Aadhaar Biometric Behaviour Intelligence System

Problem Statement
Unlocking societal trends in Aadhaar enrolment and updates by identifying meaningful patterns, anomalies, and predictive indicators to support informed decision-making and system improvements.

Project Overview
This project builds an AI-powered behavioural intelligence system that analyzes Aadhaar biometric and demographic update data to detect abnormal activity, identify high-risk regions, visualize trends, and support government and infrastructure-level decisions.

Datasets Used
Aadhaar Demographic Update Dataset
Aadhaar Biometric Update Dataset

Key columns: date, state, district, pincode, demo_age_5_17, demo_age_17_, bio_age_5_17, bio_age_17_

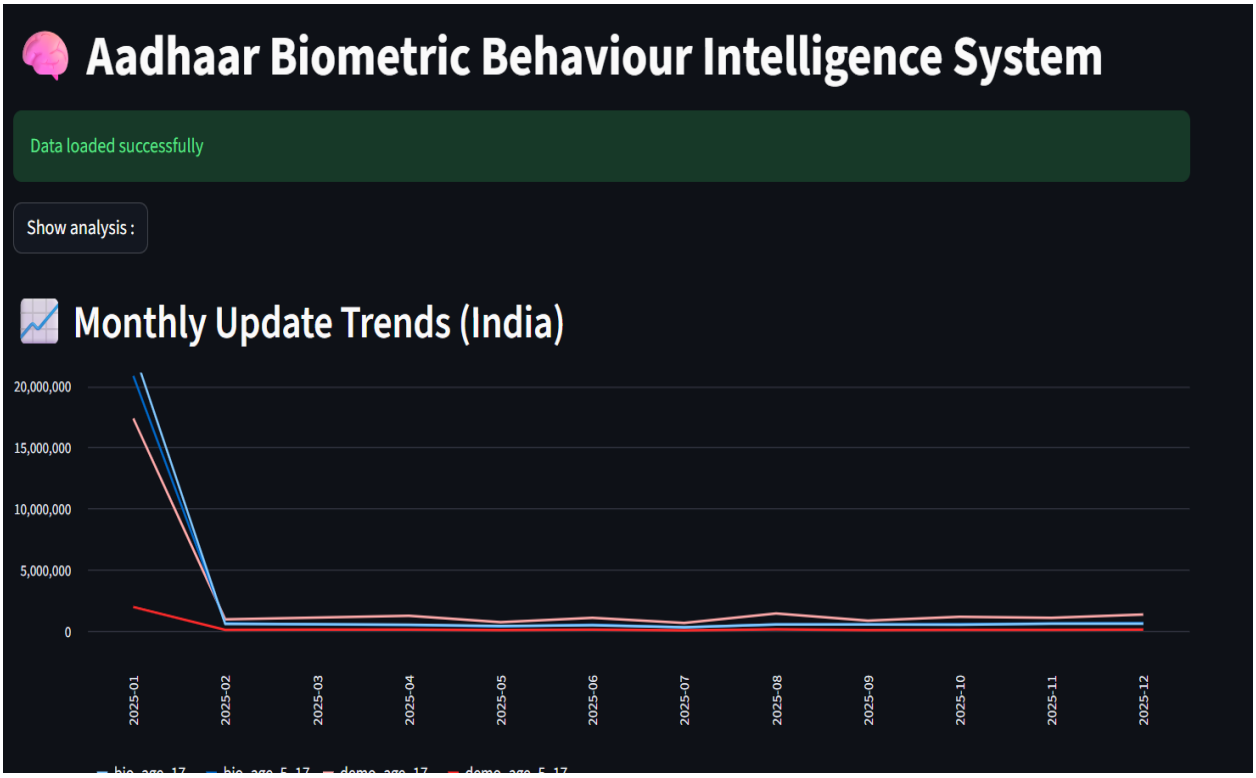
Data Pipeline
CSV merging → Data cleaning → Monthly aggregation → Feature engineering → ML anomaly detection → Risk scoring → Visualization via Streamlit dashboard

Feature Engineering
demo_total, bio_total, bio_demo_ratio, risk_score, risk_level

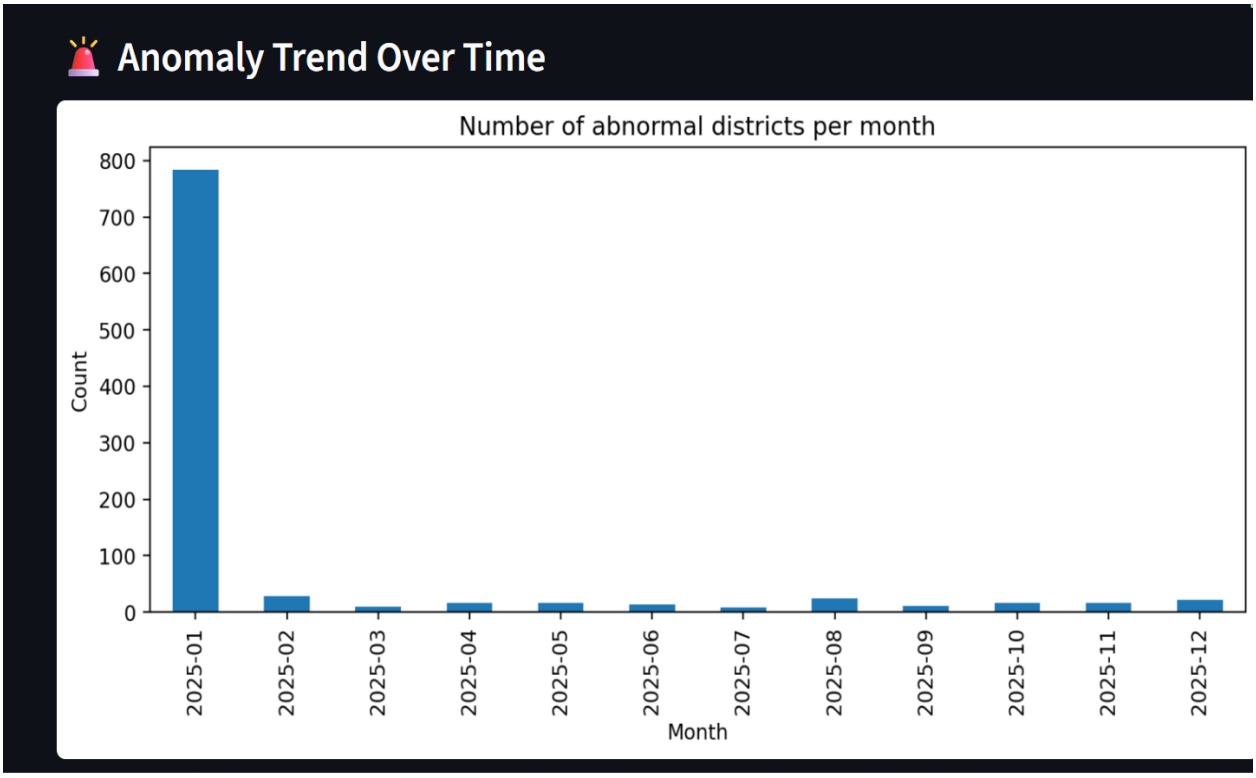
Machine Learning Model
Isolation Forest (unsupervised) used to detect abnormal biometric behaviour patterns without labeled data.

Key Insights
• Major anomaly spike in January 2025
• ~800 districts abnormal in same month
• Repeated abnormal behaviour in certain districts
• Multiple high-risk states identified

Dashboard Overview & Data Loading



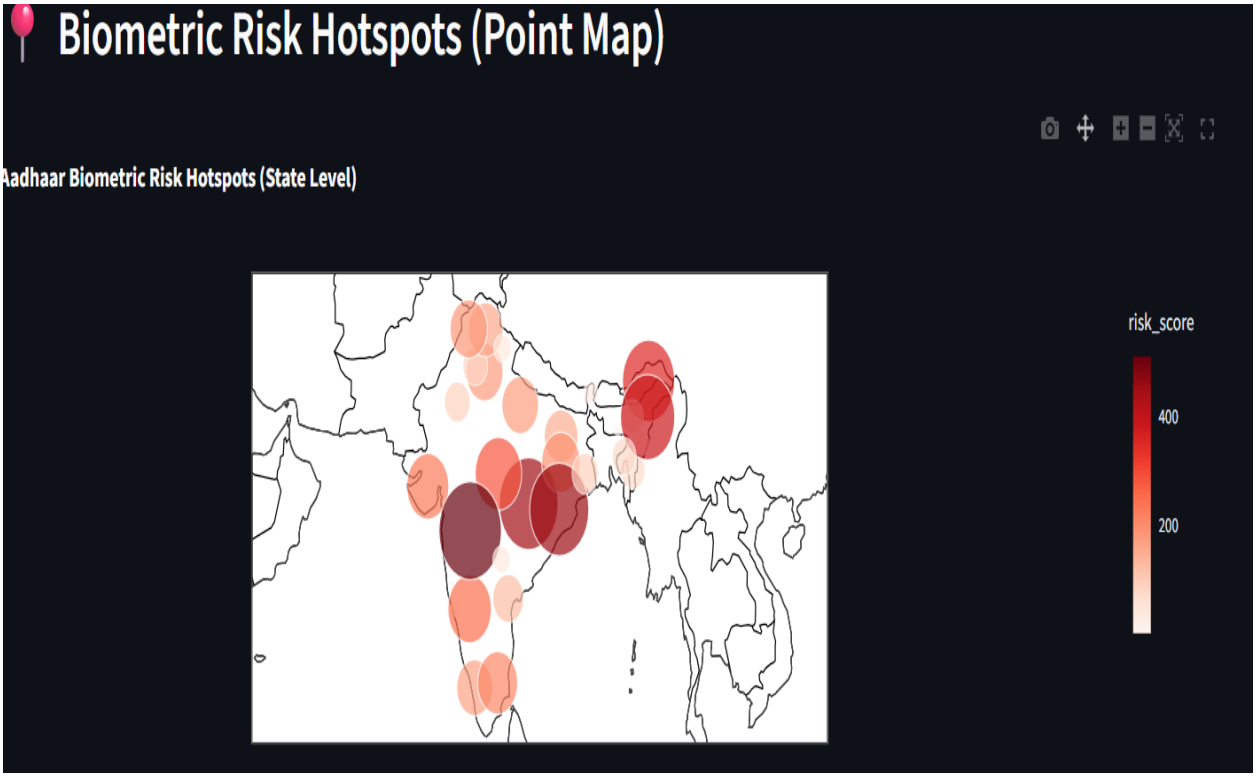
Monthly Update Trends (India)



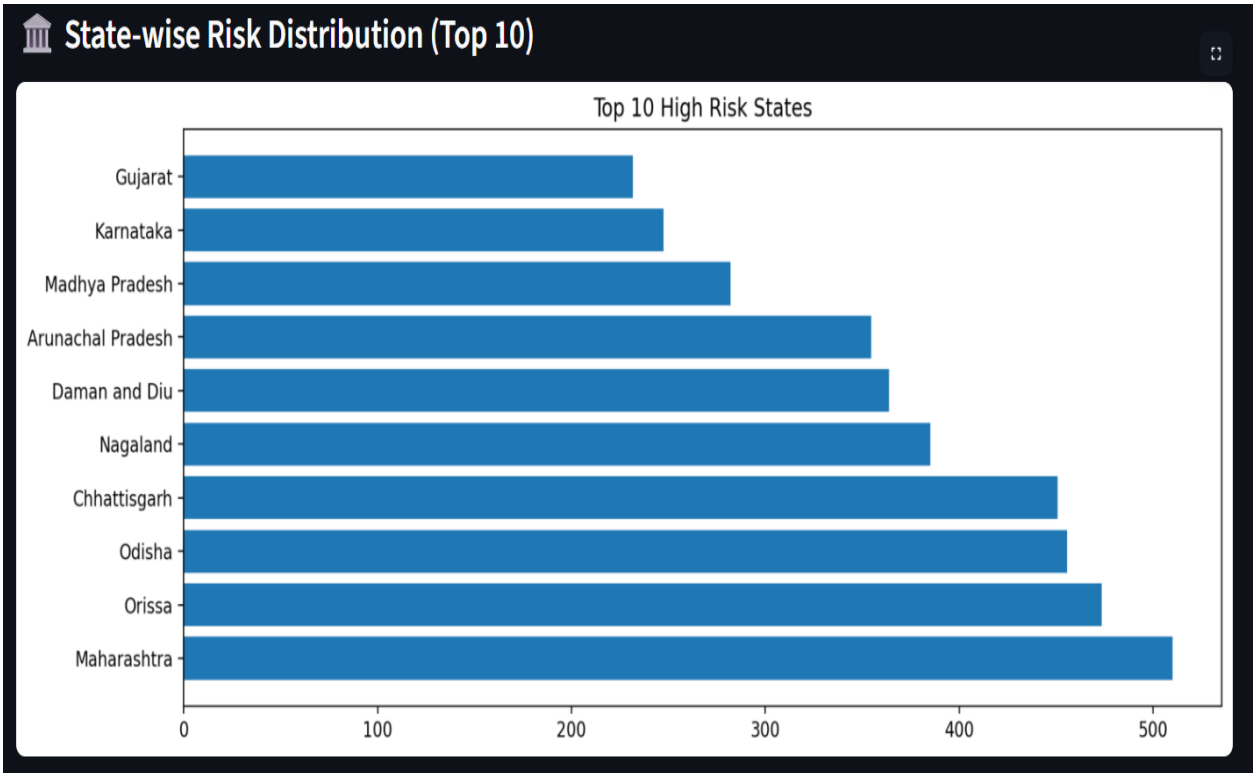
Anomaly Trend Over Time

	state	district	anomaly_count	avg_severity	max_severity	risk_score
160	Daman and Diu	Diu	1	352.8	352.8	352.8
144	Chhattisgarh	ManendragarhChirmiriBharatpur	1	312	312	312
163	Delhi	Najafgarh	1	148.5	148.5	148.5
158	Daman & Diu	Diu	1	143.3333	143.3333	143.33
495	Odisha	Baudh	1	121.84	121.84	121.84
46	Arunachal Pradesh	Kurung Kumey	3	40.0839	79.4516	120.25
355	Lakshadweep	Lakshadweep	1	113.5926	113.5926	113.59
478	Nagaland	Meluri	7	14.2578	19	99.8
154	Dadra & Nagar Haveli	Dadra & Nagar Haveli	1	92.6667	92.6667	92.67
507	Odisha	Jajpur	1	75.5	75.5	75.5

District Risk Ranking Table



Geographic Biometric Risk Hotspots Map



State-wise Risk Distribution (Top 10)

Districts with Highest Biometric Ratios

	month	state	district	bio_demo_ratio
209	2025-01	Daman and Diu	Diu	352.8
186	2025-01	Chhattisgarh	ManendragarhChirmiriBharatpur	312
212	2025-01	Delhi	Najafgarh	148.5
207	2025-01	Daman & Diu	Diu	143.3333
609	2025-01	Odisha	Baudh	121.84
437	2025-01	Lakshadweep	Lakshadweep	113.5926
201	2025-01	Dadra & Nagar Haveli	Dadra & Nagar Haveli	92.6667
58	2025-01	Arunachal Pradesh	Kurung Kumey	79.4516
622	2025-01	Odisha	Jajpur	75.5
595	2025-01	Nagaland	Shamator	70

Decision Support Framework
LOW: Routine monitoring
MEDIUM: Infrastructure review & validation
HIGH: Immediate investigation & resource allocation

Deployment
The system is deployed on Streamlit Cloud as an interactive web dashboard.

System Capabilities
• Large-scale data processing
• Behaviour modelling
• Machine learning anomaly detection
• Geographic intelligence
• Interactive analytics dashboard

Future Improvements
• District-level geo-mapping
• Time-series forecasting
• Real-time anomaly detection
• Automated alerting system
• Model explainability

Conclusion
This project demonstrates how Aadhaar update data can be transformed into actionable intelligence for national-level monitoring, planning, and system optimization using AI and data science.