

UIDAI DATA HACKATHON

Data-Driven Insights on Aadhaar Enrolment & Updates 2026

UIDAI_7370

Problem Statement

Unlocking Societal Trends in Aadhaar Enrolment and Updates Identify meaningful patterns, trends, anomalies, or predictive indicators and translate them into clear insights or solution frameworks that can support informed decision-making and system improvements.

PARTICIPANTS INFORMATION



RITIK SINGH

[guycomo](#)

NIT TRICHY
B.tech'22 - MECH



NEHAL MITTAL

[nehalsmittal](#)

NIT TRICHY
B.tech'22 - MME



TECH STACK



Python



GitHub



Google Collab



Numpy



Pandas



Matplotlib



Seaborn



Machine Learning



Glob



Prophet



Sarima

Datasets Used

• Aadhaar Enrolment Data	10,06,029	Jan 2025 - Dec 2025
• Aadhaar Demographic Data	20,71,700	Mar 2025 - Dec 2025
• Aadhaar Biometric Update Data	18,61,108	Mar 2025 - Dec 2025

ROWS

Enrolment Analysis

- **Merged** multi-part enrolment files into one dataset
- **Parsed dates** and filtered valid **2025 records**
- **Standardised state names** (spelling + UT mergers)
- **Removed invalid** and non-alphabetic locations
- **Created total enrolment** across age groups
- **Analysed state-wise** enrolment concentration
- **Computed age-group shares** (0-5, 5-17, 18+)
- Deep-dive **district analysis** for Uttar Pradesh

Demographic Analysis

- **Combined** all demographic update files
- **Fixed age-column naming** inconsistencies
- **Removed duplicate** transactional records
- **Cleaned and validated** state names
- **Filtered low-frequency/invalid states**
- **Derived** total demographic updates
- **Compared age-group** participation (5-17 vs 18+)
- **Analysed state-wise** and **monthly** trends

Bio - Update Analysis

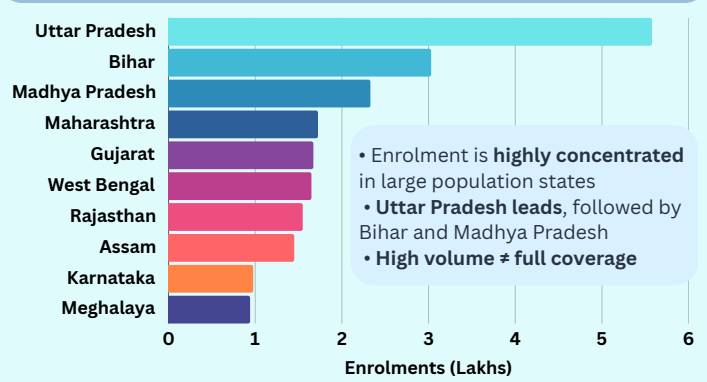
- **Concatenated** biometric update datasets
- **Corrected age-group column labels**
- **Parsed dates** and removed **duplicates**
- **Standardised state names** and UT boundaries
- **Derived total biometric update** metric
- **Compared** top vs bottom **states**
- **Analysed** age-group balance
- **Studied monthly** and **UP-level** trends

ADHAAR ENROLMENT ANALYSIS

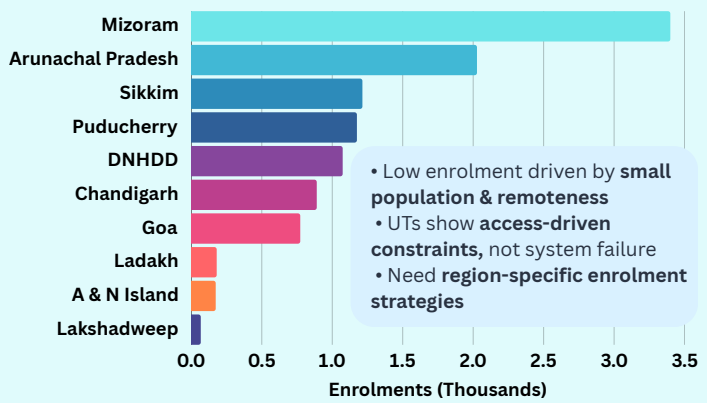
Date | State | District | Pincode | Age 0-5 | Age 5-17 | Age > 18 [Github Code Link](#) 4 Jan 2025 - 11 Dec 2025

Problem Statement To analyse Aadhaar enrolment patterns during 2025 across states and age groups, identify regional and early-age enrolment gaps, and provide data-driven insights to support targeted planning and policy interventions by UIDAI.

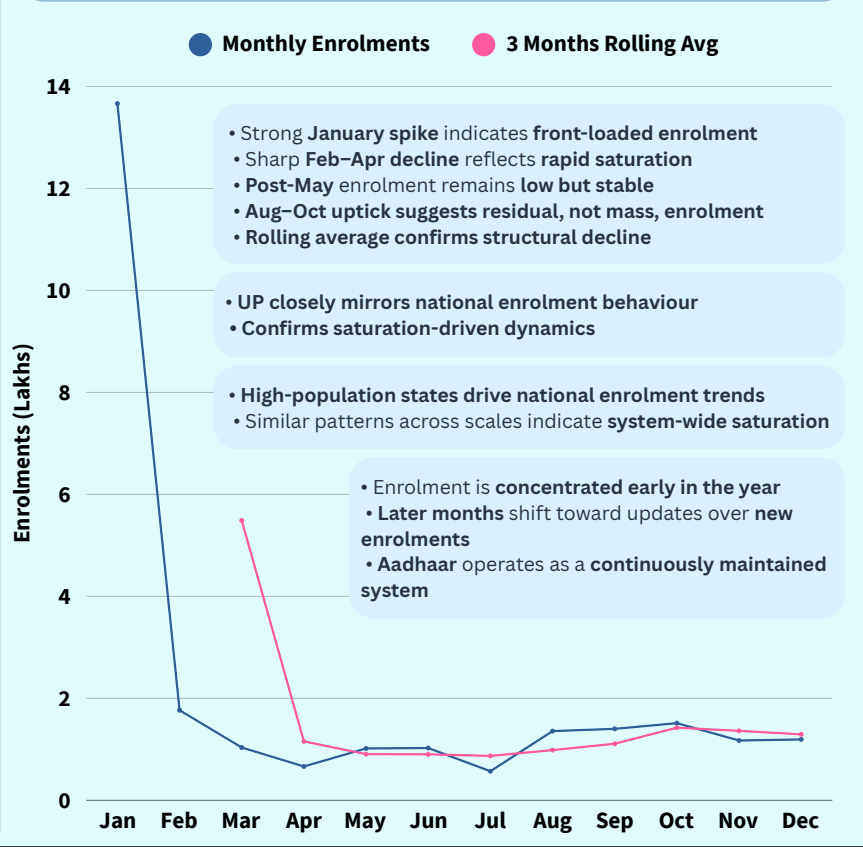
Top 10 States/UTs by Adhaar Enrolment



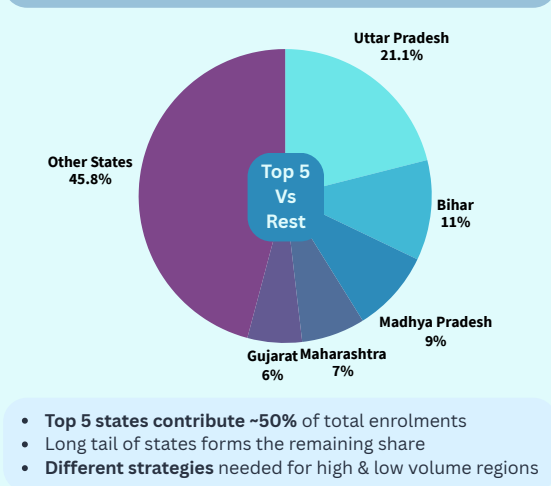
Bottom 10 States/UTs by Adhaar Enrolment



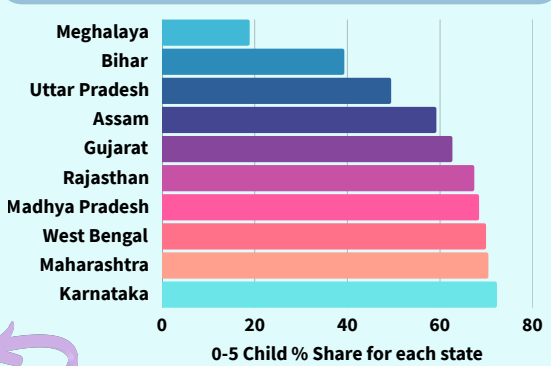
Monthly Enrolments Trends of all States/UTs



Share of Adhaar Enrolment Top 5 States vs Rest



Child(0-5) Adhaar Enrolment Share



Recommendations for UIDAI

- Launch targeted child enrolment drives in low-share states
- Integrate Aadhaar enrolment with birth registration systems
- Use Anganwadi and school networks for early enrolment
- Deploy mobile enrolment units in rural and underserved districts
- Monitor child enrolment as a key KPI alongside total enrolment

District-Level Insights - Uttar Pradesh

- Aadhaar enrolment within Uttar Pradesh is unevenly distributed across districts.
- A small number of districts contribute a disproportionately high share of total enrolments.
- Several districts show very low enrolment volumes, indicating possible access or awareness gaps.
- District-level variation highlights the importance of targeted operational planning

Why this Matters?

- Aadhaar enables access to welfare, health, and education
- Low child enrolment delays benefit access
- Regional gaps require targeted intervention

CONCLUSION

This analysis highlights significant regional and age-wise disparities in Aadhaar enrolment. While overall enrolment is high in populous states, child enrolment remains uneven across regions. Targeted, early-age enrolment strategies can help UIDAI achieve more inclusive and future-ready Aadhaar coverage.

ADHAAR DEMOGRAPHIC ANALYSIS

Date | State | District | Pincode | Age 5-17 | Age > 17

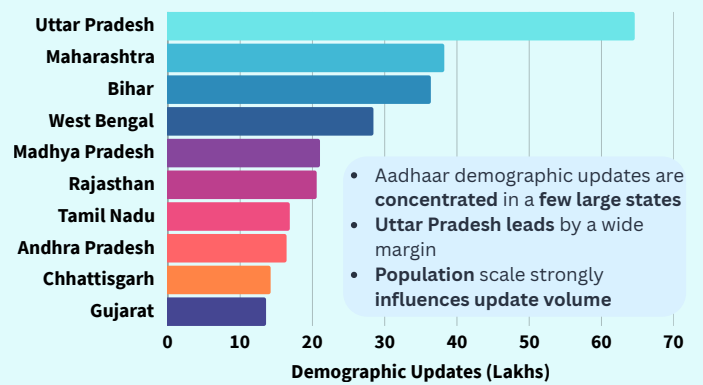
 [Github Code Link](#)

1 Mar 2025 - 29 Dec 2025

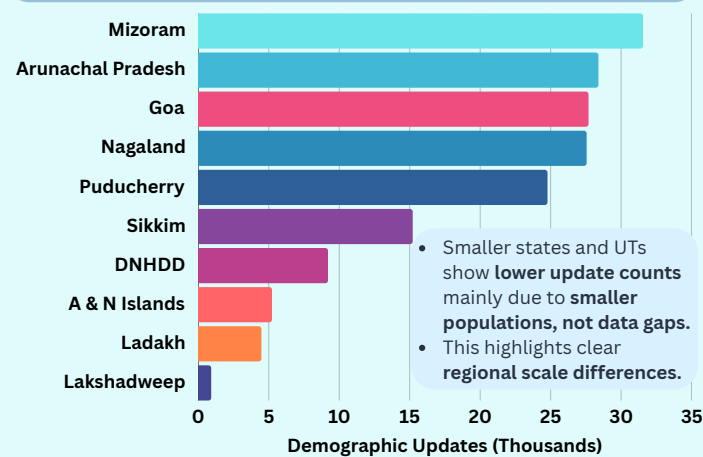
Problem Statement

This analysis examines Aadhaar demographic update patterns across states and age groups to identify regional and age-wise trends that support data-driven governance and service planning.

Top 10 States/UTs by Demographic Updates



Bottom 10 States/UTs by Demographic Updates

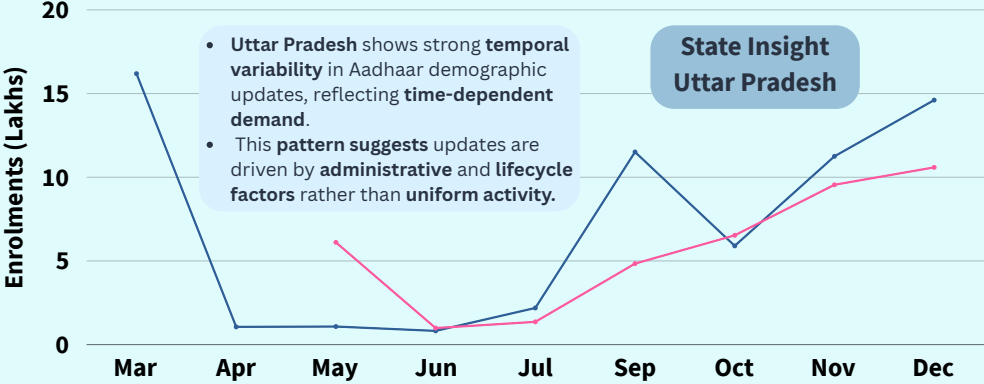
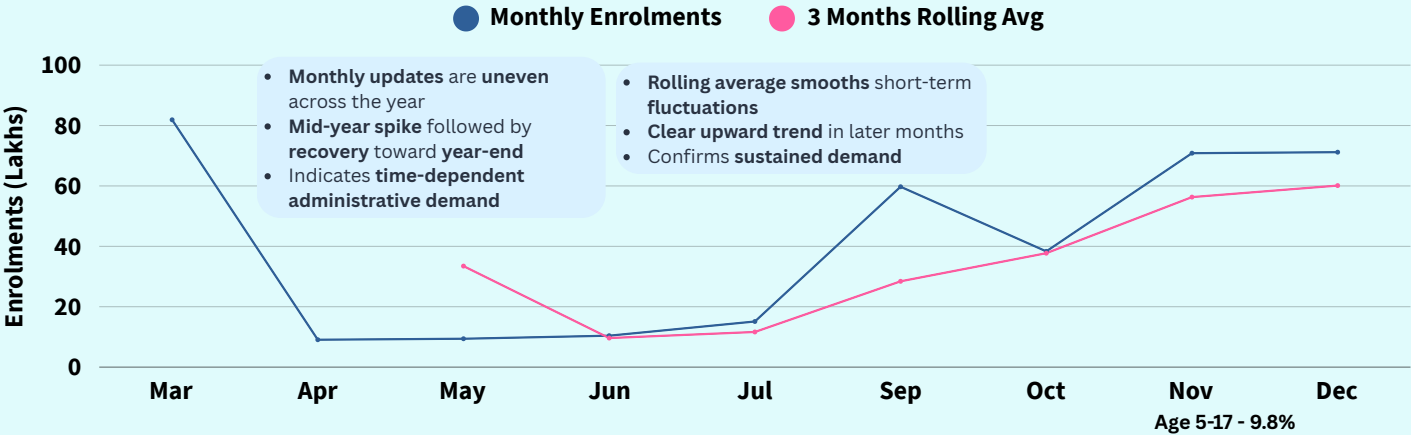


Limitations

- The dataset contains **transaction-level records**, where **multiple updates** by the same individual may occur.
- Location fields** are **partially free-text**, leading to **residual inconsistencies** despite **standardisation**.
- The **analysis does not link update trends to specific policies or events** due to **lack of external reference data**.
- Population normalisation was not applied**; results reflect **absolute update volumes**.

Monthly Demographic Update Trends

of all States/UTs



Age >18 - 90.2%

- 18+ group contributes over **90%** of updates
- Updates driven by **adult lifecycle events**
- 5-17 activity remains limited

Conclusion

- This analysis examined Aadhaar demographic updates across states, age groups, and time using **large-scale government data**.
- Adults (18+)** account for the **majority** of demographic updates, indicating that Aadhaar updates are **primarily driven by post-enrolment lifecycle needs**.
- A small number of **populous states dominate** overall update volumes, while smaller states and union territories show **lower counts due to population scale**.
- Monthly trends reveal non-uniform update activity**, highlighting periods of increased administrative demand.
- Overall, Aadhaar demographic updates represent a **continuous and dynamic lifecycle process**, rather than a **one-time administrative action**.

Future Work

- The data is **transaction-level**, so **multiple updates** may belong to the same individual, and **location fields** retain **minor inconsistencies**.
- Results reflect absolute update volumes**, as **population normalization and policy-level linkage** were not applied.

ADHAAR BIO-UPDATE ANALYSIS

Date | State | District | Pincode | Age 5-17 | Age > 17

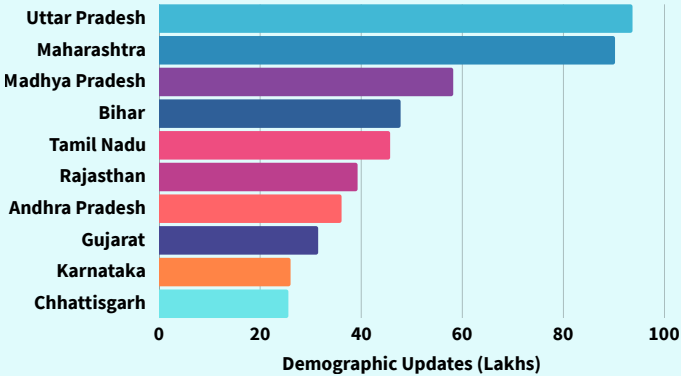
 [Github Code Link](#)

1 Mar 2025 - 29 Dec 2025

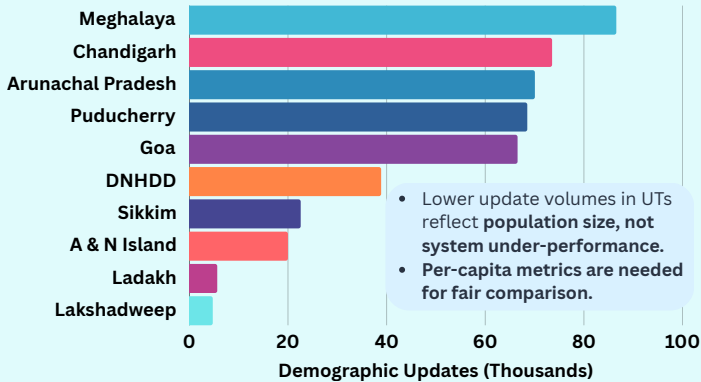
Problem Statement

This project analyses the Aadhaar lifecycle using enrolment, demographic, and biometric datasets to identify spatial, age-wise, and temporal patterns, and understand how service demand varies across regions and over time.

Top 10 States/UTs by Biometric Updates

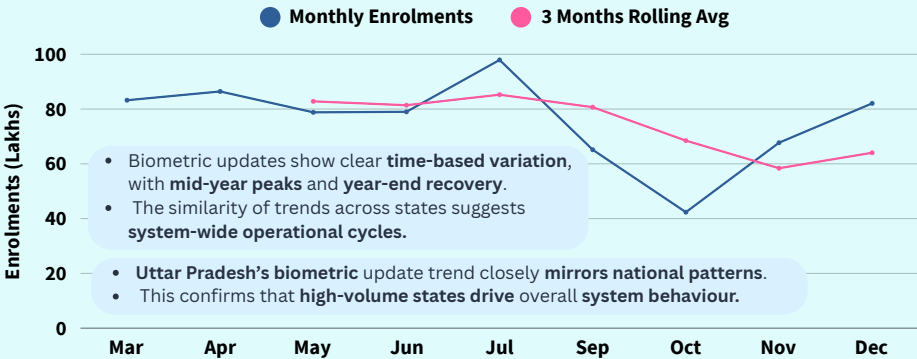


Bottom 10 States/UTs by Biometric Updates



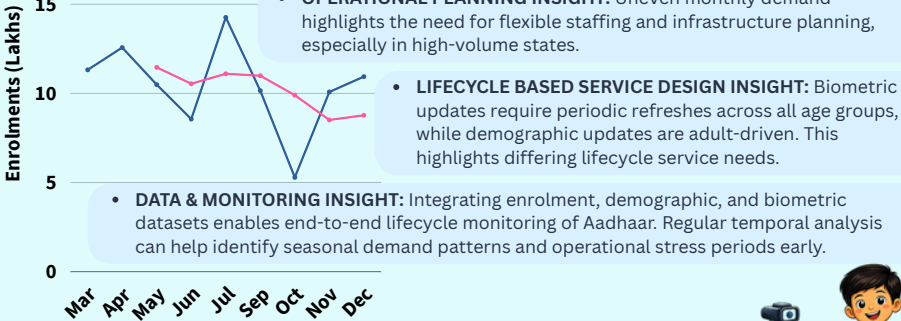
- Lower update volumes in UTs reflect **population size, not system under-performance**.
- Per-capita metrics** are needed for fair comparison.

Monthly Enrolments Trends of all States/UTs



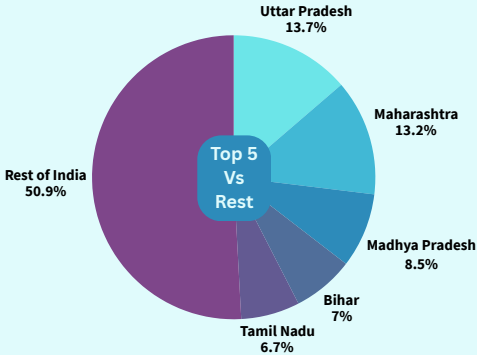
- Biometric updates show clear **time-based variation**, with **mid-year peaks** and **year-end recovery**.
- The similarity of trends across states suggests **system-wide operational cycles**.
- Uttar Pradesh's biometric update trend closely **mirrors national patterns**.
- This confirms that **high-volume states drive overall system behaviour**.

- SCALE CONSISTENCY INSIGHT:** National trends closely align with high-volume states like Uttar Pradesh, confirming scale consistency.



- OPERATIONAL PLANNING INSIGHT:** Uneven monthly demand highlights the need for flexible staffing and infrastructure planning, especially in high-volume states.
- LIFECYCLE BASED SERVICE DESIGN INSIGHT:** Biometric updates require periodic refreshes across all age groups, while demographic updates are adult-driven. This highlights differing lifecycle service needs.
- DATA & MONITORING INSIGHT:** Integrating enrolment, demographic, and biometric datasets enables end-to-end lifecycle monitoring of Aadhaar. Regular temporal analysis can help identify seasonal demand patterns and operational stress periods early.

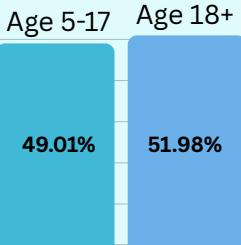
Share of Adhaar Enrolment Top 5 States vs Rest



- Biometric updates are **concentrated** in a few large states, with **Uttar Pradesh** and **Maharashtra** contributing the **highest volumes**.
- Lower update counts** in smaller states and UTs mainly reflect **population size, not data gaps**.

% Share of Age Group in Biometric Updates

- Biometric updates are **almost evenly distributed** between the 5-17 and 18+ age groups.
- This **reflects periodic biometric refresh needs** across all life stages rather than age-specific demand.



Limitations

- The datasets are **transactional** and may include **multiple records** per individual.
- Location fields** contain **free-text entries**, leading to **residual inconsistencies**.
- Analysis** is based on **absolute volumes**, without **population normalisation**.
- No direct causal inference** is made regarding policies or external events.

Recommendations

- Adopt **flexible operational planning** based on observed **monthly demand patterns**.
- Consider **periodic biometric refresh strategies** across **all age groups**.
- Use **per-capita metrics** for more **equitable regional comparisons**.
- Integrate lifecycle datasets** for **continuous performance** monitoring of Aadhaar services.

Conclusion

- Aadhaar is a dynamic, **lifecycle-driven identity system**, not a **static one-time process**.
- Enrolment ensures initial coverage, demographic updates track identity changes, and biometric updates **maintain authentication accuracy**.
- Integrated **spatial, demographic, and temporal analysis** highlights the **need for continuous service availability**, **adaptive operations**, and **data-driven governance**.



Recommendations

- Aadhaar operates as a **continuous identity lifecycle system** across enrolment, demographic, and biometric updates.
- Updates** ensure identity relevance and **authentication accuracy over time**.
- Temporal patterns** highlight the need for ongoing, **region- and time-specific operational planning**.

ADHAAR UPDATE DEMAND FORECAST

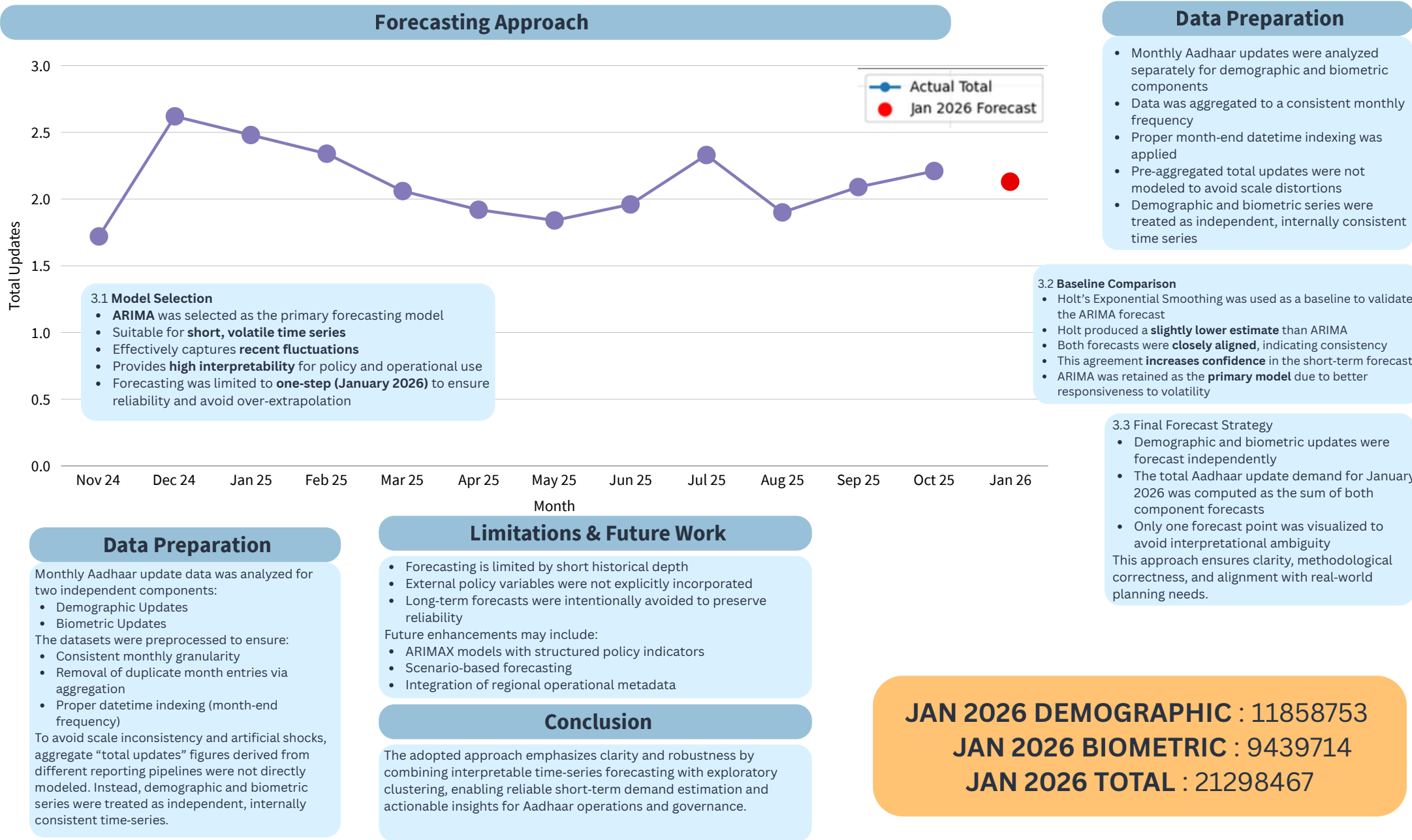
Time Series Forecast

 [Github Code Link](#)

January 2026 Forecast

Problem Statement

The objective of this study is to analyze historical Aadhaar update trends and generate a reliable short-term forecast to support near-term operational planning. Given the public-service nature of Aadhaar updates, the focus is on interpretability, data integrity, and conservative estimation, rather than long-term speculative forecasting.



STATE LEVEL ADHAAR OPERATIONAL CLUSTERING

Problem Statement National Aadhaar statistics mask regional differences in demand, stability, and growth. A one-size-fits-all operational strategy is insufficient at national scale, creating the need for data-driven analysis to identify regional behavior, high-risk states, and support proactive governance, capacity planning, and policy decision-making.

Objectives

- 1. Forecast short-term Aadhaar update demand using historical data.
- 2. Identify patterns and anomalies in Aadhaar activity across states.
- 3. Group states with similar operational behavior using unsupervised learning.
- 4. Provide interpretable and governance-oriented insights, rather than black-box predictions.
- 5. Demonstrate scalable data engineering and analytical rigor on real-world datasets.

Dataset Description

The analysis uses three primary datasets provided for the hackathon:

3.1 Enrolment Dataset

- Aadhaar enrolments by date, state, district, and pincode
- Age-wise enrolment distribution
- Captures the entry point of Aadhaar lifecycle

3.2 Demographic Update Dataset

- Demographic updates (name, address, DOB, etc.)
- Age-segmented counts
- Represents identity maintenance demand

3.3 Biometric Update Dataset

- Biometric updates (fingerprint, iris)
- Age-segmented counts
- Represents biometric quality and re-verification demand

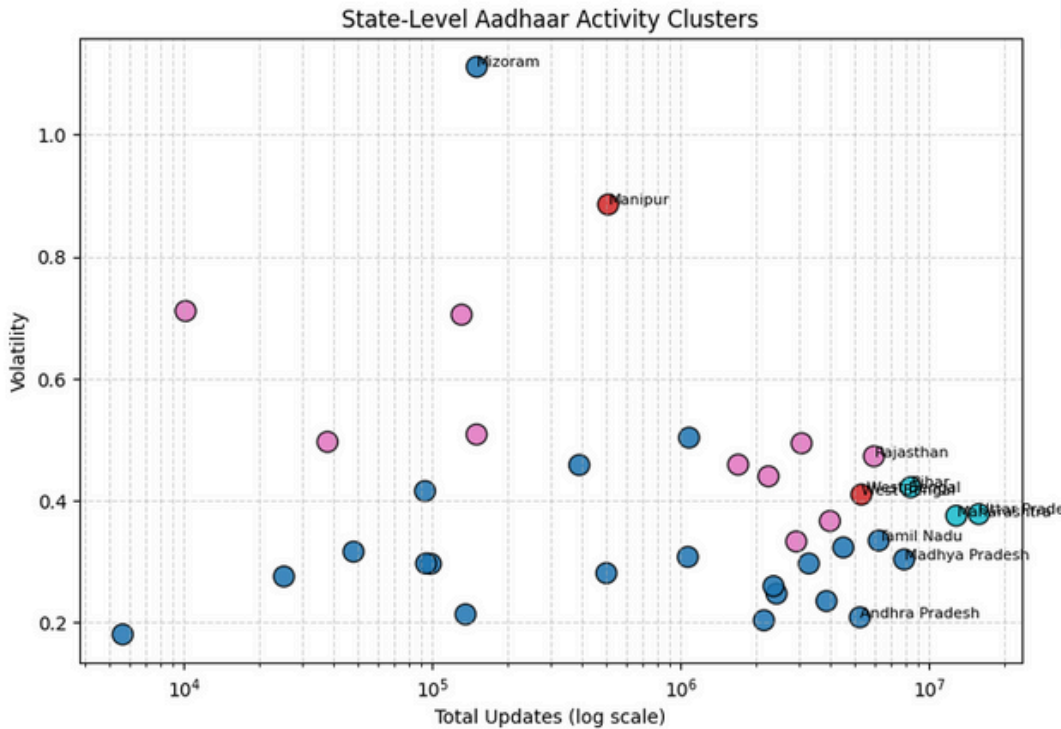
All datasets span multiple years and contain millions of records, requiring careful preprocessing and aggregation.

Data Preprocessing and Feature

- 4.1 Cleaning & Standardization
- Converted date fields to datetime format
 - Standardized state names to avoid duplication
 - Removed inconsistent casing and formatting issues
 - Handled missing values through aggregation-aware imputation
- 4.2 Aggregation Strategy To avoid noisy row-level joins and ensure scalability:
- Data was aggregated to state-month level
 - Final analysis used state-level summaries
- 4.3 Engineered Features For each state, the following features were computed:
- Total Enrolment
 - Total Demographic Updates
 - Total Biometric Updates
 - Demo Ratio (demographic share of updates)
 - Bio Ratio (biometric share of updates)
 - Volatility (month-to-month instability in updates)
 - Growth Rate (change in demand over time)
- These features capture scale, composition, stability, and trend — key dimensions of operational behavior.

State Level Clustering (Unsupervised Learning)

- 6.1 Why Clustering?
- Forecasting alone does not explain why demand differs across regions. Clustering enables discovery of latent behavioral patterns without predefined labels.
- 6.2 Methodology
- Features were standardized using StandardScaler
 - K-Means clustering was applied
 - Number of clusters selected for interpretability and separation
- 6.3 Visualization Strategy
- X-axis: Total Updates (log scale)
 - Y-axis: Volatility
 - Color: Cluster membership
- This visualization highlights operational load vs instability, a critical governance dimension.



Conclusion

1. This project demonstrates how large-scale Aadhaar operational data can be transformed into actionable intelligence through a combination of time-series forecasting and unsupervised clustering.
2. Rather than relying solely on aggregate statistics or black-box models, the approach emphasizes:
3. Interpretability
4. Operational relevance
5. Real-world data challenges
6. By identifying distinct state-level behavior patterns, the system enables data-driven decision-making, supporting scalable, efficient, and region-aware Aadhaar governance.

State-Level Aadhaar Activity (Analysis View)

INSIGHT 3: Mid-scale states show the widest behavior spread

States like:

- Assam
- Jharkhand
- Delhi
- Chhattisgarh

Have similar total updates, but very different volatility.

Volume alone is insufficient for planning, stability metrics are critical.

INSIGHT 1: Volatility is NOT proportional to scale (Very strong)

Look at the top of the graph:

Mizoram → extremely high volatility

Manipur → very high volatility

But both have low total updates.

Small states can still be operationally risky due to sudden Aadhaar activity spikes.

INSIGHT 2: High-update states split into two distinct behaviors

On the right side (high total updates):

Stable high-volume states:

- Andhra Pradesh
- Tamil Nadu
- Karnataka

These lie low on volatility.

These states have mature Aadhaar ecosystems with predictable demand.

High-volume but unstable states

- Uttar Pradesh
- West Bengal
- Rajasthan

These are higher on volatility despite large scale.

Large Aadhaar states cannot be treated uniformly, some need dynamic resource allocation.

INSIGHT 5: Ladakh & Chandigarh are special cases

- Moderate volatility
- Low scale
- Isolated behavior

Union Territories may show distinct Aadhaar usage patterns driven by administrative policies rather than population size.

INSIGHT 4: Very low-volume & stable states exist (baseline cluster)

- Lakshadweep
- Andaman & Nicobar
- Nagaland

Low updates, low volatility.

These regions require minimal intervention and can be managed with baseline operational capacity.

Volatility

1.0

0.8

0.6

0.4

0.2

10^4

10^5

10^6

10^7

Total Updates (log scale)

Mizoram

Manipur

Ladakh

Chandigarh

Sikkim

Meghalaya

Puducherry

Tripura

Uttarakhand

Assam

Delhi

Jharkhand

Rajasthan

West Bengal

Bihar

Chhattisgarh

Madhya Pradesh

Uttar Pradesh

Telangana

Gujarat

Tamil Nadu

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

Kerala

Punjab

Haryana

Uttar Pradesh

Madhya Pradesh

Chhattisgarh

Tamil Nadu

Gujarat

Odisha

Karnataka

Andhra Pradesh

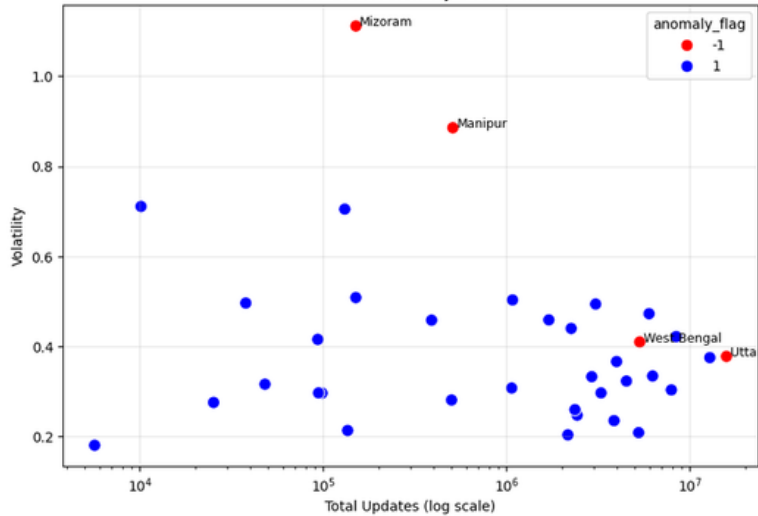
Kerala

STATE LEVEL ANOMALIES

Problem Statement

This notebook identifies anomalous state-level Aadhaar update behavior and localized district- and pincode-level hotspots to support targeted operational planning.

State-Level Anomaly Detection



Insight 3: High-volume \neq high volatility

- Some pincodes have:
- Very high total_updates
- Moderate volatility
- High demand regions are not necessarily unstable; instability is localized.

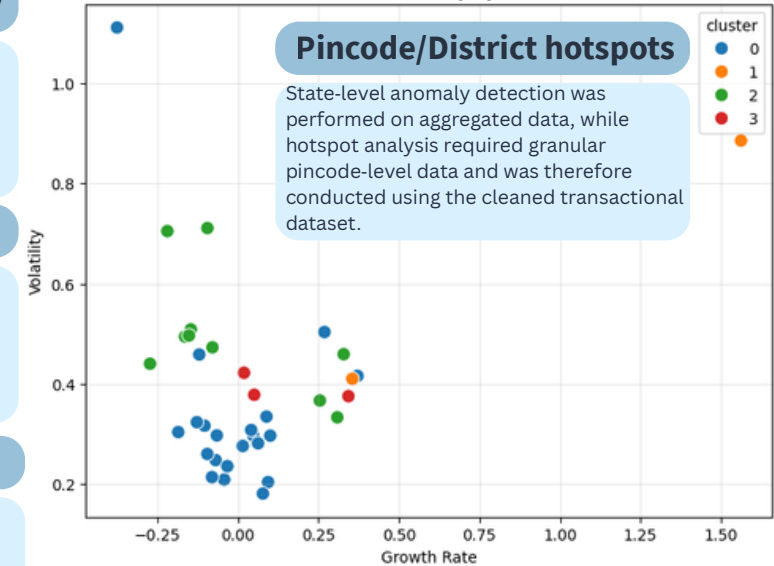
Insight 4: Zero-activity pincodes are

- Especially in:
- Rural
- Newly created
- Administrative pincodes
- Zero-activity pincodes reflect structural or demographic factors, not data issues.

Insight 1: Urban concentration

- Many top hotspots are:
 - Delhi (1100xx)
 - Uttar Pradesh (24xxxx, 20xxxx)
 - Maharashtra (43xxxx, 42xxxx)
- Aadhaar update demand is highly concentrated in dense urban and peri-urban pincodes.

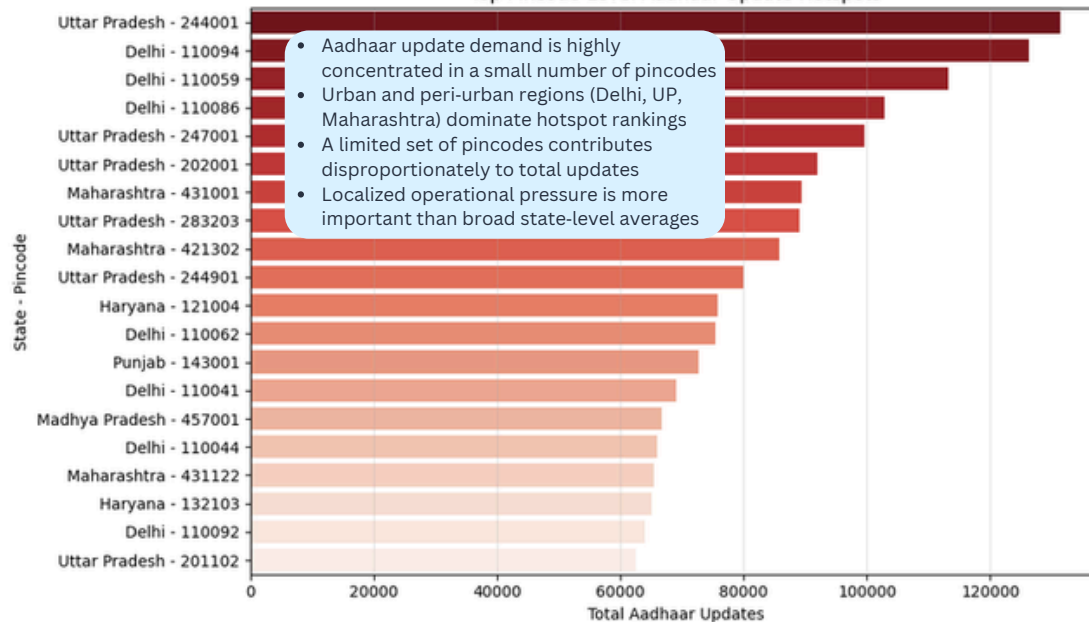
Growth vs Volatility by Cluster



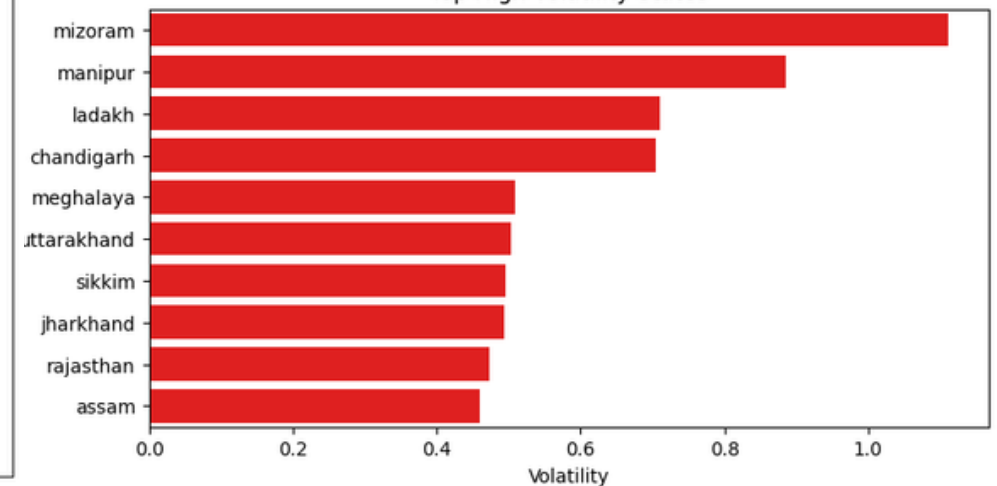
Insight 2: Few pincodes dominate volume

- Top ~20 pincodes contribute lakhs of updates, while most pincodes contribute very little.
- A small fraction of pincodes accounts for a disproportionate share of total Aadhaar activity.

Top Pincode-Level Aadhaar Update Hotspots



Top High-Volatility States



CONCLUSION

- **Multi-Layered Forecasting**
- Utilizes time-series models to project short-term demand.
- Facilitates near-term capacity planning for manpower and hardware allocation.
- **State-Level Operational Clustering**
- Groups states by behavioral patterns (stability, growth, and update types).
- Identifies regional differences in how citizens interact with the Aadhaar system.
- **Anomaly Detection & Risk Assessment**
- Flags states with unusually high volatility or outlier growth rates.
- Designates "operational risk zones" to preempt system bottlenecks or service failures.
- **Granular Hotspot Analysis**
- Performs deep dives at district and pincode levels.
- Reveals spatial concentration, where a few geographic units drive the majority of total activity.
- **Strategic Governance Shift**
- Moves away from uniform national planning toward targeted, data-driven interventions.
- Redefines operational risk as a combination of scale, volatility, and location.