Goverment Of India

Unique Identification Authority of India (UIDAI)

ANALYSIS OF AADHAAR ENROLMENT AND UPDATE PATTERNS FOR OPERATIONAL WORKLOAD INSIGHT AND SYSTEM IMPROVEMENT

A Data-Driven Analytical Framework Using UIDAI Aggregated Datasets

Submitted as part of the Data Analytics Challenge hosted on data.gov.in

January 2026

Submitted by:

Priyanshu Singh

Kriti Anand Singh

Nitish Kumar Bind

Affiliated Institute: Dr. Shyama Prasad Mukherjee International Institute of Information Technology, Naya Raipur

1. Executive Summary

The Aadhaar ecosystem is a critical pillar of India's digital public infrastructure, supporting identity verification for a wide range of government services and welfare programs. Due to the scale of the population covered and the continuous requirement for demographic and biometric updates, enrolment and update operations place varying levels of operational demand across different regions and time periods. Understanding how this demand is distributed is essential for improving service efficiency and preventing infrastructure stress.

This study analyzes UIDAI-provided aggregated Aadhaar enrolment, demographic update, and biometric update datasets to identify meaningful patterns related to workload concentration, update intensity, and demographic influence. By examining enrolment and update activity across temporal, geographic, and age-based dimensions, the analysis moves beyond absolute counts and introduces normalized indicators that highlight regions experiencing disproportionate operational pressure. The findings reveal that Aadhaar enrolment activity is highly concentrated in a limited number of states, while demographic and biometric update workloads vary independently of enrolment volume. Several states exhibit significantly higher update pressure relative to their enrolment share, indicating hidden operational stress points that are not visible from enrolment data alone. Age-wise analysis further shows that although enrolment activity is dominated by younger age groups, the majority of update workload is driven by the adult population, underscoring the lifecycle-driven nature of Aadhaar data maintenance.

From an implementation perspective, this study proposes a data-driven framework for identifying high-pressure regions and periods within the Aadhaar ecosystem. The insights generated can support proactive resource allocation, targeted staffing, and infrastructure planning by enabling administrators to anticipate enrolment- and update-driven workload surges. The analytical framework is scalable and can be extended into real-time dashboards or early warning systems, supporting continuous monitoring and evidence-based decision-making for system improvements within UIDAI operations.

2. Problem Statement and Analytical Approach

2.1 Problem Statement

The Aadhaar programme functions as a foundational digital identity system that supports a wide range of public services and welfare delivery mechanisms across India. Due to the scale of coverage and the diversity of geographic and demographic contexts, Aadhaar enrolment and update operations generate substantial and continuous operational demand. Managing this demand efficiently is a critical administrative challenge.

A key limitation in current operational planning is the lack of a systematic, data-driven mechanism to identify regions experiencing disproportionate workload pressure. While enrolment volumes are often used as a primary indicator for resource allocation, they do not fully capture the operational burden created by frequent demographic and biometric updates. In several regions, update activity may be significantly higher relative to enrolment levels, leading to service delays, infrastructure strain, and inefficient utilization of enrolment and update centers.

Without analytical indicators that normalize update activity against enrolment volume and demographic composition, these hidden stress points remain difficult to detect. As a result, administrative decisions related to staffing, infrastructure deployment, and scheduling are often reactive rather than proactive.

The problem addressed in this study is the absence of an integrated analytical framework that can (a) identify uneven distribution of enrolment and update workloads across states and time periods, (b) quantify update-driven operational pressure relative to enrolment activity, and (c) translate these patterns into actionable insights for improved planning and system optimization. Addressing this gap is essential for enabling evidence-based decision-making and ensuring efficient, timely Aadhaar service delivery at scale.

2.2 Analytical Approach

To address the identified problem of uneven enrolment and update workload distribution, this study adopts a structured, exploratory analytical approach with a strong emphasis on normalization, comparability, and decision relevance. Rather than relying solely on absolute enrolment or update counts, the approach is designed to uncover hidden operational pressure by relating update activity to enrolment volume and demographic composition.

The analysis is conducted across three primary dimensions: time, geography, and age group. Temporal analysis is used to identify fluctuations and periodic surges in enrolment and update activity, while geographic analysis focuses on state-level concentration and disparity in workload distribution. Age-group analysis provides additional context by examining how demographic composition influences enrolment patterns and lifecycle-driven update behavior.

A key component of the analytical approach is the derivation of normalized indicators, such as demographic update intensity and biometric update pressure. These indicators are computed by relating update counts to corresponding enrolment volumes, enabling meaningful comparison across states with significantly different population sizes and enrolment scales. This normalization allows the identification of regions where update workload is disproportionately high, even when absolute enrolment numbers appear moderate.

Exploratory data analysis techniques, including univariate, bivariate, and trivariate analysis, are applied to capture both isolated trends and interaction effects between enrolment, updates, time, and age groups. Visual analytics are used extensively to support interpretability and to clearly communicate patterns related to workload concentration, temporal variability, and demographic influence.

Overall, the analytical approach is designed not only to describe historical patterns but also to support actionable insights. By transforming raw enrolment and update data into interpretable pressure indicators, the approach enables administrators to move from reactive monitoring to proactive operational planning and system optimization within the Aadhaar enrolment and update framework.

3. Proposed Solution Framework

To translate analytical insights into actionable system improvements, this study proposes an integrated solution framework that enables continuous monitoring of Aadhaar enrolment and update workload across regions and time periods. The framework is designed to move from descriptive analysis toward operational decision support by systematically identifying high-pressure regions and periods.

3.1 Data Aggregation and Alignment Layer

At the foundation of the framework, aggregated Aadhaar enrolment, demographic update, and biometric update datasets are aligned temporally and geographically at the state–month level. This ensures consistency across datasets and enables integrated analysis of enrolment and update activity without reliance on individual-level data, thereby maintaining privacy compliance.

3.2 Normalized Indicator Computation

Instead of using absolute counts alone, the framework computes normalized workload indicators that capture relative operational pressure. Key indicators include demographic update intensity and biometric update pressure, calculated by relating update volumes to corresponding enrolment counts. These indicators allow meaningful comparison across states with varying enrolment scales and population sizes.

3.3 Workload Stress Identification

Using the derived indicators, states and time periods are evaluated to identify disproportionate workload conditions. Regions exhibiting high update pressure relative to enrolment volume are flagged as potential operational stress points. This step helps surface hidden pressure areas that may not be evident from enrolment data alone.

3.4 Temporal Surge Detection

The framework incorporates time-series analysis to detect periodic surges in enrolment or update activity. By monitoring changes in normalized indicators over time, the system can identify emerging workload spikes and seasonal patterns, enabling early intervention rather than reactive response.

3.5 Decision Support and Administrative Action Mapping

The identified stress regions and surge periods are mapped to potential administrative actions such as targeted staffing, temporary infrastructure scaling, or specialized update-focused service drives. This transforms analytical outputs into decision-support signals that can directly inform operational planning.

3.6 Scalability and Extension

The framework is designed to be scalable and adaptable. It can be extended to finer geographic levels (district or PIN code), integrated into real-time dashboards, or enhanced with predictive models to forecast future workload. Such extensions can further strengthen proactive planning and system optimization within the Aadhaar ecosystem.

Overall, the proposed framework bridges the gap between data analysis and operational decision-making by providing a structured mechanism to identify, monitor, and respond to enrolment- and update-driven workload pressure in a timely and evidence-based manner.

4. Datasets Used

This study is based on official datasets released by the Unique Identification Authority of India (UIDAI) related to Aadhaar enrolment and update activities. All datasets used in the analysis are aggregated in nature and publicly available, ensuring compliance with privacy and data protection requirements. The datasets collectively enable the examination of enrolment patterns, demographic changes, and biometric update behavior across geographic regions and time periods.

4.1 Aadhaar Enrolment Dataset

The Aadhaar Enrolment dataset provides aggregated information on Aadhaar enrolments across multiple geographic and demographic levels. It includes attributes such as the date of enrolment, state, district, PIN code, and age-wise enrolment categories. The age-wise classification segments enrolments into three groups: children aged 0–5 years, individuals aged 5–17 years, and adults aged 18 years and above.

This dataset captures both temporal and spatial variations in enrolment activity, making it suitable for analyzing long-term trends, seasonal fluctuations, and regional concentration of enrolments. In

this study, the enrolment data is primarily used to examine state-wise and monthly enrolment patterns and to derive age-group-wise enrolment contributions. The dataset also serves as the baseline against which update-related workload indicators are normalized.

The enrolment data used in this analysis corresponds to records available up to 31st December 2025, including the latest state-wise aggregated information.

4.2 Aadhaar Demographic Update Dataset

The Aadhaar Demographic Update dataset contains aggregated information related to changes made to residents' demographic details linked to Aadhaar. These updates include modifications to attributes such as name, address, date of birth, gender, and mobile number. The dataset reflects the dynamic nature of Aadhaar records, where demographic details may change over time due to life events or data corrections.

This dataset is available at multiple geographic levels, including state, district, and PIN code, and is updated periodically. In the context of this study, demographic update data is used to analyze the frequency and distribution of demographic changes across states and time periods. By comparing demographic update volumes with enrolment counts, the analysis highlights regions experiencing relatively higher demographic update activity, which may indicate increased operational workload.

The demographic update data considered in this analysis includes aggregated records available up to 31st December 2025, using the latest available state-wise information.

4.3 Aadhaar Biometric Update Dataset

The Aadhaar Biometric Update dataset provides aggregated information on updates made to residents' biometric attributes, including modalities such as fingerprints, iris scans, and facial data. Biometric updates are particularly significant in cases such as children transitioning into adulthood, where biometric characteristics undergo natural changes and require periodic revalidation.

This dataset captures the volume and distribution of biometric updates across different regions and time periods. In this study, biometric update data is used to assess biometric update pressure by comparing update volumes relative to enrolment levels. Such analysis helps identify states or periods where biometric update activity is disproportionately high, potentially indicating greater operational stress on enrolment and update infrastructure.

The biometric update data used in this analysis corresponds to aggregated records available up to 31st December 2025.

4.4 Summary of Datasets Used

Dataset Name	Description	Key Dimensions	Key Attributes Used
Aadhaar Enrolment Dataset	Aggregated Aadhaar enrolment records	Time, State, Age Group	Total Enrolments, Age-wise Enrolments
Aadhaar Demographic Update Dataset	Aggregated demographic update records	Time, State	Demographic Update Counts
Aadhaar Biometric Update Dataset	Aggregated biometric update records	Time, State	Biometric Update Counts

Table 4.1: Summary of Datasets Used in the Analysis

All three datasets are temporally aligned and aggregated at compatible geographic levels, enabling integrated analysis across enrolment, demographic updates, and biometric updates. Together, they form the foundation for the analytical methods and insights presented in the subsequent sections of this report.

Columns Used for Analysis:

Enrolment Dataset: date, state, age_0_5, age_5_17, age_18_plus

Demographic Update Dataset: date, state, demo_age_5_17, demo_age_18_plus

Biometric Update Dataset: date, state, bio_age_5_17, bio_age_18_plus

5. Methodology

This study follows a structured and reproducible methodology designed to support integrated analysis of Aadhaar enrolment, demographic updates, and biometric updates, with the objective of identifying operational workload patterns and stress points. The methodology is aligned with the proposed solution framework and emphasizes comparability, interpretability, and administrative relevance.

5.1 Data Ingestion and Initial Examination

Three UIDAI-provided aggregated datasets were used: Aadhaar Enrolment, Aadhaar Demographic Updates, and Aadhaar Biometric Updates. Each dataset was independently loaded and examined to understand its temporal coverage, geographic granularity, and available attributes. Initial inspection focused on verifying the presence of consistent state identifiers, time variables, and count-based metrics required for integrated analysis.

date	state	district	pincode	age_0_5	age_5_17	age_18_plus
02-03-2025	Meghalaya	East Khasi Hills	793121	11	61	37
09-03-2025	Karnataka	Bengaluru Urban	560043	14	33	39
09-03-2025	Uttar Pradesh	Kanpur Nagar	208001	29	82	12
09-03-2025	Uttar Pradesh	Aligarh	202133	62	29	15
09-03-2025	Karnataka	Bengaluru Urban	560016	14	16	21

Table 5.1: Sample of Aadhaar Enrolment Dataset

date	state	district	pincode	demo_age_5_17	demo_age_18_plus
01-03-2025	Uttar Pradesh	Gorakhpur	273213	49	529
01-03-2025	Andhra Pradesh	Chittoor	517132	22	375
01-03-2025	Gujarat	Rajkot	360006	65	765
01-03-2025	Andhra Pradesh	Srikakulam	532484	24	314
01-03-2025	Rajasthan	Udaipur	313801	45	785

Table 5.2: Sample of Aadhaar Demographic Update Dataset

date	state	district	pincode	bio_age_5_17	bio_age_18_plus
01-03-2025	Haryana	Mahendragarh	123029	280	577
01-03-2025	Bihar	Madhepura	852121	144	369
01-03-2025	Jammu and Kashmir	Punch	185101	643	1091
01-03-2025	Bihar	Bhojpur	802158	256	980
01-03-2025	Tamil Nadu	Madurai	625514	271	815

Table 5.3: Sample of Aadhaar Biometric Update Dataset

5.2 Data Cleaning and Standardization

To ensure consistency across datasets, state names and categorical fields were standardized prior to integration. Numerical fields related to enrolment and update counts were validated to avoid aggregation inconsistencies. Temporal attributes were normalized into a unified monthly format to support consistent time-series analysis across all datasets.

Since the datasets are aggregated in nature, cleaning efforts focused on structural and semantic consistency rather than record-level correction.

5.3 Temporal Alignment and Aggregation

All datasets were temporally aligned at a common monthly resolution. Aggregation operations were applied to compute state-wise monthly totals for enrolments, demographic updates, and biometric updates. Age-group-wise enrolment and update counts were also aggregated to support demographic analysis and lifecycle-based interpretation.

5.4 Dataset Integration

Following aggregation, the enrolment dataset was merged with demographic and biometric update datasets using state and monthly time period as common keys. This integration resulted in a unified analytical dataset containing enrolment volumes, update volumes, and age-group information for each state and time period.

5.5 Feature Engineering and Indicator Derivation

To enable meaningful comparison across states with varying enrolment scales, derived indicators were computed from aggregated counts. These include enrolment share, demographic update intensity, and biometric update pressure, calculated by relating update activity to corresponding enrolment volumes. Age-group contribution metrics were also derived to analyze demographic composition and its influence on update behavior.

These engineered features form the core analytical inputs for identifying relative workload pressure and hidden operational stress points.

Table 5.4 summarizes the derived analytical features used to normalize enrolment and update activity and to identify relative operational workload patterns across states.

Feature Name	Description
Enrolment Share	Proportion of total Aadhaar enrolments contributed by a state during the analysis period

Demographic Update Intensity	Ratio of total demographic updates to total enrolments for a given state
Biometric Update Pressure	Ratio of total biometric updates to total enrolments for a given state
Age-group Contribution	Proportion of enrolments contributed by each age group

Table 5.4: Summary of Derived Features Used in Analysis

These derived features enable meaningful comparison across states with significantly different enrolment volumes. By normalizing update activity relative to enrolment scale and demographic composition, the analysis reveals workload pressure patterns that are not visible through absolute counts alone.

5.6 Exploratory and Comparative Analysis

Univariate, bivariate, and trivariate analytical techniques were applied to examine enrolment and update patterns across time, geography, and age groups. Visual analytics were used to support interpretation of trends, concentration effects, temporal variability, and interaction patterns between enrolment and update activity.

5.7 Validation and Analytical Scope

Validation checks were applied to ensure logical consistency of derived metrics, including verification of non-negative values and reasonable ranges for ratio-based indicators. The analysis is exploratory and descriptive in nature, focusing on identifying trends and comparative patterns rather than causal inference. The methodology is designed to support decision-making and operational planning rather than predictive certainty.

6. Data Analysis and Visualisation

This section presents a comprehensive analysis of Aadhaar enrolment, demographic update, and biometric update datasets using univariate, bivariate, and trivariate analytical approaches. The objective is to understand individual data distributions, explore relationships between variables, and examine how enrolment and update patterns interact across time, geography, and age groups. Visualisations and summary tables are used to support interpretation and clearly communicate findings.

6.1 Univariate Analysis

Univariate analysis focuses on examining individual variables independently to understand their distribution, magnitude, and variability.

6.1.1 Aadhaar Enrolment Distribution Over Time

The first univariate analysis examines the temporal distribution of Aadhaar enrolments. Monthly enrolment totals were analyzed to identify overall trends and fluctuations over the study period.

Figure 6.1 illustrates the month-wise trend in Aadhaar enrolment, highlighting fluctuations in enrolment activity over the analysis period.
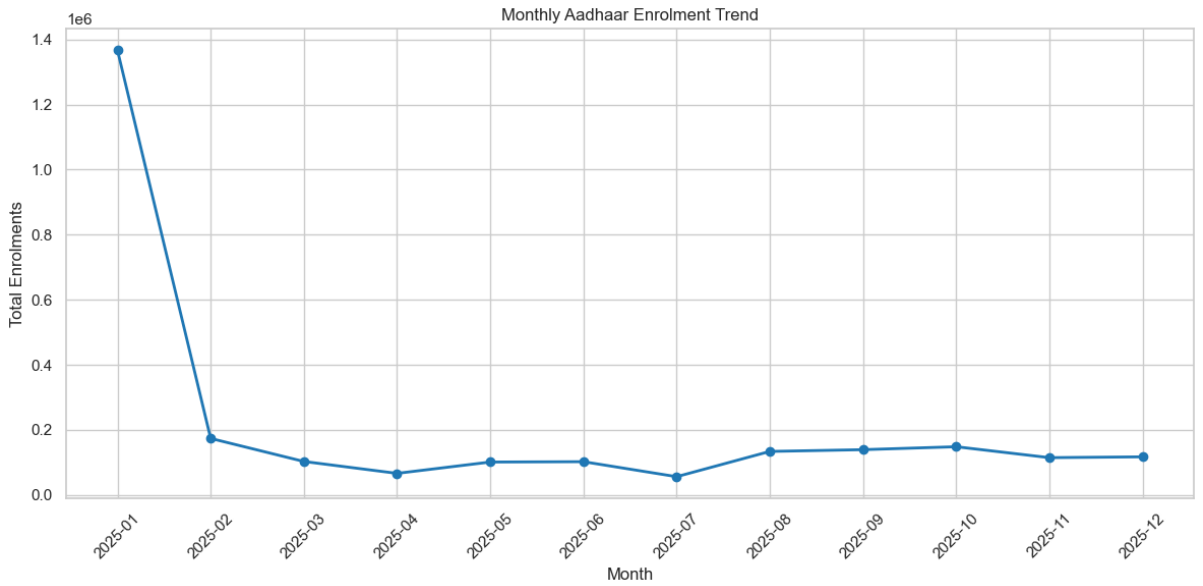
Figure 6.1: Monthly Aadhaar enrolment trend

The observed variability suggests that enrolment demand is not uniform across months, emphasizing the need for flexible operational planning.

6.1.2 State-wise Aadhaar Enrolment Distribution

To understand geographic variation, total enrolments were aggregated at the state level.

Table 6.1 presents the top ten states by total Aadhaar enrolments during the analysis period, highlighting the concentration of enrolment activity across regions.

Rank	State	Total Enrolments
1	Uttar Pradesh	1,002,631
2	Bihar	593,753

3	Madhya Pradesh	487,892
4	West Bengal	369,249
5	Maharashtra	363,446
6	Rajasthan	340,591
7	Gujarat	275,042
8	Assam	225,359
9	Karnataka	219,618
10	Tamil Nadu	215,710

Table 6.1: State-wise Aadhaar enrolment summary

The table indicates that a small number of states account for a substantial share of Aadhaar enrolments, confirming the presence of geographic concentration in enrolment-driven operational demand.

Figure 6.2 presents the distribution of Aadhaar enrolments across the top ten states, highlighting the extent of geographic concentration in enrolment activity.
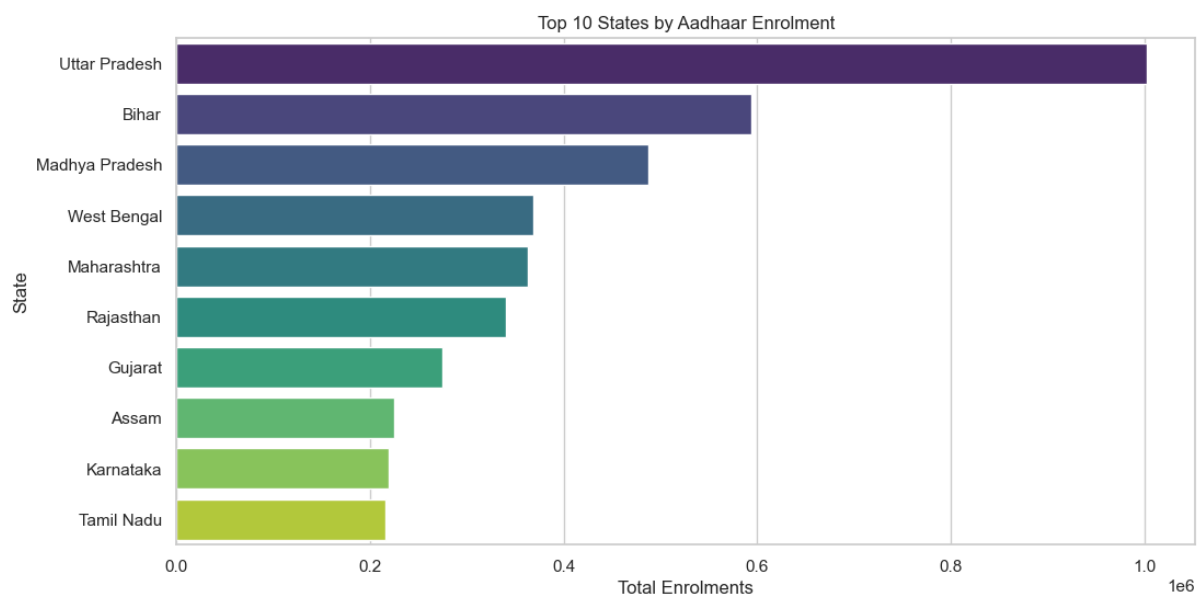
Figure 6.2: *Top 10 State-wise Aadhaar enrolment distribution*

The figure clearly shows that a small number of states dominate Aadhaar enrolment volumes, confirming the presence of strong geographic concentration.

6.1.3 Age-wise Enrolment Composition

Age-group-wise enrolment data was analyzed to understand demographic participation in Aadhaar enrolment.

Table 6.2 summarizes Aadhaar enrolment distribution across age groups, providing demographic context for subsequent update-related analysis.

Age Group	Total Enrolments	Percentage (%)
0–5 years	3,546,965	65.25
5–17 years	1,720,384	31.65
18+ years	168,353	3.10

Table 6.2: Age-wise Aadhaar enrolment distribution

The distribution indicates that Aadhaar enrolment during the analyzed period is overwhelmingly driven by early-age registration, with adult enrolments forming only a small fraction of total activity.

Figure 6.3 illustrates the contribution of different age groups to total Aadhaar enrolments, highlighting the demographic composition of enrolment activity.
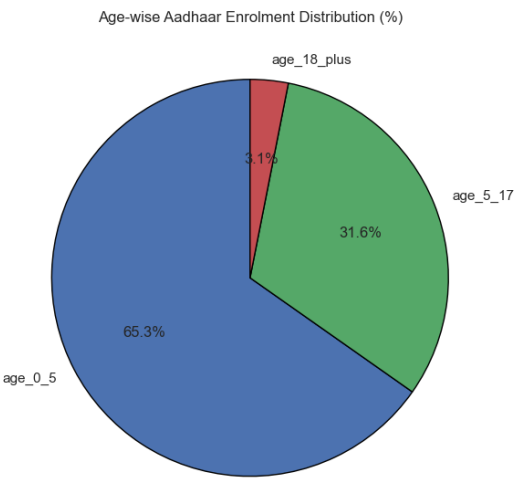

Figure 6.3: Age-wise contribution to Aadhaar enrolments

The figure shows that Aadhaar enrolment during the analyzed period is dominated by younger age groups, particularly children aged 0–5 years, while adult enrolments contribute only a small share.

6.2 Bivariate Analysis

Bivariate analysis explores relationships between two variables to identify dependencies, contrasts, and comparative patterns.

6.2.1 Enrolment vs Demographic Update Activity

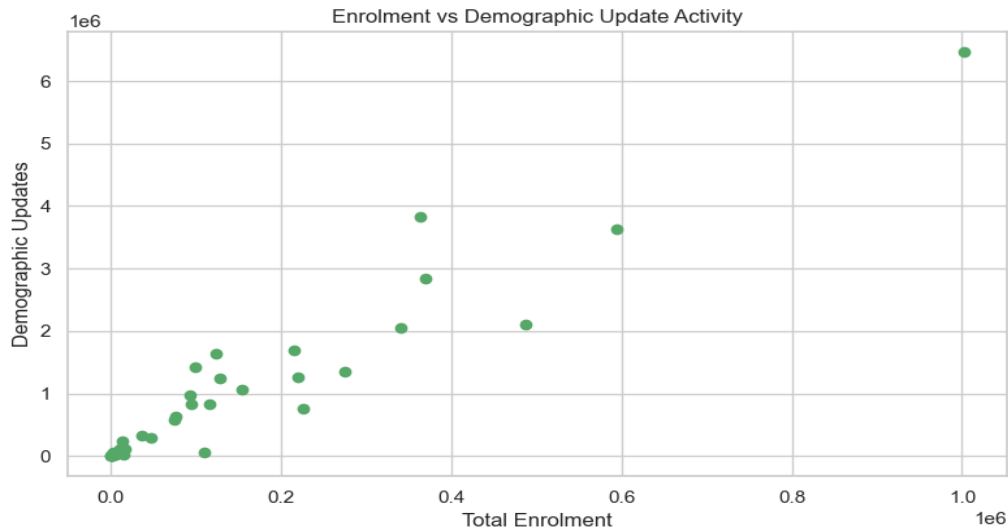This analysis examines how demographic update activity relates to Aadhaar enrolment across states.

Table 6.3 compares Aadhaar enrolment volumes with demographic update activity across major states to examine whether update workload scales proportionally with enrolment.

State	Total Enrolments	Enrolment Share	Cumulative Share	Demographic Updates (5–17)	Demographic Updates (18+)
Uttar Pradesh	1,002,631	0.188	0.188	603,453	5,857,058
Bihar	593,753	0.111	0.299	285,883	3,352,961
Madhya Pradesh	487,892	0.092	0.391	289,711	1,814,924
West Bengal	369,249	0.069	0.460	177,260	2,667,086
Maharashtra	363,446	0.068	0.528	204,293	3,620,598

Table 6.3: State-wise enrolments and demographic updates

The table shows that states with similar enrolment volumes can exhibit markedly different demographic update counts, indicating that demographic update workload is not directly proportional to enrolment activity.

Figure 6.4 compares Aadhaar enrolment volumes with demographic update activity across states to examine whether update workload scales proportionally with enrolment.

Figure 6.4: Enrolment vs demographic update activity

The figure demonstrates that states with comparable enrolment volumes can experience significantly different levels of demographic update activity, confirming that update workload does not scale linearly with enrolment.

6.2.2 Enrolment vs Biometric Update Activity

A similar bivariate comparison was conducted between enrolments and biometric updates.

Table 6.4 compares Aadhaar enrolment volumes with biometric update activity across major states to assess relative biometric update workload.

State	Total Enrolments	Enrolment Share	Cumulative Share	Biometric Updates (5–17)	Biometric Updates (18+)
Uttar Pradesh	1,002,631	0.188	0.188	6,076,420	3,290,663
Bihar	593,753	0.111	0.299	2,160,544	2,618,424
Madhya Pradesh	487,892	0.092	0.391	3,148,670	2,671,066
West Bengal	369,249	0.069	0.460	1,023,454	1,458,738
Maharashtra	363,446	0.068	0.528	3,437,083	5,583,627

Table 6.4: State-wise enrolments and biometric updates

The table highlights substantial variation in biometric update activity across states, even among those with comparable enrolment volumes, indicating that biometric updates represent a distinct and potentially intensive source of operational workload.

Figure 6.5 compares Aadhaar enrolment volumes with biometric update activity across states to assess the relative intensity of biometric update workload.
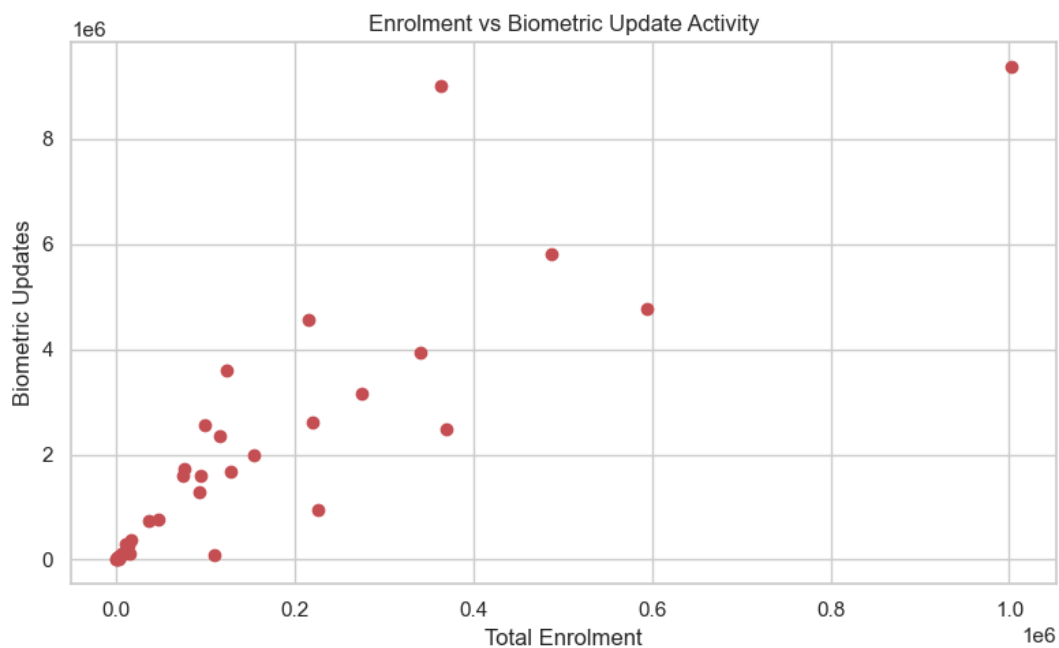

Figure 6.5: Enrolment vs biometric update activity

The figure highlights states where biometric update volumes are disproportionately high relative to enrolment levels, revealing potential hidden operational stress points.

6.2.3 Comparison of Demographic and Biometric Updates

To understand the relative contribution of different update types, demographic and biometric updates were compared directly.

Figure 6.6 compares the relative volumes of demographic and biometric updates across states to identify the dominant type of update contributing to operational workload.
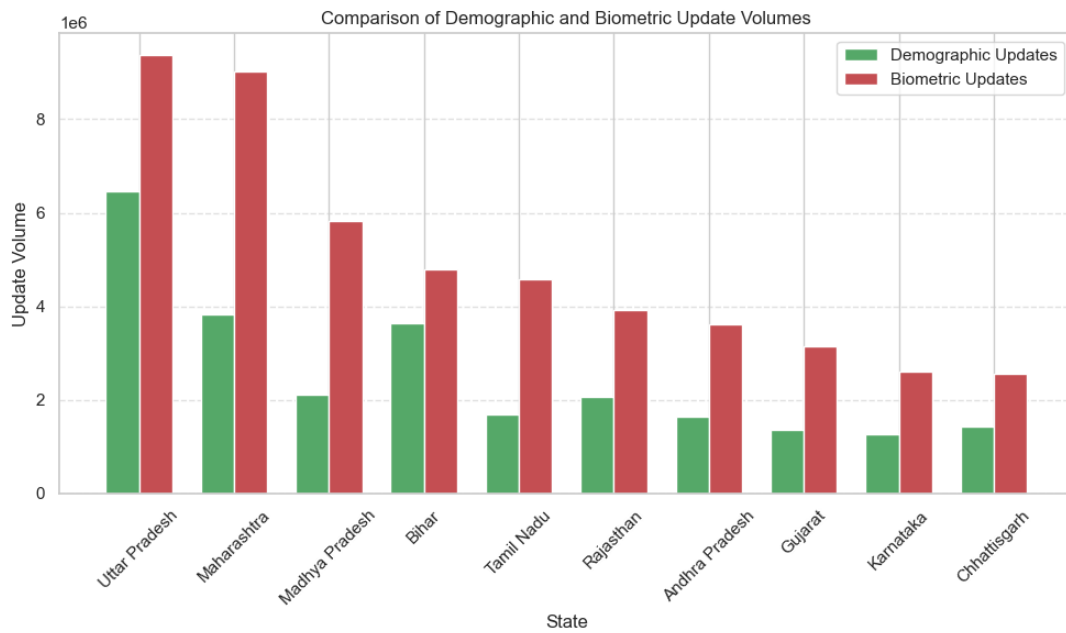
Figure 6.6: Comparison of demographic and biometric update volumes

The figure shows that the dominant type of update varies across states, suggesting that update-specific operational strategies may be more effective than a uniform management approach.
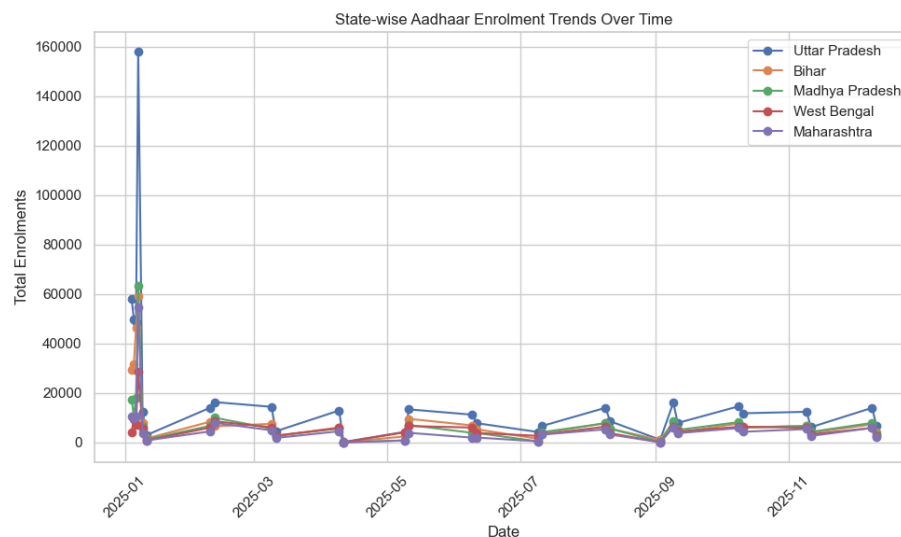
6.3 Trivariate Analysis

Trivariate analysis incorporates three dimensions simultaneously to capture more complex patterns involving time, geography, and age.

6.3.1 State-wise Enrolment Trends Over Time

This analysis examines how enrolment trends evolve over time across major states.

Figure 6.7 illustrates state-wise Aadhaar enrolment trends over time for major states, highlighting temporal variability and state-specific enrolment patterns.

The figure reveals that enrolment trends vary significantly across states, with some exhibiting relatively stable patterns and others showing pronounced temporal fluctuations.

6.3.2 Age Group and Update Activity Interaction

To explore how age composition influences update behavior, age-wise demographic and biometric updates were analyzed.

Table 6.5 presents the distribution of demographic and biometric updates across age groups to examine the lifecycle-driven nature of Aadhaar data maintenance.

Age Group	Demographic Updates	Biometric Updates
5–17 years	3,597,737	33,456,647
18+ years	32,999,820	34,804,412

Table 6.5: Age-wise demographic and biometric update distribution

Unlike enrolment activity, which is dominated by younger age groups, both demographic and biometric update workloads are primarily driven by the adult population, underscoring the lifecycle-dependent and continuous nature of Aadhaar updates.

Figure 6.8 compares demographic and biometric update activity across age groups to examine how update workload varies with age.
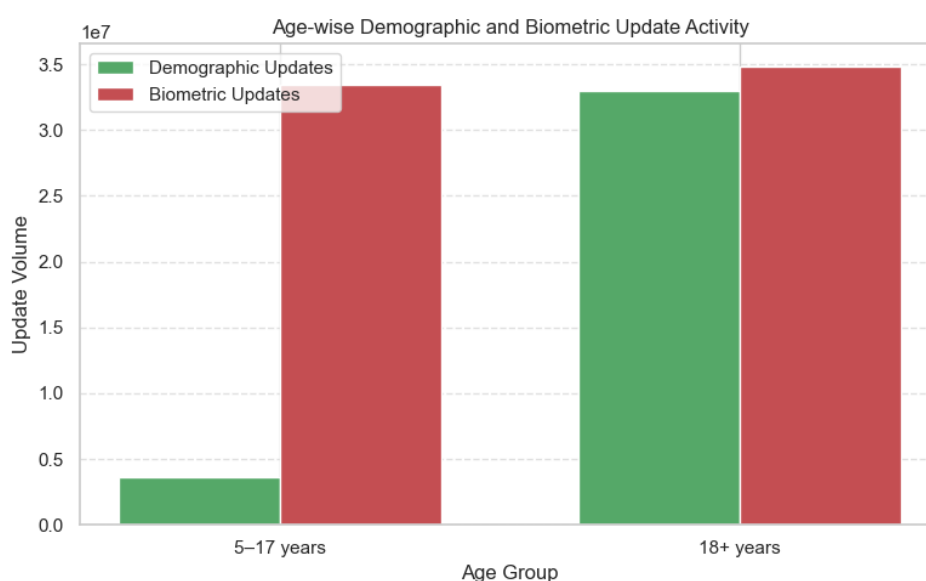
Figure 6.8: Age-wise update activity comparison

The figure clearly shows that both demographic and biometric update workloads are substantially higher in the adult population, reinforcing the lifecycle-driven nature of Aadhaar data maintenance.

6.3.3 Geographic–Temporal Interaction in Update Activity

Finally, update activity was examined across both geography and time to identify dynamic workload patterns.

Figure 6.9 illustrates state-wise trends in Aadhaar update activity over time, highlighting temporal fluctuations and periods of intensified update demand.
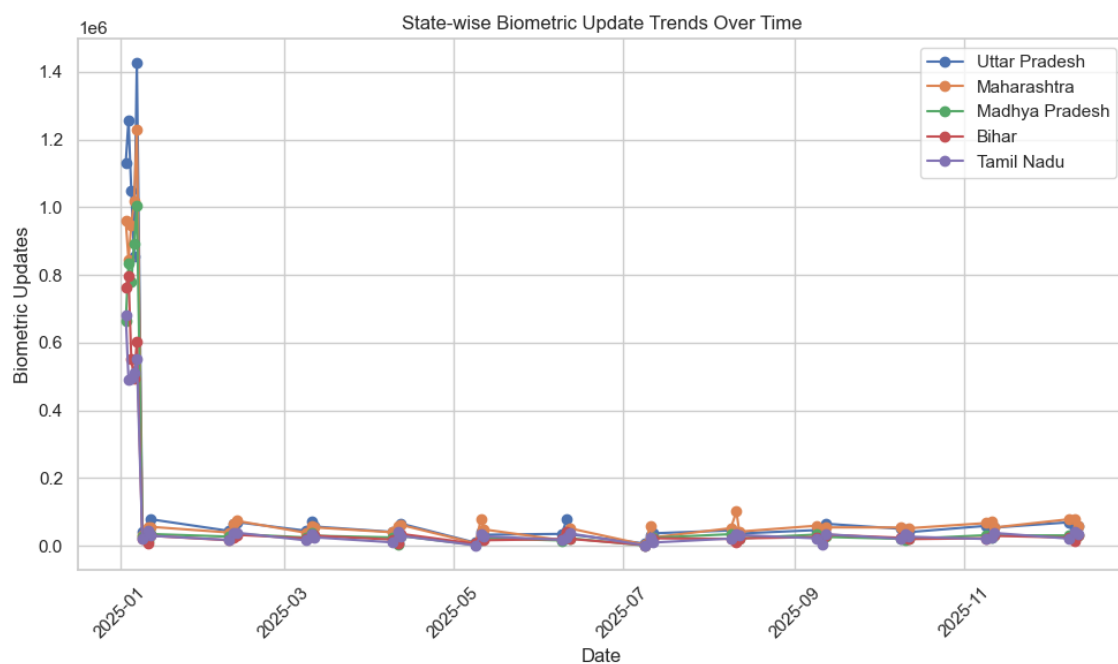

Figure 6.9: State-wise update trends over time

The figure shows noticeable temporal variability in update activity across states, with certain periods experiencing sharp surges, underscoring the importance of proactive workload monitoring and flexible operational planning.

6.4 Summary of Analytical Findings

Across univariate, bivariate, and trivariate analyses, the results consistently indicate that Aadhaar enrolment and update activity is unevenly distributed across states, age groups, and time periods. While enrolment volumes are concentrated in specific regions, demographic and biometric update workloads vary independently and are strongly influenced by age composition and temporal factors. The combination of tables and visualizations provides a multi-dimensional understanding of operational patterns within the Aadhaar ecosystem.

7. Key Insights and Observations

This section consolidates the most important insights derived from the univariate, bivariate, and trivariate analyses of Aadhaar enrolment, demographic updates, and biometric updates. The insights focus on workload distribution, demographic influence, and operational pressure patterns relevant to administrative planning and system optimization.

7.1 Enrolment Activity Is Highly Concentrated Across States

The analysis reveals a strong concentration of Aadhaar enrolment activity within a limited number of states. A small group of states contributes a disproportionately large share of total enrolments, as reflected by enrolment share and cumulative share metrics. This concentration pattern indicates that enrolment-related operational demand is not evenly distributed across the country, highlighting the need for region-specific planning rather than uniform allocation of resources.

7.2 Update Workload Is Not Proportional to Enrolment Volume

Bivariate analysis demonstrates that higher enrolment volumes do not necessarily correspond to higher demographic or biometric update activity. Several states with moderate enrolment levels exhibit comparatively high update volumes and update pressure ratios. This indicates that update workload is influenced by factors beyond enrolment scale alone, such as demographic structure, population mobility, and lifecycle-driven data changes.

7.3 Adult Population Drives the Majority of Update Activity

Age-wise analysis shows that demographic and biometric updates are heavily concentrated in the 18 years and above age group. While enrolment activity during the analyzed period is dominated by children and adolescents, update activity is primarily driven by the adult population. This reflects the ongoing need for demographic corrections and biometric revalidation over an individual's lifecycle, underscoring that update operations represent a sustained and significant workload independent of new enrolments.

7.4 Biometric Update Pressure Reveals Hidden Operational Stress Points

The biometric update pressure analysis highlights states that experience high update intensity relative to their enrolment volume. These hidden stress points may not be apparent from absolute enrolment counts alone. States exhibiting high biometric update ratios require targeted operational attention, as they may face infrastructure and staffing pressure despite moderate enrolment activity.

7.5 Temporal Variability Suggests Periodic Operational Surges

State-wise time-series analysis reveals noticeable temporal variability in both enrolment and update activity. Certain states exhibit sharp fluctuations over time, indicating periodic surges in operational

demand. These patterns suggest opportunities for proactive planning, such as temporary resource scaling or targeted awareness campaigns during high-activity periods.

7.6 Enrolment and Update Dynamics Require Integrated Monitoring

The combined insights from enrolment concentration, age-wise update behavior, and update pressure metrics indicate that enrolment and update processes should be monitored as interconnected components rather than isolated activities. An integrated monitoring framework can help identify regions where enrolment-driven and update-driven workloads overlap, enabling more efficient administrative response.

7.7 Implications for Administrative Planning

Overall, the analysis highlights that operational demand within the Aadhaar ecosystem is shaped by a combination of enrolment concentration, demographic structure, and temporal dynamics. Effective planning therefore requires moving beyond aggregate enrolment figures to include update intensity and demographic composition as key indicators for decision-making and system improvement.

8. Code and Implementation

Code Snippet 8.1: Data Loading and Initial Inspection

```python
DATA_PATH = "Data/"

enrolment_df = pd.concat([
    pd.read_csv(DATA_PATH + "api_data_aadhar_enrolment_0_500000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_enrolment_500000_1000000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_enrolment_1000000_1006029.csv")
], ignore_index=True)

biometric_df = pd.concat([
    pd.read_csv(DATA_PATH + "api_data_aadhar_biometric_0_500000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_biometric_500000_1000000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_biometric_1000000_1500000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_biometric_1500000_1861108.csv")
], ignore_index=True)

demographic_df = pd.concat([
    pd.read_csv(DATA_PATH + "api_data_aadhar_demographic_0_500000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_demographic_500000_1000000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_demographic_1000000_1500000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_demographic_1500000_2000000.csv"),
    pd.read_csv(DATA_PATH + "api_data_aadhar_demographic_2000000_2071700.csv")
], ignore_index=True)
```

Figure 8.1: Python code used to load UIDAI aggregated datasets and perform initial inspection.

Code Snippet 8.2: State Name Standardization

```python
def clean_state(series):
    series = (
        series
        .astype(str)
        .str.strip()
        .str.replace("&", "and", regex=False)
        .str.title()
    )

    corrections = {
        # Merged UTs
        "Dadra And Nagar Haveli": "Dadra And Nagar Haveli And Daman And Diu",
        "Daman And Diu": "Dadra And Nagar Haveli And Daman And Diu",
        "The Dadra And Nagar Haveli And Daman And Diu": "Dadra And Nagar Haveli And Daman And Diu",

        # West Bengal variants
        "West  Bengal": "West Bengal",
        "West Bangal": "West Bengal",
        "Westbengal": "West Bengal"
    }

    series = series.replace(corrections)

    # Drop clearly invalid numeric-like states
    series = series.where(~series.str.isnumeric())

    return series
```

Figure 8.2: Function used to normalize state names to ensure consistent aggregation.

Code Snippet 8.3: Feature Engineering and Normalization

```python
df['demo_update_intensity'] = (
    df['demo_age_5_17'] + df['demo_age_18_plus']
) / df['total_enrolments']

df['bio_update_pressure'] = (
    df['bio_age_5_17'] + df['bio_age_18_plus']
) / df['total_enrolments']
```

Figure 8.3: Computation of demographic update intensity and biometric update pressure.

Code Snippet 8.4: Monthly Aggregation of Enrolment, Demographic, and Biometric Update Data

```python
monthly_enrolment = (
    enrolment_df
    .groupby("year_month", as_index=False)["total_enrolment"]
    .sum()
    .sort_values("year_month")
)

monthly_enrolment.head()
```

```
monthly_biometric = (
    biometric_df
    .groupby("year_month", as_index=False)[["bio_age_5_17", "bio_age_17_"]]
    .sum()
    .sort_values("year_month")
)

monthly_biometric.head()
```

```
monthly_demographic = (
    demographic_df
    .groupby("year_month", as_index=False)[["demo_age_5_17", "demo_age_17_"]]
    .sum()
    .sort_values("year_month")
)

monthly_demographic.head()
```

This code aggregates Aadhaar enrolment, demographic update, and biometric update records at a monthly level using year–month grouping to enable consistent temporal analysis.

Code Snippet 8.5: Identification and Visualization of States with High Biometric Update Pressure

```
top_pressure = state_enrol_demo_bio.sort_values(
    "bio_update_ratio", ascending=False
).head(10)

sns.barplot(
    x="bio_update_ratio",
    y="state",
    data=top_pressure,
    palette="rocket"
)

plt.title("Top 10 States by Biometric Update Pressure")
plt.xlabel("Biometric Update Ratio")
plt.ylabel("State")
plt.show()
```

This code identifies the top ten states with the highest biometric update pressure by sorting states based on the biometric update ratio and visualizes the results using a horizontal bar chart.

Code Snippet 8.6: State-wise Comparison of Enrolment and Demographic Update Activity

```
demo_state = (
    demographic_df
    .groupby('state', as_index=False)[
        ['demo_age_5_17', 'demo_age_17_']
    ]
    .sum()
)

demo_state['total_demo_updates'] = (
    demo_state['demo_age_5_17'] +
    demo_state['demo_age_17_']
)

enrol_state = (
    enrolment_df
    .groupby('state', as_index=False)['total_enrolment']
    .sum()
)

demo_compare = enrol_state.merge(demo_state, on='state')
```

```
plt.figure(figsize=(8, 5))
plt.scatter(
    demo_compare['total_enrolment'],
    demo_compare['total_demo_updates'],
    color= '#55A868'
)

plt.xlabel('Total Enrolment')
plt.ylabel('Demographic Updates')
plt.title('Enrolment vs Demographic Update Activity')
plt.grid(True)

plt.tight_layout()
plt.show()
```

This code aggregates enrolment and demographic update data at the state level, merges the datasets, and visualizes the relationship between total enrolment and demographic update activity using a scatter plot.