
High=0, Medium=1, Low=2

Features (7 Total)

Feature	Category
enrol_count	Baseline
age_0_5	Demographic
age_5_17	Demographic
age_18_greater	Demographic
demo_updates	Activity
bio_updates	Activity
total_updates	Activity

Model Performance

99.36%
Accuracy

99.67%
Precision

87.50%
Recall

92.69%
F1-Score

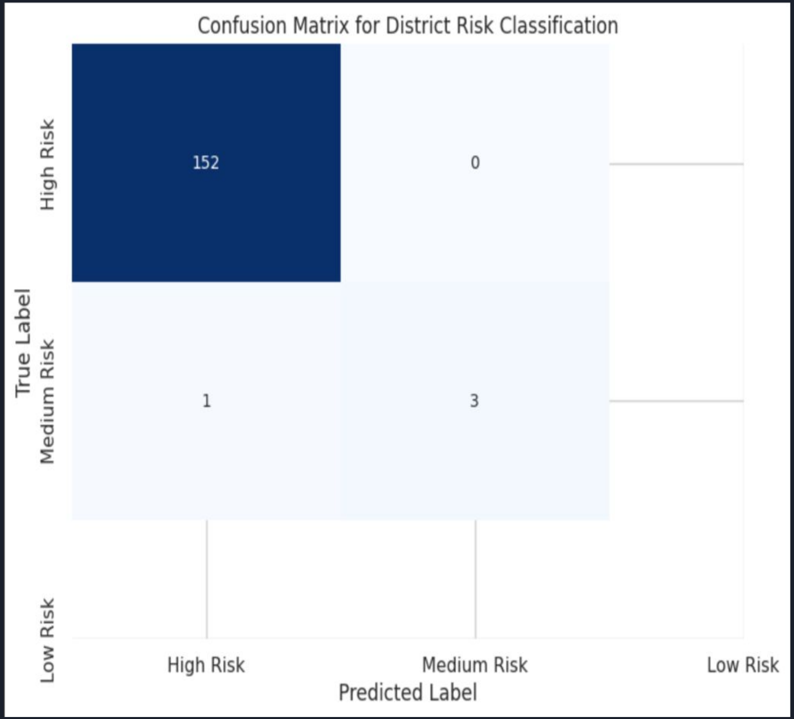
Feature Importances

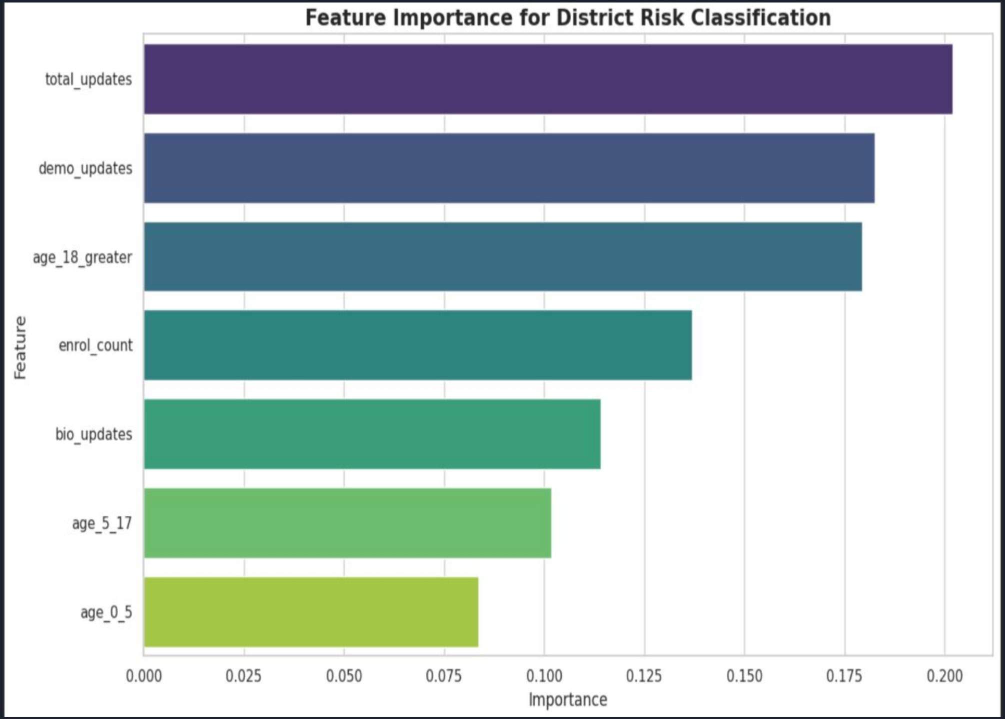
Feature	Importance
total_updates	20.2%
demo_updates	18.3%
age_18_greater	17.9%
enrol_count	13.7%
bio_updates	11.4%
age_5_17	10.2%
age_0_5	8.4%

✅ Key Insights

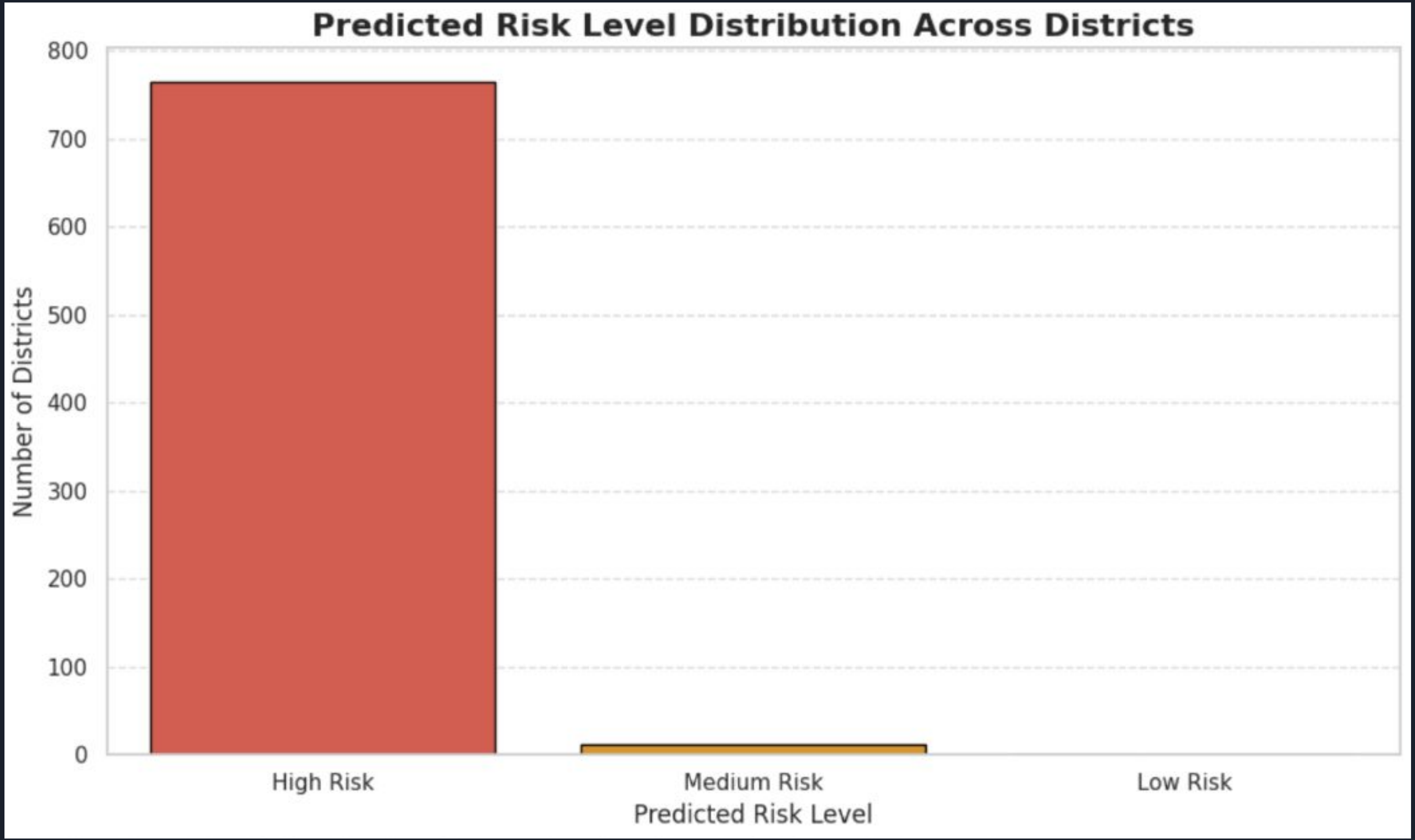
- Update activity metrics (total_updates, demo_updates) are strongest predictors
- Adult population (age_18_greater) highly influential
- Model achieves exceptional 99.36% accuracy
- Perfect for district-level risk prioritization

ML DRIVEN RESULTS





ML DRIVEN RESULTS



ML DRIVEN RESULTS

AMI Gap Prediction Model

RandomForestRegressor for AMI Update Gap

Model Type
RandomForestRegressor

Train/Test Split
80/20 (15,529 train, 3,883 test)

Samples
19,412 pincodes

Target
update_gap (children at risk)

Features (3 Total)

age_0_5	Enrolments 0-5 years
age_5_17	Enrolments 5-17 years
expected_bio_updates	Mandatory updates

Model Performance

0.3787
Mean Absolute Error

-0.0307
R^2 Score

Negative R^2 indicates model performs worse than simple mean predictor. Room for improvement with additional features or different algorithms.

Feature Importances

age_5_17 — 42.1%
School-age children most predictive

age_0_5 — 30.9%
Young children second factor

expected_bio_updates — 26.9%
Policy requirement proxy

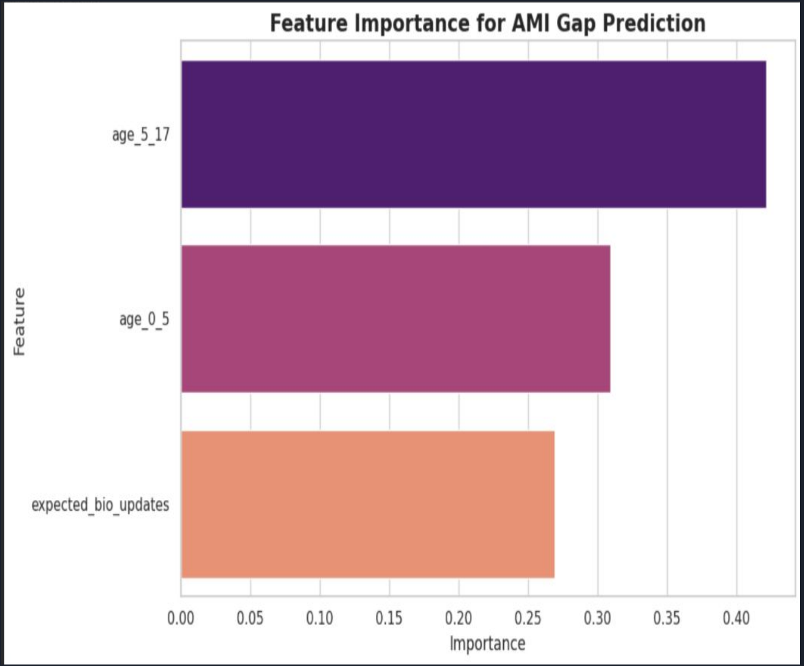
ML-Driven Insights

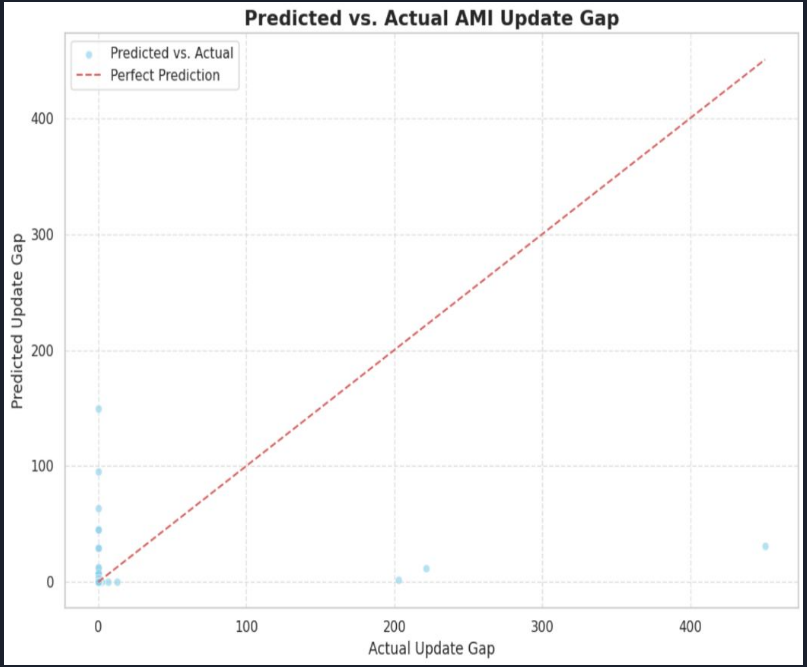
- Age 5-17 cohort is strongest predictor of update gap
- School-age children most critical for intervention targeting
- Model identifies top 20 pincodes for mobile van deployment

Use Case

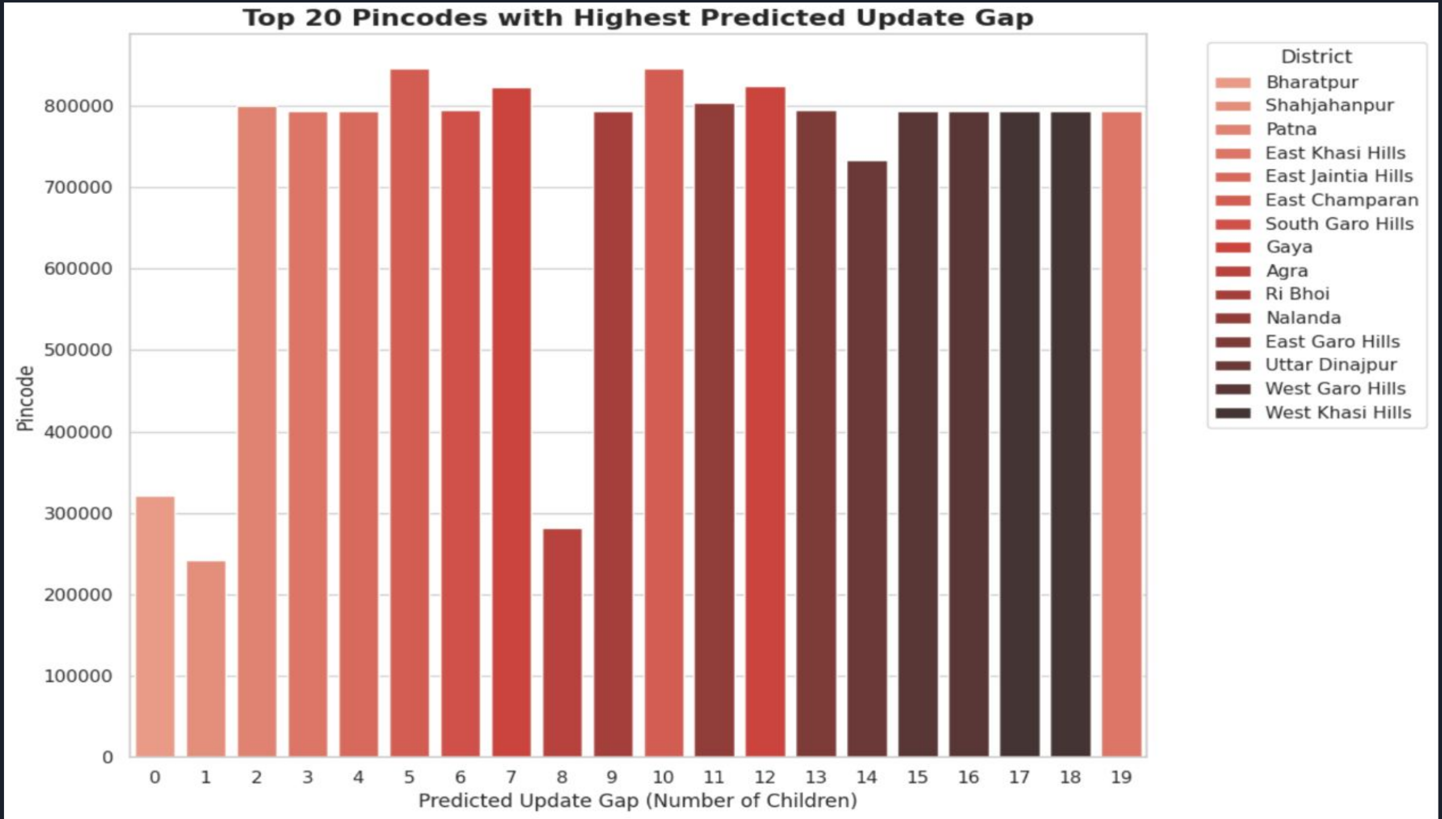
Despite performance limitations, model successfully identifies pincodes with highest predicted update_gap for **priority resource allocation** and mobile van deployment planning.

ML DRIVEN RESULTS





ML DRIVEN RESULTS

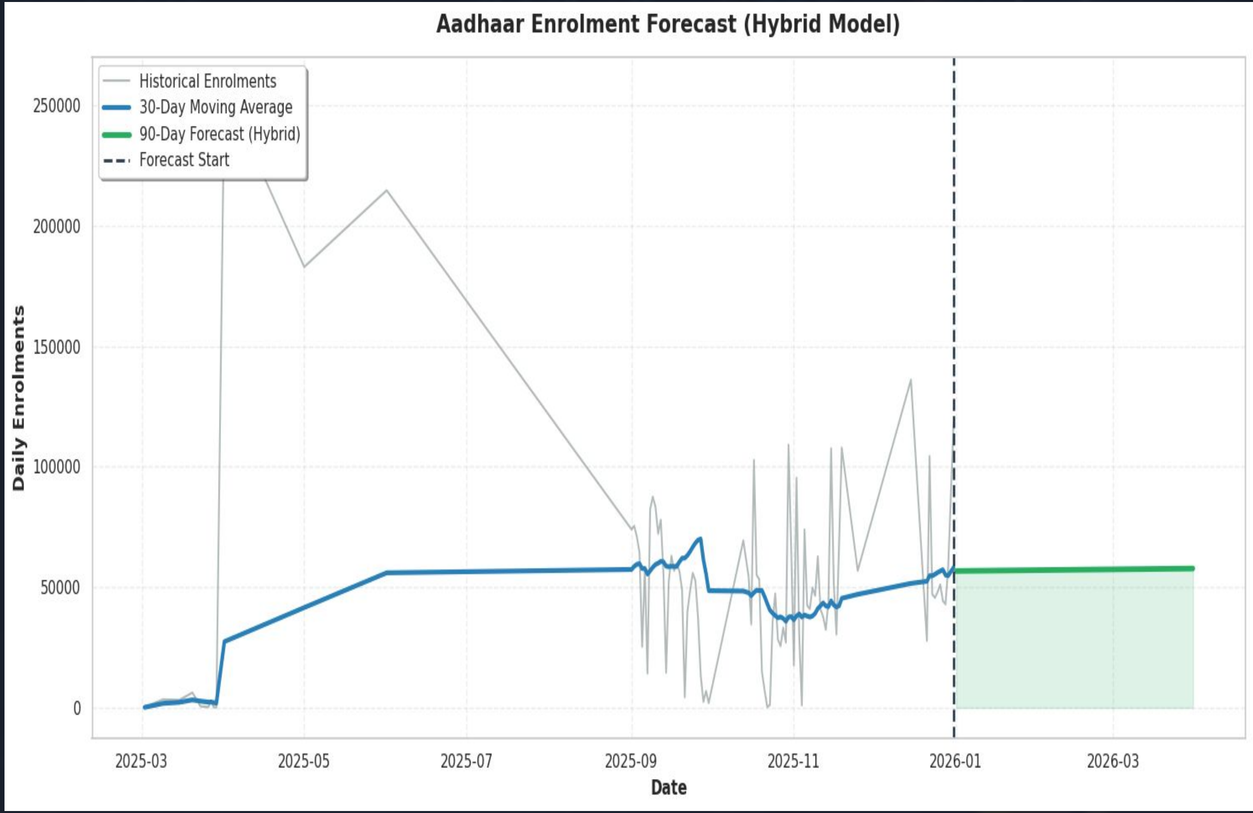


ML DRIVEN RESULTS

A hybrid forecasting approach combining a 30-day moving average with a recent linear trend was applied to model Aadhaar daily enrolments. The model smooths short-term volatility while retaining recent growth patterns, producing a stable 90-day forecast window.

Inference

The forecast indicates **steady enrolment volumes with controlled variability**, suggesting operational demand is expected to remain consistent in the near term. This enables **proactive capacity planning and resource allocation** for enrolment infrastructure across districts.



10 Strategic Recommendations

Actionable Interventions for UIDAI

STRATEGIC RECOMMENDATIONS

Priority 1-3: Immediate Actions (0-3 months)

🚨 Priority 1: Mobile Enrolment Van Deployment

Target

Deploy to critical pincodes identified by AMI analysis

Objective

Reach children missing mandatory biometric updates

Impact

Prevent future authentication failures for 4,902 at-risk children

🔍 Priority 2: Data Quality Audits

Target

Conduct audits in 764 high-risk districts (Trust Score < 40)

Objective

Pinpoint data entry errors, fraudulent enrollments, systemic issues

Impact

Improve data integrity and system reliability

⚠️ Priority 3: Biometric Refresh Camps

Target

Deploy in 15 high-friction districts (Friction Score > 1.0)

Objective

Address paper updates without biometric refresh

Impact

Prevent authentication failures at PDS/pension distribution

Immediate Impact: These three interventions target the most critical vulnerabilities: children missing biometric updates, widespread data quality issues, and the "paperwork trap" leading to authentication failures. Implementation within 3 months can prevent **thousands of citizens from service denial** at PDS shops and pension distribution points.

STRATEGIC RECOMMENDATIONS

Priority 4-6: Short to Medium-term Actions

Priority 4: Adult Enrolment Campaign

Current State	Target	Actions
Only 3.1% adult enrolment vs 96.9% children	Increase to 50% coverage within 6 months	Workplace drives, evening centers, digital outreach

Priority 5: Infrastructure Scaling (Migration Hubs)

Target	Signal	Actions
Top 10 migration hubs (Pune, Thane, Murshidabad, etc.)	High demographic update volume indicates migration pressure	Additional enrollment centers and staff in top 10 hubs

Priority 6: Hardware Upgrades (High-Failure Districts)

Target	Issue	Actions
10 districts with >50% operational failure rates	Exorbitant failure rates (7,500% to 17,100%) indicate hardware/training gaps	Biometric scanner upgrades, operator retraining, certification programs

Medium-term impact: These interventions address systemic gaps: demographic imbalance (96.9% children), infrastructure pressure from migration, and severe operational failures. Implementation over 6-12 months will create a **more inclusive, resilient, and efficient Aadhaar system** capable of serving all citizens effectively.

COMPREHENSIVE OVERVIEW

Master Overview Dashboard

Panel 1: Age Distribution

Pie Chart

0–5 years, 5–17 years, 18+ years breakdown showing 96.9% children dominance

Panel 2: Top 10 States

Horizontal Bar Chart

States ranked by total enrolments with value labels on bars

Panel 3: Update Volume

Bar Chart

Biometric vs Demographic update comparison (67.8M vs 36.1M)

Panel 4: Geographic Coverage

Text Summary Box

States: 39, Districts: 777, Pincodes: 19,412
Total Enrolments: 5.24M
Total Updates: 103.85M
Coverage Ratio: 19.82x

Panel 5: Risk Distribution

Bar Chart with Labels

High/Medium/Low risk districts with counts and percentages

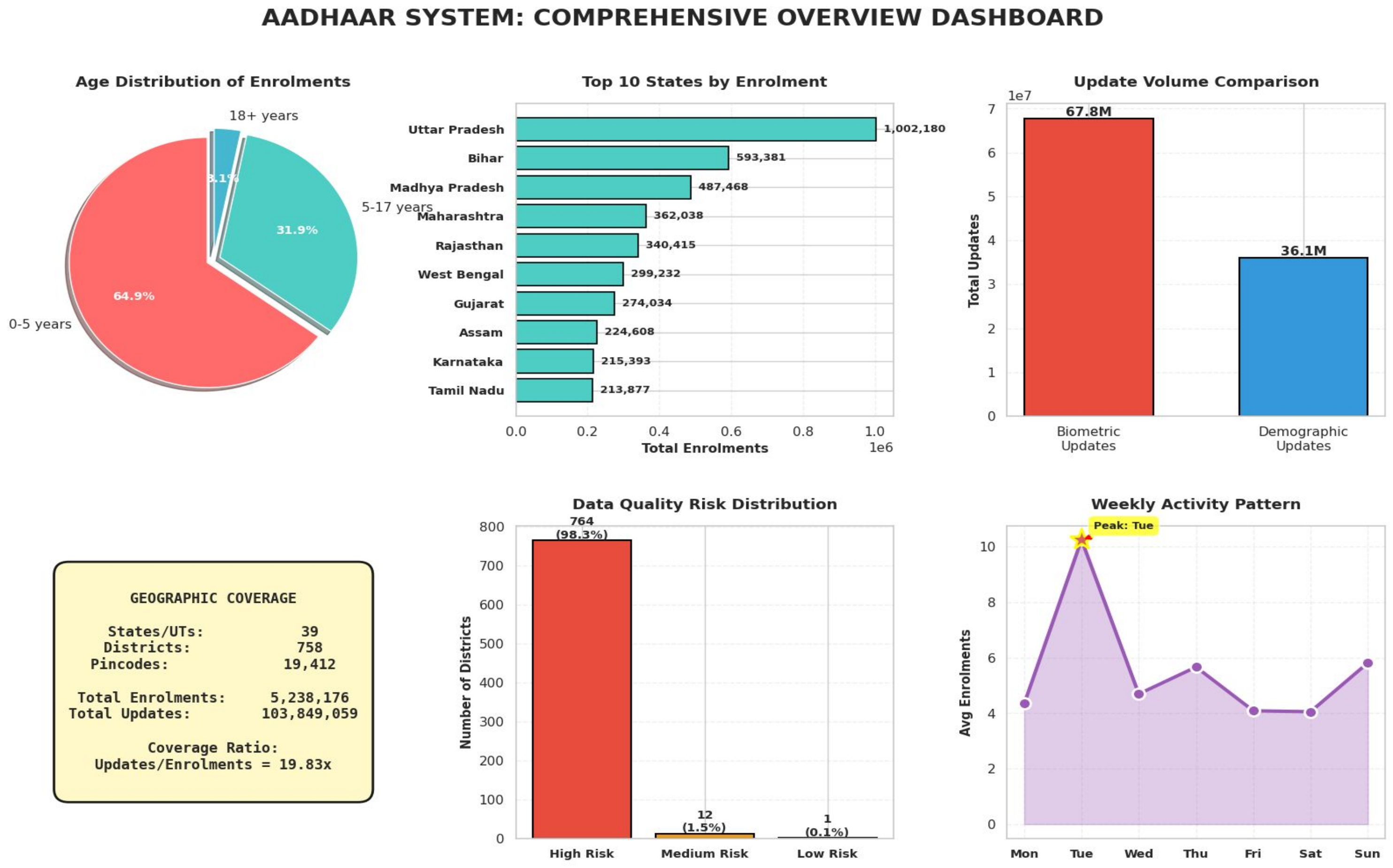
Panel 6: Weekly Activity

Line Chart

Average enrolments by day of week with peak day highlighted

Key Metrics Summary

5.24M	103.85M	777	19,412	19.82x
Total Enrolments	Total Updates	Districts	Pincodes	Updates/Enrolments


AADHAAR SYSTEM: COMPREHENSIVE OVERVIEW DASHBOARD
Age Distribution of Enrolments
18+ years
5-17 years
0-5 years
64.9%
31.9%
3.1%
Top 10 States by Enrolment
Uttar Pradesh 1,002,180
Bihar 593,381
Madhya Pradesh 487,468
Maharashtra 362,038
Rajasthan 340,415
West Bengal 299,232
Gujarat 274,034
Assam 224,608
Karnataka 215,393
Tamil Nadu 213,877
Total Enrolments
1e6
Update Volume Comparison
1e7
Total Updates
67.8M
36.1M
Biometric Updates
Demographic Updates
GEOGRAPHIC COVERAGE
States/UTs: 39
Districts: 758
Pincodes: 19,412
Total Enrolments: 5,238,176
Total Updates: 103,849,059
Coverage Ratio:
Updates/Enrolments = 19.83x
Data Quality Risk Distribution
Number of Districts
764 (98.3%)
12 (1.5%)
1 (0.1%)
High Risk
Medium Risk
Low Risk
Weekly Activity Pattern
Peak: Tue
Avg Enrolments
Mon
Tue
Wed
Thu
Fri
Sat
Sun

Impact Potential & Next Steps

At-Risk Children

4,902

Children missing mandatory biometric updates reachable through mobile van deployment to 99 update deserts.

Authentication Failures

40-60%

Reduction in authentication failures at PDS/pension distribution through biometric refresh camps in high-friction districts.

Data Integrity

764

High-risk districts targeted for data quality audits, improving overall system reliability and trust.

System Stability

+2.26

Trust score improvement indicates successful transition from mass enrolment chaos to stable maintenance phase.

Call to Action

Immediate (0-3 months)

Deploy mobile vans to update deserts, launch data audits in 764 high-risk districts, establish biometric refresh camps

Short-term (3-6 months)

Scale adult enrolment campaigns, expand infrastructure in migration hubs, complete hardware upgrades

Medium-term (6-12 months)

Establish continuous monitoring using AMI and trust score frameworks, scale successful interventions nationally

CODE LINK

COLAB LINK
https://drive.google.com/file/d/1N7yQ9SbX2K1VR4as5S6i1cVj8amEtylw/view?usp=sharing

GitHub REPO LINK
https://github.com/Harshithaviswanathan/UIDAI_DATA_HACKATHON_2026

COLAB CODE PDF LINK
https://drive.google.com/file/d/1ANRnvB5dAIMtyW60XncrqjUFrWo8oXnp/view?usp=sharing