Unlocking Societal Trends in Aadhaar Enrolment and Updates

Data-Driven Analysis of Aadhaar Enrolment Patterns (2025)

UIDAI Data Hackathon 2026

Organised by

UIDAI — NIC — MeitY

Team Id

(UIDAI_9403)

Team Members

Sahil Shukre
Sharvil Kulkarni
Aayush Singh

Tools Used

Python (Numpy, Pandas, Matplotlib, Seaborn)
Jupyter Notebook

1 Problem Statement and Objectives

1.1 Problem Statement

The Aadhaar ecosystem generates large-scale enrolment data that can provide valuable insights into demographic coverage, regional disparities, and enrolment demand. Analyzing this data can help in understanding how Aadhaar services are accessed across different regions and population groups.

This project aims to analyze anonymised Aadhaar enrolment data to identify meaningful patterns, trends, and disparities. The insights derived from this analysis can support informed decision-making and contribute to improvements in enrolment processes and system efficiency.

1.2 Objectives

The key objectives of this project are as follows:

- To analyze district-wise and state-wise Aadhaar enrolment patterns

- To study age-wise enrolment distribution across regions

- To identify demographic and geographic disparities in enrolment

- To propose actionable and data-driven recommendations based on analysis

2 Dataset Description

2.1 Dataset Used

The dataset used for this project is the **Aadhaar Enrolment Dataset** provided by the Unique Identification Authority of India (UIDAI).

- **Year Covered:** 2025

- **Data Type:** Aggregated and anonymised

The dataset contains enrolment information collected across different states, districts, and PIN code regions in India. Since the data is anonymised and aggregated, it ensures privacy while still allowing meaningful analysis of enrolment trends and demographic patterns.

2.2 Key Columns

The table below describes the important columns used in the analysis:

Column	Description
date	Date of Aadhaar enrolment
state	Name of the state
district	Name of the district
pincode	PIN code of the enrolment location
age_0_5	Number of enrolments for children aged 0–5 years
age_5_17	Number of enrolments for individuals aged 5–17 years
age_18_greater	Number of enrolments for individuals aged 18 years and above

3 Methodology

3.1 Data Preprocessing

Before performing analysis, the raw data was cleaned and prepared to ensure accuracy and consistency. The following preprocessing steps were carried out:

- Merged multiple CSV files into a single dataset

- Parsed and standardized the date format for uniformity

- Removed data inconsistencies and validated numeric fields

- Created a derived metric named *Total Enrolments*

These preprocessing steps helped in improving data quality and reliability for further analysis.

3.2 Analytical Approach

A structured analytical approach was adopted to explore the dataset and extract meaningful insights:

- Univariate analysis to study individual variable distributions

- Bivariate analysis to compare enrolment patterns across regions and age groups

- Trivariate analysis combining age group, geographic region, and enrolment intensity

4 Data Analysis and Insights

4.1 Section A: Univariate Analysis

Univariate analysis was performed to understand the overall trends and distribution of Aadhaar enrolment activity across different regions. This analysis focuses on examining one variable at a time to identify patterns and variations in the data.

4.1.1 Q1. How does Aadhaar enrolment intensity vary across districts?

To analyze enrolment intensity, district-wise total Aadhaar enrolments were examined. The analysis highlights significant variation in enrolment counts across districts, indicating differences in population density, access to enrolment centers, and regional demand.

Some districts show consistently high enrolment intensity, while others exhibit relatively lower enrolment activity. These variations suggest the presence of geographic and infrastructural disparities in Aadhaar enrolment coverage.
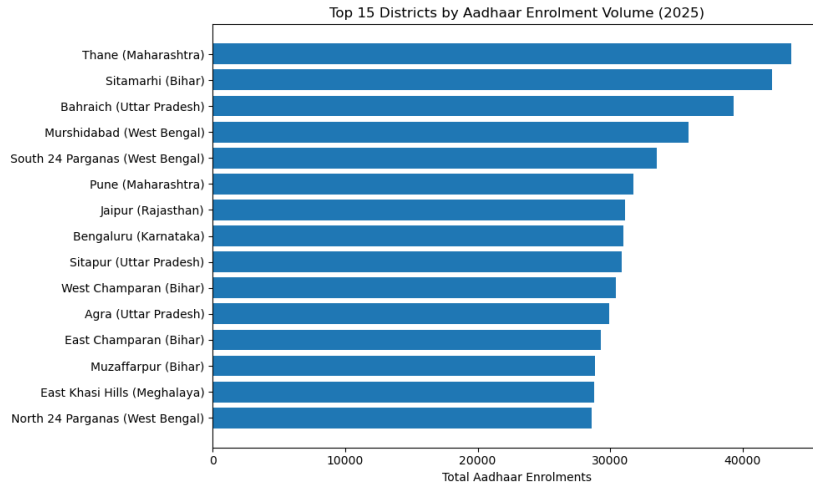

Figure 1: District-wise Aadhaar Enrolment Intensity

The visualization clearly demonstrates uneven enrolment distribution across districts, which can help authorities identify high-demand regions as well as areas requiring additional enrolment support.

Insights for District-wise Aadhaar Enrolment Intensity

It is evident that Aadhaar enrolment is not evenly distributed across districts. A significant concentration of enrolment activity is observed in major metropolitan areas such as Thane, Pune, and Bengaluru, as well as in densely populated rural regions across states like Bihar, Uttar Pradesh, and West Bengal.

This pattern indicates a high enrolment demand in both urban centers and high-population rural districts. The findings highlight the importance of district-specific resource planning and the need for targeted allocation of enrolment infrastructure to manage demand efficiently and ensure balanced coverage.

4.1.2 Q2. Which age group contributes the most to Aadhaar enrolments?

This analysis focuses on identifying the demographic group that contributes the most to Aadhaar enrolment activity. Age-wise enrolment data was examined to understand which population segment is currently driving new registrations.

The distribution of enrolments across age groups shows noticeable variation, reflecting the stage of Aadhaar coverage among different sections of the population.
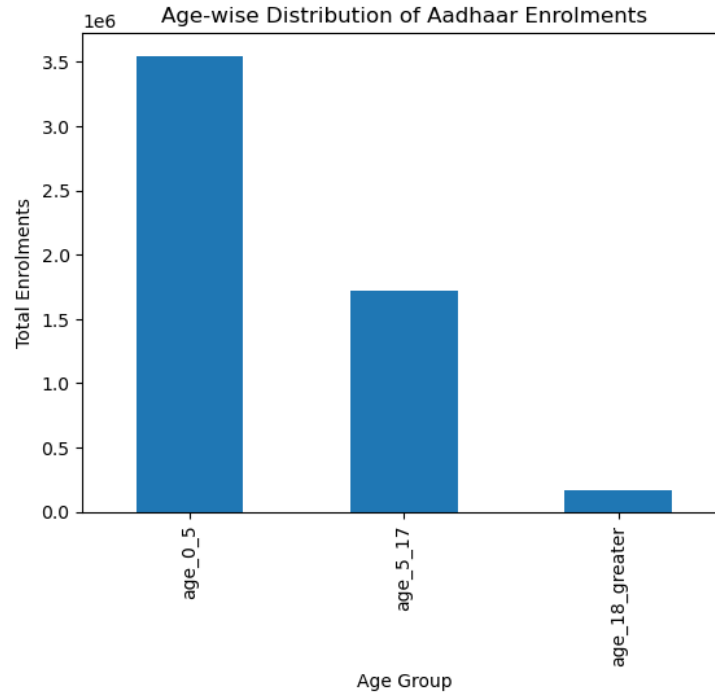

Figure 2: Age-wise Aadhaar Enrolment Distribution

Insights for Age-wise Aadhaar Distribution

From the chart, it is observed that:

- The 0–5 age group contributes the highest number of Aadhaar enrolments

- The 5–17 age group also shows substantial enrolment activity

- The 18+ age group records comparatively lower enrolments in this dataset

This suggests that enrolments among adults (18 years and above) are relatively lower, indicating that Aadhaar coverage in the adult population is largely saturated. Current enrolment activity is primarily driven by early-age registrations, highlighting the focus on enrolling children and adolescents.

4.1.3   Q3. Which states record the highest Aadhaar enrollments?

This analysis examines state-wise Aadhaar enrolment totals to identify regions with the highest enrolment activity. Understanding state-level enrolment patterns helps in assessing the impact of population size and enrolment infrastructure effectiveness.

The comparison highlights significant differences in enrolment volumes across states, reflecting demographic and geographic variations.

Insights for Highest Aadhaar Enrollments

From the chart, the following observations can be made:

- Uttar Pradesh records the highest Aadhaar enrolments by a significant margin

- Bihar and Madhya Pradesh also show high enrolment activity

- Large and populous states dominate the top enrolment rankings

- Southern and western states appear with comparatively lower enrolment volumes

This suggests that states with larger populations naturally dominate enrolment volumes, indicating a strong relationship between population size and Aadhaar enrolment activity. The findings highlight the importance of proportionate infrastructure and resource allocation in high-population states to ensure efficient and accessible Aadhaar services.
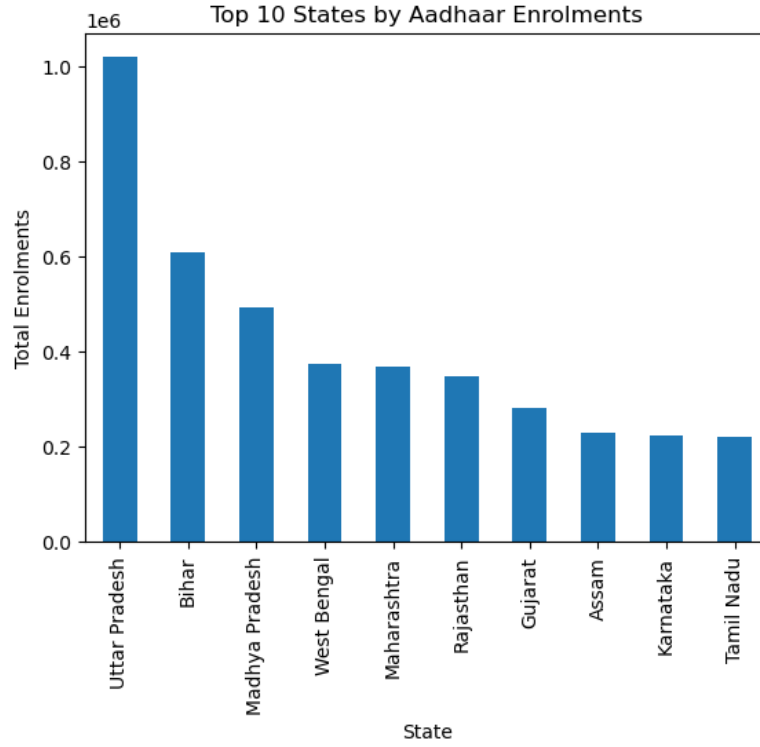
Figure 3: State-wise Aadhaar Enrolment Distribution

4.2 Section B: Bivariate Analysis

Bivariate analysis was conducted to examine the relationship between two variables simultaneously, allowing deeper insights into Aadhaar enrolment patterns across demographic and geographic dimensions.

4.2.1 Q1. How does age-wise Aadhaar enrolment vary across states?

This analysis compares Aadhaar enrolment patterns of different age groups across states to identify demographic disparities and regional differences. By analyzing age-wise enrolment at the state level, it becomes possible to understand which population segments are driving enrolment activity in different regions.

Insights for age-wise Aadhaar enrollment across states

From the chart, the following conclusions can be drawn:
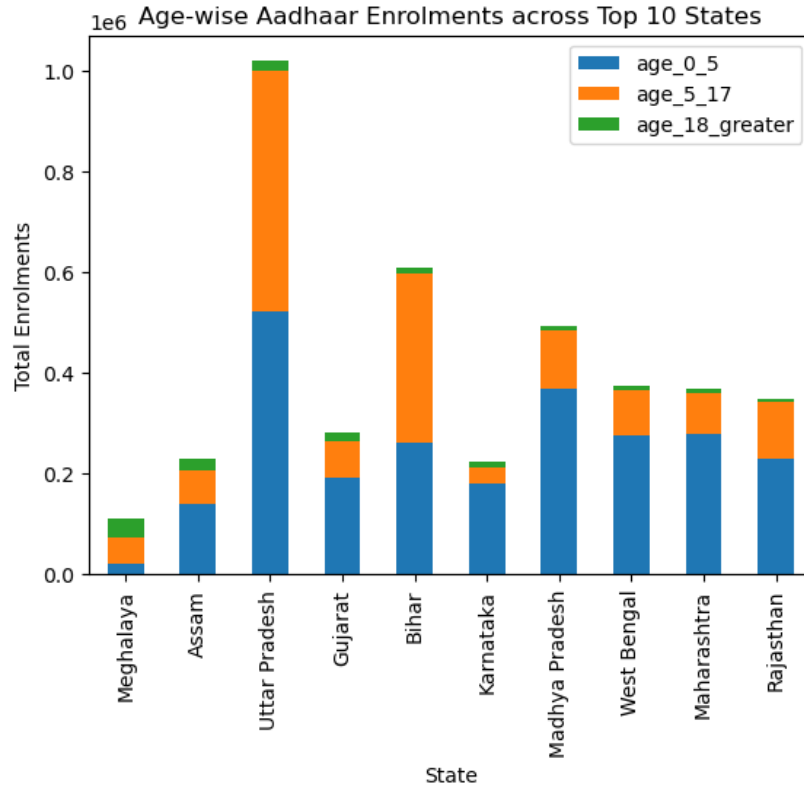
7

Figure 4: State-wise Age Group Distribution of Aadhaar Enrolments

- **Youth Dominance:** Enrolments are heavily concentrated in the 0–5 and 5–17 age groups across all states

- **Adult Saturation:** The 18+ age group consistently represents the smallest segment, indicating that adult Aadhaar registration is largely complete

- **State Leader:** Uttar Pradesh records the highest overall enrolment volume, followed by Bihar and Madhya Pradesh. The high volumes in these states reflect larger child populations and active enrolment drives for minors

4.2.2 Q2. What is the percentage contribution of each age group in Aadhaar enrolment?

This analysis examines the percentage contribution of different age groups to total Aadhaar enrolments. Studying proportional contributions provides a clearer understanding of which demographic segments dominate enrolment activity.
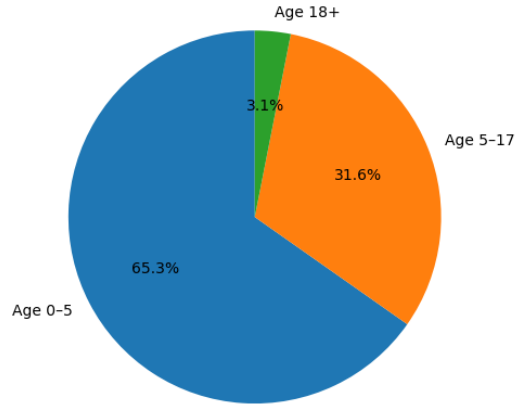
Figure 5: Percentage Contribution of Age Groups in Aadhaar Enrolment

Insights for percentage contribution of age groups in 2025

From the chart, the following observations are made:

- **Dominant Group:** Children aged 0–5 constitute the largest share of Aadhaar enrolments at 65.3%, indicating a strong focus on registration at birth

- **School-Age Contribution:** The 5–17 age group accounts for 31.6% of total enrolments, reflecting continued registration during schooling years

- **Minimal Adult Enrolment:** Adults aged 18 years and above contribute only 3.1% of enrolments, suggesting that adult Aadhaar coverage has reached near-total saturation

4.2.3 Q3. Which states show relatively higher child enrolment compared to adult enrolment?

This analysis focuses on identifying states and Union Territories where early-age Aadhaar enrolment is comparatively stronger than adult enrolment. The comparison helps in evaluating the effectiveness of child-focused enrolment initiatives across regions.
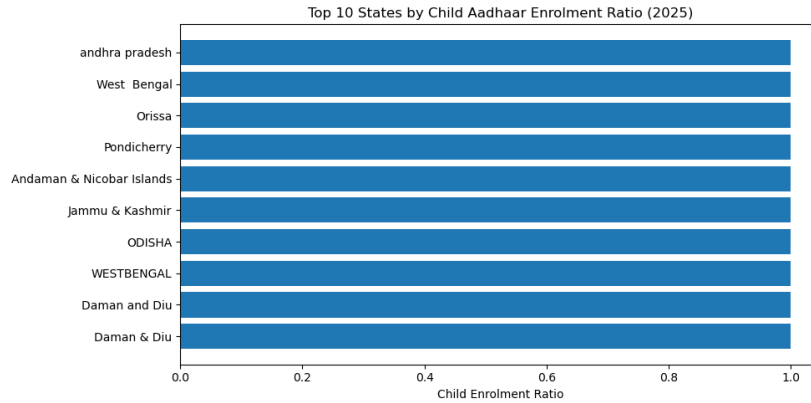
Figure 6: Child to Adult Aadhaar Enrolment Ratio Across States

Insights

The following insights are derived from the chart:

- **Perfect Saturation:** Several states and Union Territories, including Andhra Pradesh, West Bengal, and Odisha, have achieved a child enrolment ratio close to 1.0, indicating near-universal Aadhaar coverage among children

- **Effective Child Outreach:** The presence of both large states and smaller Union Territories such as Puducherry and Daman & Diu suggests that child-focused registration drives are effective across diverse geographic scales

- **Near-Zero Adult Gap:** In these top-performing regions, child enrolment no longer lags behind adult enrolment, reflecting a fully matured and balanced Aadhaar registration ecosystem across age groups

4.2.4 Q4. Is there a seasonal pattern in Aadhaar enrolments?

This analysis explores monthly Aadhaar enrolment trends to determine whether enrolment activity follows a seasonal pattern. Understanding temporal variations helps in planning registration drives and optimizing resource deployment throughout the year.

Insights

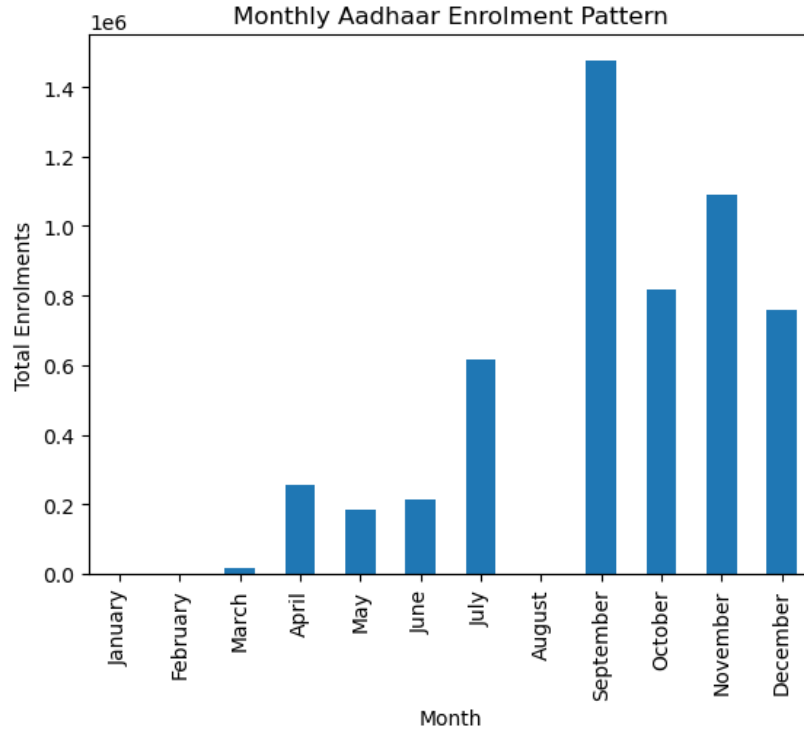The chart reveals a clear seasonal pattern in Aadhaar enrolments:

Figure 7: Monthly Aadhaar Enrolment Trends

- **Year-End Surge:** A significant concentration of enrolments occurs between September and December, indicating heightened registration activity during the latter part of the year

- **Educational Alignment:** The increase beginning around July to September likely corresponds with school admission cycles and requirements for updated student documentation

- **First Quarter Lull:** The minimal enrolment activity observed during January to March suggests a seasonal slowdown or a specific administrative or reporting cycle

- **September Peak:** September emerges as the most critical month, recording the highest total Aadhaar enrolments within the observed period

4.3   Section C: Trivariate Analysis

Trivariate analysis examines the interaction between three variables simultaneously to uncover deeper patterns and insights. In this context, age-wise Aadhaar

enrolments are analyzed across states to identify regions with significant demographic imbalance.

4.3.1 Q1. Which states exhibit significant imbalance in age-wise Aadhaar enrolments?

The analysis explores age-wise enrolment distributions across states to detect imbalances and outliers, highlighting states that may require targeted attention for certain age groups.

Insights from Figure 8

Key observations from the chart include:

- **Demographic Imbalance:** Most states are in a "maintenance" phase, with 80–90% of enrolments among children (0–17 years), as adult coverage is already saturated

- **Data Anomaly:** An outlier entry labeled "100000" shows almost exclusively adult enrolments, which contradicts the national trend of child-led enrolment growth

- **Regional Variation:** Meghalaya demonstrates ongoing adult enrolment activity, suggesting the state is still progressing toward universal coverage for mature populations

- **Infant Focus:** In high-performing states such as Goa and Karnataka, the majority of new registrations are among newborns (0–5 age group), indicating a shift in focus to early-age enrolment

4.3.2 Q2. Are there districts with unusually high Aadhaar enrolment activity?

This analysis identifies districts that exhibit unusually high enrolment activity, highlighting urban and rural hotspots as well as regional leaders in Aadhaar registration.

Insights from Table 1

From the analysis, the following conclusions can be drawn:

- **Urban-Rural Hotspots:** High enrolment activity is observed both in metropolitan centers such as Thane and Pune, and in densely populated rural districts in Bihar and West Bengal, reflecting broad-based registration drives
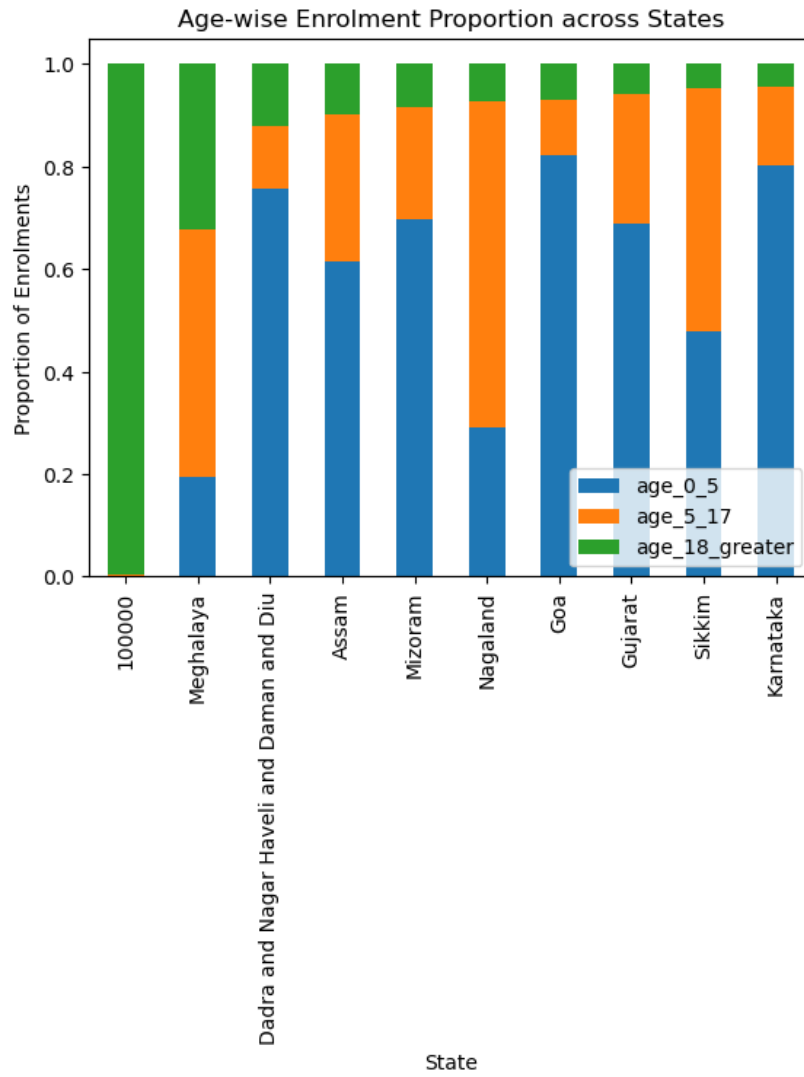
Figure 8: State-wise Age Group Imbalance in Aadhaar Enrolments

District	Total Enrolments
Thane	43,688
Sitamarhi	42,232
Bahraich	39,338
Murshidabad	35,911
South 24 Parganas	33,540
Pune	31,763
Jaipur	31,146
Bengaluru	30,980
Sitapur	30,854
Hyderabad	30,830

Table 1: Top 10 Districts by Aadhaar Enrolment Activity

- **Regional Leaders:** Thane and Sitamarhi stand out as the primary drivers of enrolment volume, significantly surpassing other major districts

- **Sustained Demand:** All top 10 districts maintain a high baseline of 30,000+ enrolments, indicating robust administrative capacity and consistent demand for Aadhaar registration

4.3.3 Q3. How does child enrolment vary across districts within high-enrolment states?

This analysis investigates intra-state variations in child (0–5 years) Aadhaar enrolments, focusing on districts within states that have high overall enrolment volumes. The aim is to identify local disparities and areas needing targeted enrolment efforts.

State	District	Age 0–5 Enrolments
100000	100000	0
Lakshadweep	Lakshadweep	192
Maharashtra	Ahilyanagar	12
Maharashtra	Ahmadnagar	9,369
Maharashtra	Ahmed Nagar	277

Table 2: Child (0–5 Years) Enrolments Across Selected Districts

Insights from Table 2

Key observations include:

- **Intra-State Imbalance:** Even high-performing states show internal "blind spots" where child enrolment is much lower in certain districts compared to state averages

- **Localized Enrollment Drives:** Districts with high enrolments, such as Ahmadnagar, likely benefit from permanent enrolment centers or targeted camps in schools and anganwadis

- **Saturation Gaps:** Low enrolments in districts like Ahilyanagar may reflect smaller target populations or limited accessibility to enrolment services

- **Administrative Focus:** Disparities highlight areas where state governments may need to redeploy staff to ensure universal coverage for children entering school age

5 Solution Framework & Recommendations

Based on the analyses conducted in Sections A–C, the following solution framework and recommendations are proposed to optimize Aadhaar enrolment coverage, reduce disparities, and improve operational efficiency:

5.1 Key Recommendations

1. **Targeted District-Level Resource Allocation:**

   - Focus resources on districts with high enrolment demand (e.g., Thane, Sitamarhi)
   - Deploy additional staff and enrolment centers in high-population rural districts
   - Address intra-state blind spots where enrolment is low despite high state averages

2. **Strengthening Child Enrolment Programs:**

   - Expand early-age enrolment drives in districts with lower 0–5 enrolment
   - Utilize schools, anganwadis, and health centers to capture newborn registrations
   - Maintain successful strategies in high-performing regions (e.g., Goa, Karnataka)

3. **Age-Specific Outreach Strategies:**

   - Focus adult enrolment efforts only where saturation has not yet been achieved
   - Tailor campaigns for school-age children to align with educational enrollment cycles
   - Use seasonal insights to plan enrolment drives, especially from July to December

4. **Data-Driven Monitoring Indicators:**

   - Establish dashboards tracking district, state, and age-group enrolments
   - Identify outliers and anomalies (e.g., unusually high adult enrolments in specific districts)
   - Monitor seasonal trends to optimize resource deployment

5. **Regional Infrastructure Optimization:**

   - Prioritize setting up enrolment centers in high-demand urban and rural districts
   - Ensure adequate staffing and facility availability during peak months
   - Integrate mobile enrolment units for remote or under-served districts

6. **Equity and Saturation Focus:**

   - Address regional disparities in high-population states such as UP, Bihar, and MP
   - Ensure uniform coverage for children and adolescents in all districts
   - Plan interventions where adult enrolment gaps persist, like Meghalaya

7. **Policy and Administrative Actions:**

   - Use insights to guide periodic audits and resource reallocation
   - Align enrolment drives with school and health program cycles for efficiency
   - Incorporate learnings into long-term Aadhaar operational planning

Conclusion

The analysis demonstrates that Aadhaar enrolment patterns are influenced by age, region, and seasonality. Targeted, data-driven strategies focusing on district-specific allocation, child enrolment programs, and seasonal planning can help achieve universal coverage, optimize administrative efficiency, and reduce disparities across the country.

6 Limitations & Assumptions

6.1 Limitations

While this study provides valuable insights into Aadhaar enrolment patterns, several limitations should be noted:

- **Single-Year Dataset:** The analysis uses data from the year 2025 only, which limits the ability to detect multi-year trends or long-term changes in enrolment patterns

- **Aggregated Data:** Data is aggregated at district, state, and age-group levels, preventing individual-level inference and detailed household or demographic analysis

- **Population Normalization:** Absence of population denominators restricts the ability to compute per-capita enrolment rates and compare districts proportionally

- **Data Anomalies:** Outliers (e.g., district entry "100000") may skew certain trends and require careful interpretation

- **Seasonal/Administrative Reporting Bias:** Monthly variations might reflect reporting cycles rather than true enrolment activity

6.2 Assumptions

The analysis is conducted under the following assumptions:

- **Enrolment Volume as Demand Proxy:** The total enrolment counts are assumed to reflect actual enrolment demand in each region

- **Aggregated Data Represents Regional Activity:** District- and state-level aggregates are considered representative of local enrolment activity patterns

- **Consistency of Data Reporting:** Data is assumed to be consistently collected and accurately reported across all districts and states

- **Age Groups Capture Key Demographic Trends:** The age bins (0–5, 5–17, 18+) are assumed sufficient to capture the major demographic trends in enrolment

7 Future Scope

Building on the insights and recommendations from this study, several avenues can be explored to enhance Aadhaar enrolment analysis and operational efficiency in the future:

- **Multi-Year Analysis:** Extend the study to include multiple years of enrolment data to identify long-term trends, growth patterns, and policy impacts over time

- **Integration with Demographic and Biometric Update Datasets:** Combine enrolment data with demographic and biometric update information to gain a more comprehensive understanding of population coverage and update needs

- **Real-Time Dashboards for UIDAI Administrators:** Develop interactive dashboards to monitor enrolment activity at district, state, and age-group levels, enabling timely interventions and resource allocation

- **Population-Adjusted Indicators:** Introduce per-capita and population-adjusted enrolment metrics to account for differences in district and state populations, allowing more equitable resource planning

- **Predictive Modelling:** Use historical trends to forecast future enrolment demand and optimize enrolment drive scheduling and staffing

- **Enhanced Data Quality Measures:** Implement automated anomaly detection to identify and correct outliers or inconsistencies in real-time

These future directions aim to strengthen Aadhaar enrolment monitoring, improve policy planning, and ensure equitable access across all regions and demographics.

8 Conclusion

This analysis demonstrates how anonymised Aadhaar enrolment data can be leveraged to uncover meaningful demographic and geographic patterns. Key findings include:

- Significant regional disparities in enrolment, with certain urban and high-population rural districts driving most activity

- Strong early-age enrolment focus, indicating successful child-focused registration initiatives and near-saturation among adults

- Seasonal and demographic patterns that can inform resource allocation and operational planning

The proposed solution framework offers practical, data-driven recommendations, including targeted district-level resource allocation, age-specific outreach strategies, and monitoring indicators. These measures can support UIDAI in improving enrolment coverage, operational efficiency, and policy planning, while addressing demographic and regional gaps to achieve near-universal Aadhaar coverage.

Code Files

January 15, 2026

```
[52]: # Unlocking Societal Trends in Aadhaar Enrolment

      ## Hackathon: UIDAI Data Hackathon 2026
      ## Dataset: Aadhaar Enrolment Dataset
      ## Objective: To identify demographic, regional, and temporal patterns in␣
       ↪Aadhaar enrolment
```

```
[53]: ### STAGE 0 - ENVIRONMENT SETUP

      # Importing libraries

      import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
      import seaborn as sns

      plt.style.use("default")
```

```
[54]: # Loading and merging Dataset

      paths = [
          "api_data_aadhar_enrolment_0_500000.csv",
          "api_data_aadhar_enrolment_500000_1000000.csv",
          "api_data_aadhar_enrolment_1000000_1006029.csv",
      ]

      dfs = [pd.read_csv(p) for p in paths]
      df = pd.concat(dfs, ignore_index=True)

      df.head()
```

```
[54]:         date           state          district  pincode  age_0_5  age_5_17  \
      0  02-03-2025       Meghalaya  East Khasi Hills   793121       11        61
      1  09-03-2025       Karnataka   Bengaluru Urban   560043       14        33
      2  09-03-2025   Uttar Pradesh      Kanpur Nagar   208001       29        82
      3  09-03-2025   Uttar Pradesh           Aligarh   202133       62        29
      4  09-03-2025       Karnataka   Bengaluru Urban   560016       14        16
```

```
     age_18_greater
0                 37
1                 39
2                 12
3                 15
4                 21
```

[55]:
```
### STAGE 1 - DATA CLEANING AND PREPROCESSING

df.shape
```

[55]: (1006029, 7)

[56]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1006029 entries, 0 to 1006028
Data columns (total 7 columns):
 #   Column          Non-Null Count    Dtype
---  ------          --------------    -----
 0   date            1006029 non-null  object
 1   state           1006029 non-null  object
 2   district        1006029 non-null  object
 3   pincode         1006029 non-null  int64
 4   age_0_5         1006029 non-null  int64
 5   age_5_17        1006029 non-null  int64
 6   age_18_greater  1006029 non-null  int64
dtypes: int64(4), object(3)
memory usage: 53.7+ MB
```

[57]:
```
## CHECKING MISSING VALUES

df.isnull().sum()
```

[57]:
```
date              0
state             0
district          0
pincode           0
age_0_5           0
age_5_17          0
age_18_greater    0
dtype: int64
```

[58]:
```
## CONVERTING DATE COLUMN TO DATETIME

df['date'] = pd.to_datetime(df['date'],format='%d-%m-%Y', errors='coerce')
```

```
[59]:  ## CHECKING NULL VALUES AGAIN

       df['date'].isnull().sum()
```

[59]: 0

```
[60]:  ## REMOVING INVALID ROWS IF THEY EXIST

       df = df.dropna(subset=['date'])
```

```
[61]:  ## CHECKING ON AGE WISE ENROLLMENT COLUMNS

       (df[['age_0_5','age_5_17','age_18_greater']]< 0 ).sum()
```

```
[61]: age_0_5         0
      age_5_17        0
      age_18_greater  0
      dtype: int64
```

```
[62]:  ## CREATING TOTAL ENROLLMENT COLUMNS FOR CLEAN ANALYSIS

       df['total_enrolments'] = (
           df['age_0_5'] +
           df['age_5_17'] +
           df['age_18_greater']
       )
```

```
[63]:  ## FEATURE ENGINEERING (EXTRACTING DATE, MONTH, YEAR)

       df['year'] = df['date'].dt.year
       df['month'] = df['date'].dt.month
       df['month_name'] = df['date'].dt.month_name()
```

```
[64]:  ## HANDLING UNWANTED SPACES IN DATASET

       df['state'] = df['state'].str.strip()
       df['district'] = df['district'].str.strip()
```

```
[65]:  ## FINAL CHECK

       df.head()
```

```
[65]:        date          state          district  pincode  age_0_5  age_5_17  \
      0 2025-03-02      Meghalaya  East Khasi Hills   793121       11        61
      1 2025-03-09      Karnataka   Bengaluru Urban   560043       14        33
      2 2025-03-09  Uttar Pradesh      Kanpur Nagar   208001       29        82
      3 2025-03-09  Uttar Pradesh           Aligarh   202133       62        29
      4 2025-03-09      Karnataka   Bengaluru Urban   560016       14        16
```

|	age_18_greater	total_enrolments	year	month	month_name
0	37	109	2025	3	March
1	39	86	2025	3	March
2	12	123	2025	3	March
3	15	106	2025	3	March
4	21	51	2025	3	March

```
[66]:   ## STAGE 3 - UNIVARIATE ANALYSIS

        # This analysis aims to understand the overall trend in Aadhaar enrolment
         →activity.
        # Q1. How does Aadhaar enrolment intensity vary across districts?


        # total enrollment columns
        df["total_enrolment"] = (
            df["age_0_5"] + df["age_5_17"] + df["age_18_greater"]
        )


        #district level
        district_enrolment = (
            df.groupby(["state", "district"])["total_enrolment"]
            .sum()
            .reset_index()
        )


        # the plot
        top_districts = district_enrolment.sort_values(
            "total_enrolment", ascending=False
        ).head(15)

        plt.figure(figsize=(10, 6))
        plt.barh(
            top_districts["district"] + " (" + top_districts["state"] + ")",
            top_districts["total_enrolment"]
        )
        plt.xlabel("Total Aadhaar Enrolments")
        plt.title("Top 15 Districts by Aadhaar Enrolment Volume (2025)")
        plt.gca().invert_yaxis()
        plt.tight_layout()
        plt.show()
```

Top 15 Districts by Aadhaar Enrolment Volume (2025)
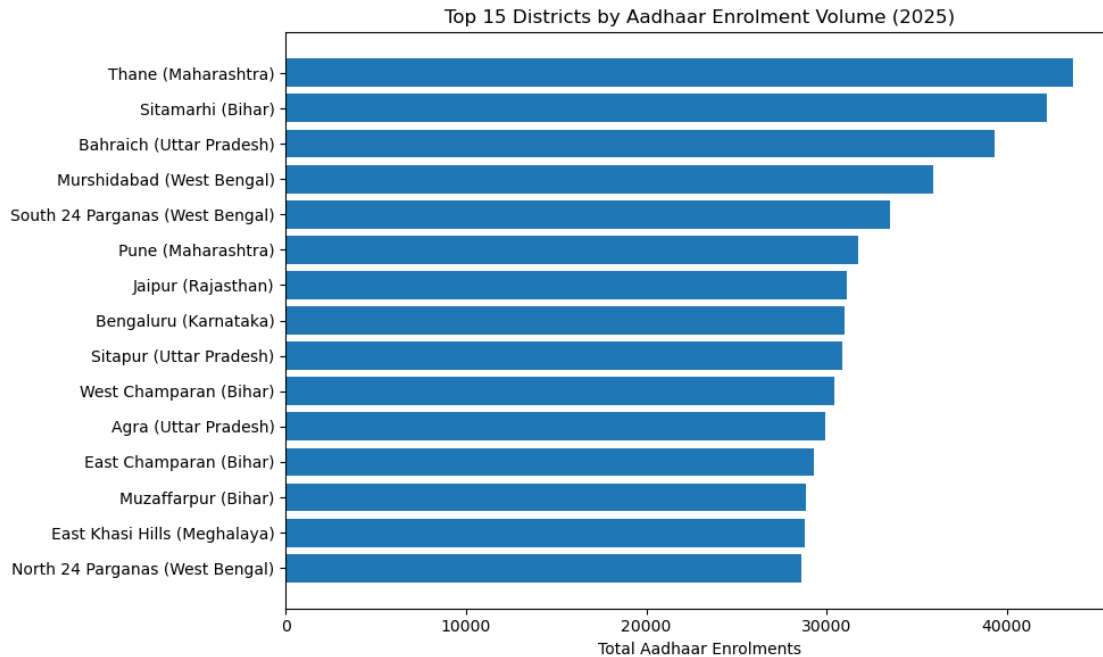
[67]:
```
# Insight:
# It's clear that Aadhaar enrolment isn't spread out evenly.
# We're seeing a massive rush in two main areas: big cities like Thane, Pune,
↪and Bengaluru, and heavily populated rural spots across Bihar, UP, and West
↪Bengal.
# This indicates concentrated enrolment demand in both metropolitan and
↪high-population rural regions, highlighting the need for district-specific
↪resource planning and enrolment infrastructure allocation.
```

[68]:
```
# This analysis helps identify which demographic group is driving Aadhaar
↪enrolment activity.
### Q2. Which age group contributes the most to Aadhaar enrolments?

age_distribution = df[['age_0_5', 'age_5_17', 'age_18_greater']].sum()

plt.figure()
age_distribution.plot(kind='bar')
plt.xlabel('Age Group')
plt.ylabel('Total Enrolments')
plt.title('Age-wise Distribution of Aadhaar Enrolments')
plt.show()
```

5

Age-wise Distribution of Aadhaar Enrolments
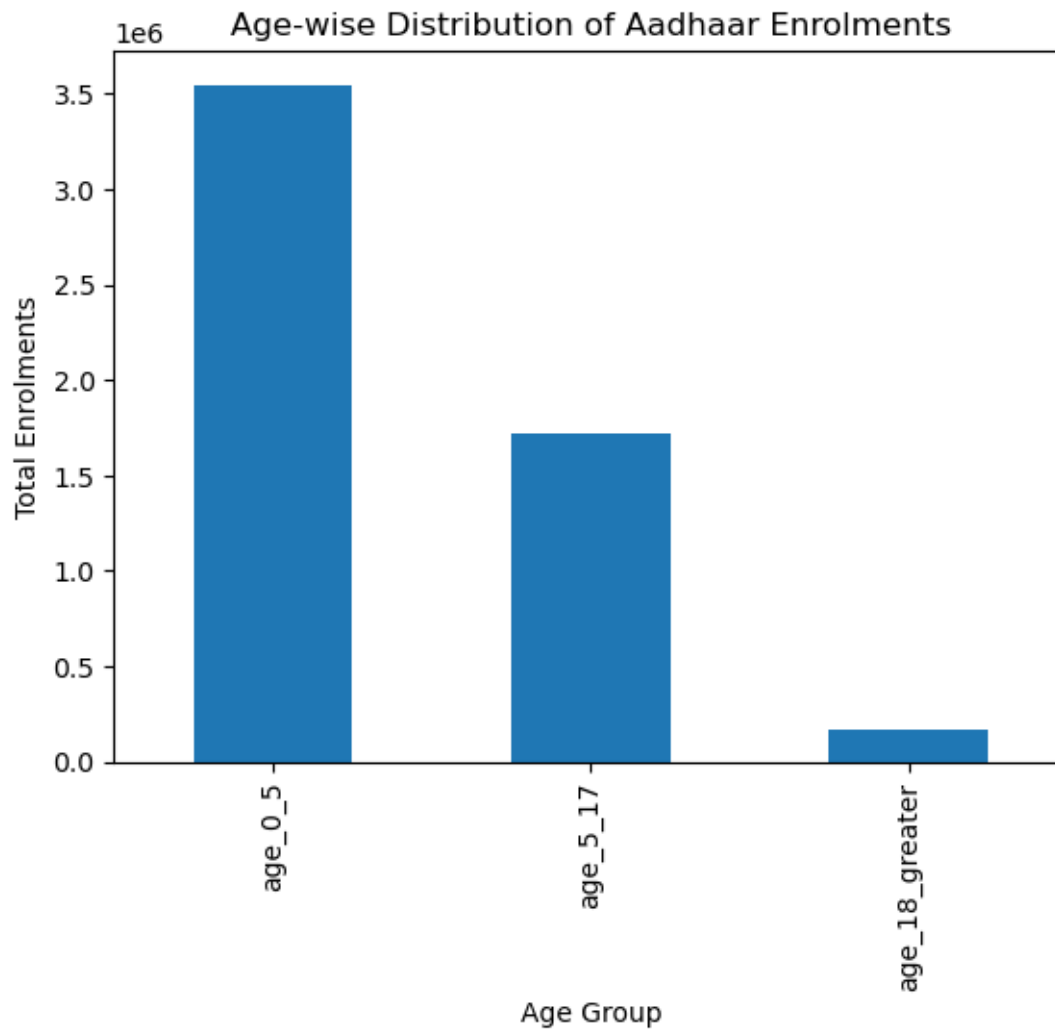
```
[69]: # Insight:
      # From the chart, we observe:

      # 0-5 age group contributes the highest number of enrolments

      # 5-17 age group also shows substantial enrolment activity

      # 18+ age group has significantly lower enrolments in this dataset

      ### This suggests that:
      # Enrolments among adults (18 years and above) are comparatively lower,␣
      ↪suggesting that Aadhaar coverage among the adult population is largely␣
      ↪saturated, while enrolment efforts are currently driven by early-age␣
      ↪registrations.
```

```
[70]: # This analysis identifies states with the highest enrolment activity,␣
      ↪indicating population density or effective enrolment infrastructure.
      # Q3. Which states record the highest Aadhaar enrolments?

      state_enrolment = (
          df.groupby('state')['total_enrolments']
          .sum()
          .sort_values(ascending=False)
          .head(10)
      )

      plt.figure()
      state_enrolment.plot(kind='bar')
      plt.xlabel('State')
      plt.ylabel('Total Enrolments')
      plt.title('Top 10 States by Aadhaar Enrolments')
      plt.show()
```
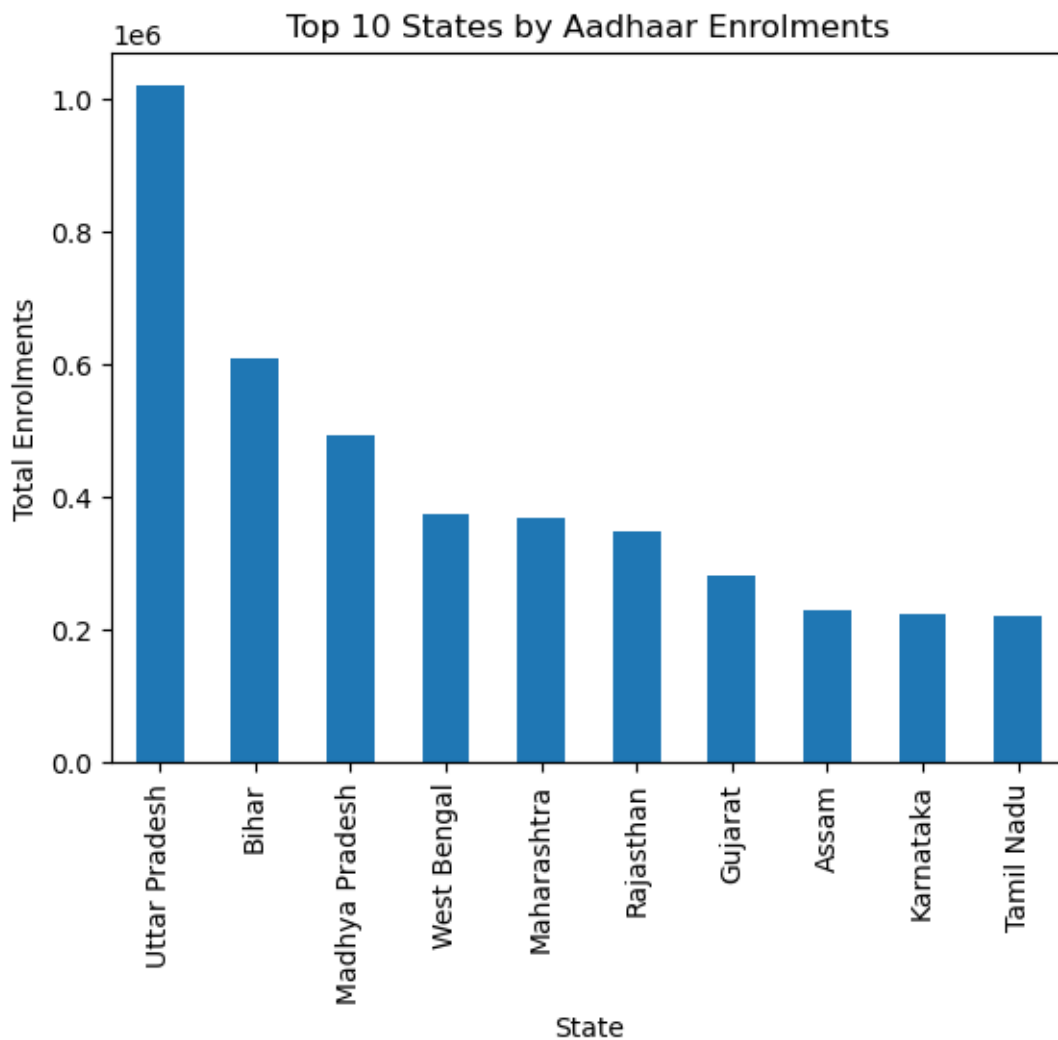
```
[71]: ### Insight:
      # From the chart:

      # Uttar Pradesh records the highest Aadhaar enrolments by a significant margin
      # Bihar and Madhya Pradesh also show high enrolment activity
      # Large and populous states dominate the top rankings
      # Southern and western states appear with comparatively lower enrolment volumes

      ### This suggests
      # States with larger populations dominate enrolment volumes, indicating a strong␣
       ↪relationship between population size and enrolment activity.
      # This highlights the need for proportionate infrastructure and resource␣
       ↪allocation in high-population states to ensure efficient Aadhaar services.

[72]: ## STAGE 4 - BIVARIATE ANALYSIS
      # This stage examines the relationship between two variables to uncover deeper␣
       ↪patterns in Aadhaar enrolment across demographic, geographic, and temporal␣
       ↪dimensions.


      # This analysis compares enrolment patterns of different age groups across␣
       ↪states to identify demographic disparities.
      ### Q1. How does age-wise Aadhaar enrolment vary across states?


      state_age = (
          df.groupby('state')[['age_0_5', 'age_5_17', 'age_18_greater']]
          .sum()
          .sort_values(by='age_18_greater', ascending=False)
          .head(10)
      )

      state_age.plot(kind='bar', stacked=True)
      plt.xlabel('State')
      plt.ylabel('Total Enrolments')
      plt.title('Age-wise Aadhaar Enrolments across Top 10 States')
      plt.show()
```
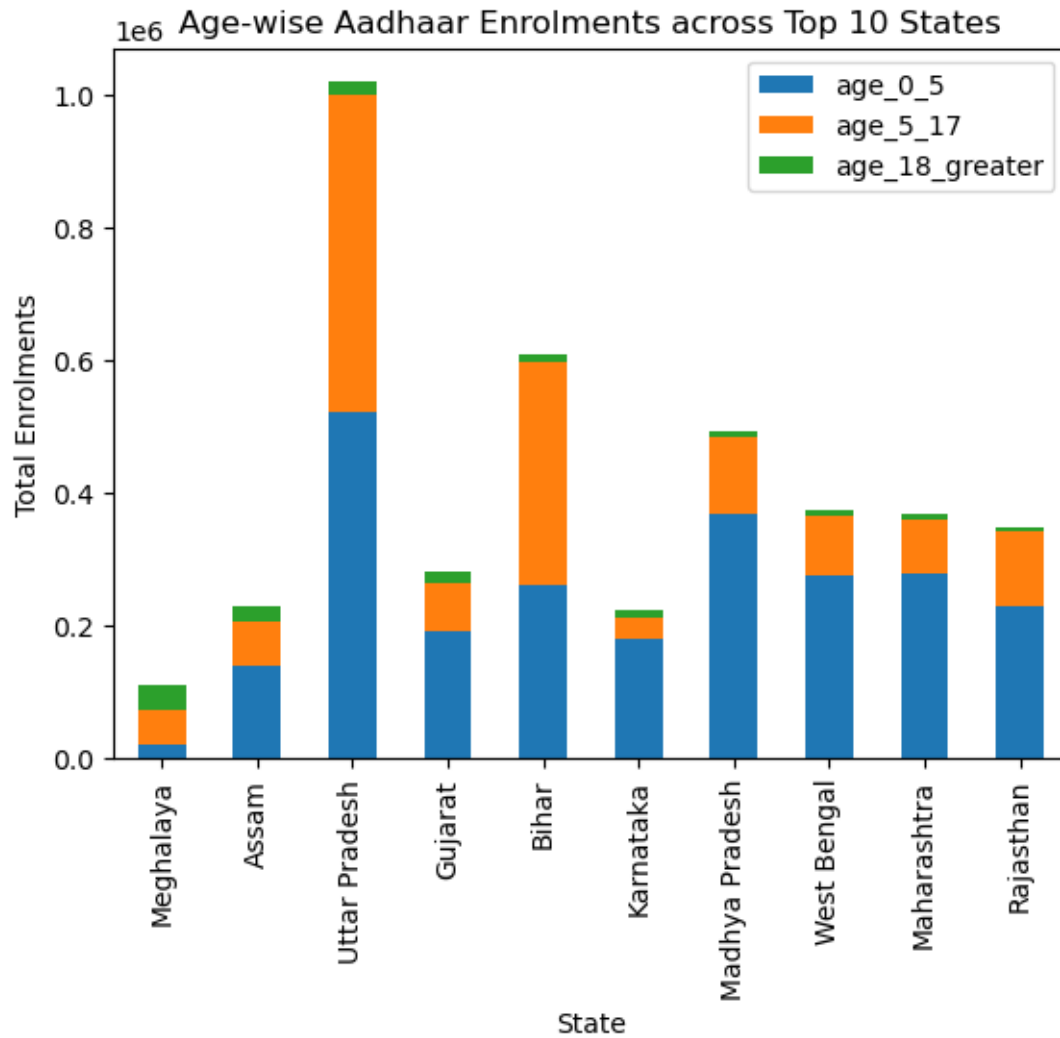
Age-wise Aadhaar Enrolments across Top 10 States

```
# Calculate totals
total_enrolments = df[["age_0_5", "age_5_17", "age_18_greater"]].sum()

# Plot
plt.figure()
plt.pie(
    total_enrolments,
    labels=["Age 0-5", "Age 5-17", "Age 18+"],
    autopct="%1.1f%%",
    startangle=90
)
plt.title("Percentage Contribution of Age Groups in Aadhaar Enrolment (2025)")
plt.tight_layout()
plt.show()
```

Percentage Contribution of Age Groups in Aadhaar Enrolment (2025)
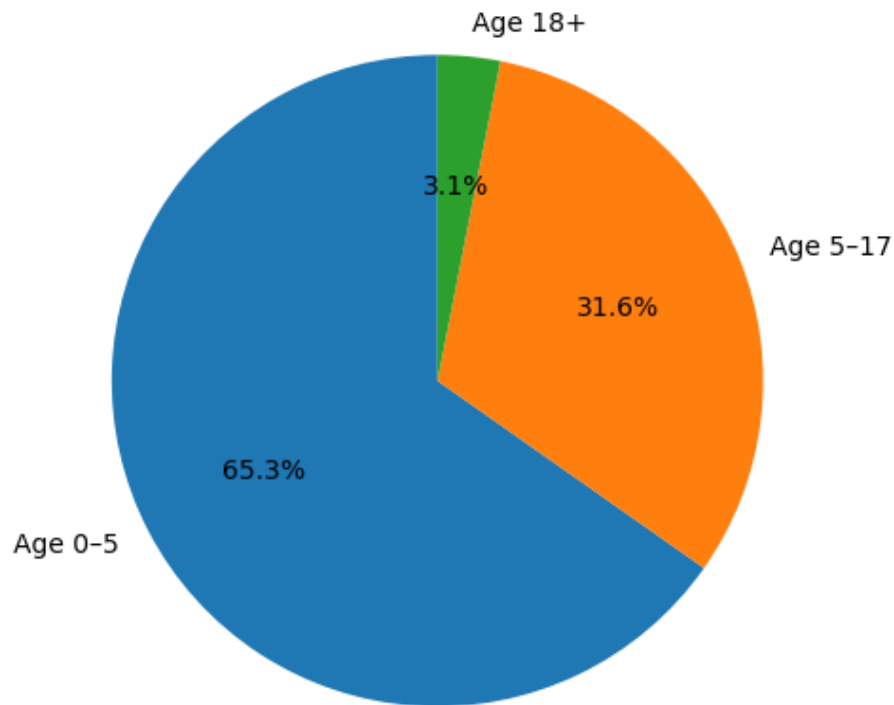
[75]:
```
### Insight:

# From the chart:
```

```python
# Dominant Group: Children aged 0-5 make up the largest share of enrolments at
↪65.3%. This shows a clear shift toward capturing citizens at birth.
# School-Age Contribution: The 5-17 age group accounts for nearly a third of
↪total enrolments (31.6%).
# Minimal Adult Enrolment: Adults (18+) contribute the smallest fraction at only
↪3.1% This indicates that adult enrolment has reached a state of near-total
↪saturation.
```

[76]:
```python
# This analysis identifies states where early-age Aadhaar enrolment is
↪comparatively stronger.
### Q3. Which states show relatively higher child enrolment compared to adult
↪enrolment?

# Aggregate state-level data
state_data = df.groupby("state")[["age_0_5", "age_5_17", "age_18_greater"]].sum()

# Calculate child enrolment ratio
state_data["child_ratio"] = (
    state_data["age_0_5"] + state_data["age_5_17"]
) / (
    state_data["age_0_5"] + state_data["age_5_17"] + state_data["age_18_greater"]
)

# Select top 10 states
top_states = state_data.sort_values("child_ratio", ascending=False).head(10)

# Plot
plt.figure(figsize=(10, 5))
plt.barh(top_states.index, top_states["child_ratio"])
plt.xlabel("Child Enrolment Ratio")
plt.title("Top 10 States by Child Aadhaar Enrolment Ratio (2025)")
plt.gca().invert_yaxis()
plt.tight_layout()
plt.show()
```
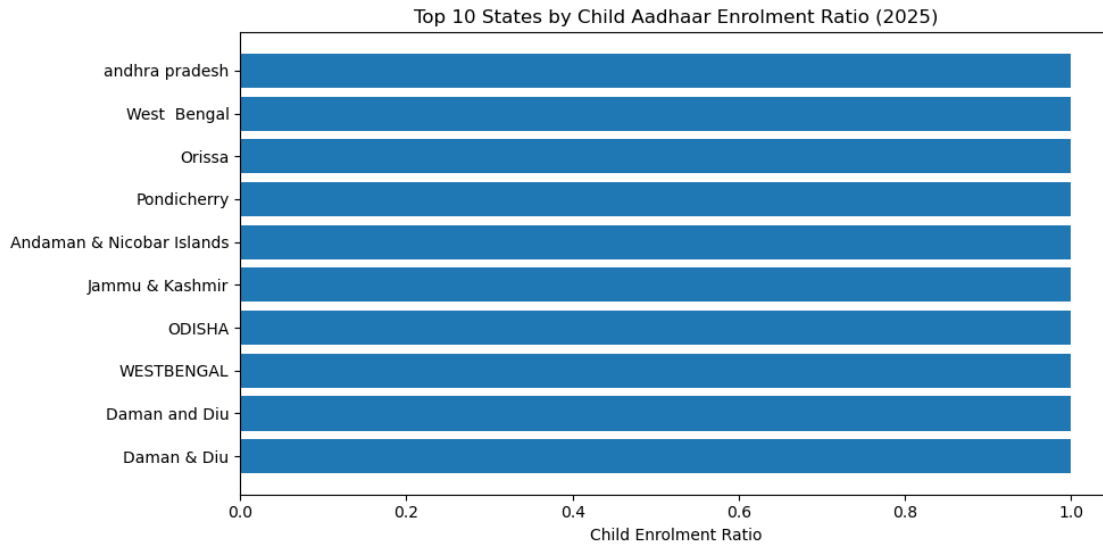
Top 10 States by Child Aadhaar Enrolment Ratio (2025)

[77]: ### Insight:

Perfect Saturation: Several states and UTs, including Andhra Pradesh, West␣
↪Bengal, and Odisha, have reached a 1.0 Child Enrolment Ratio, indicating␣
↪near-universal Aadhaar coverage for children in these regions.
Effective Child Outreach: The presence of both large states and smaller UTs␣
↪(like Pondicherry and Daman & Diu) suggests that child-focused registration␣
↪drives are successful across varying geographic scales.
Near-Zero Adult Gap: In these top-performing regions, child enrollment is no␣
↪longer lagging behind adult enrollment, representing a fully matured␣
↪registration ecosystem for all age groups.

[78]: # This analysis explores whether enrolment activity varies across different␣
↪months.
### Q4. Is there a seasonal pattern in Aadhaar enrolments?

monthly_enrolment = df.groupby('month_name')['total_enrolments'].sum()

monthly_enrolment = monthly_enrolment.reindex([
    'January','February','March','April','May','June',
    'July','August','September','October','November','December'
])

plt.figure()
monthly_enrolment.plot(kind='bar')
plt.xlabel('Month')
plt.ylabel('Total Enrolments')
plt.title('Monthly Aadhaar Enrolment Pattern')

```
plt.show()
```

Monthly Aadhaar Enrolment Pattern
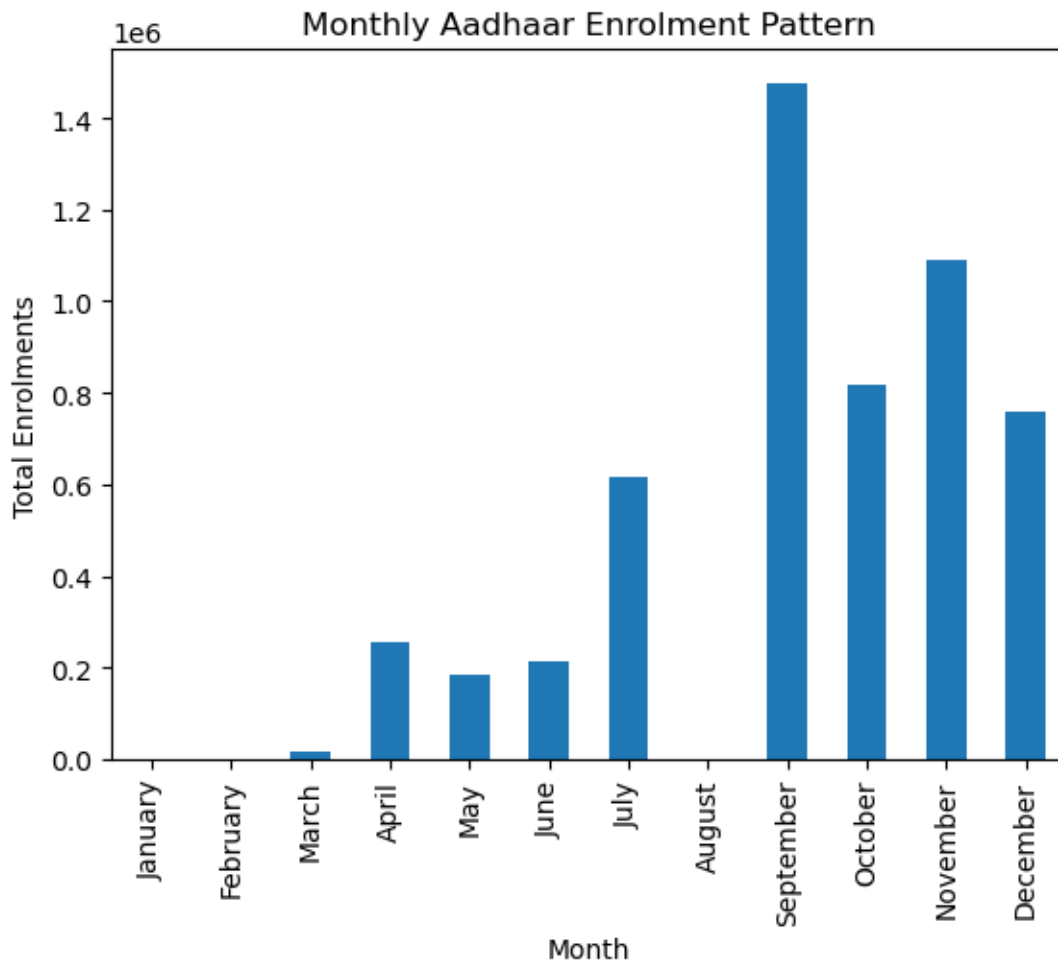
[79]: ### Insight:

```
# Year-End Surge: A strong seasonal trend is visible, with the vast majority of␣
↪enrolments occurring between September and December.

# Educational Alignment: The spike starting in July/September likely aligns with␣
↪school admission cycles and requirements for updated student documentation.

# First Quarter Lull: The near-absence of data in Q1 (Jan-Mar) suggests either a␣
↪seasonal pause in registration drives or a specific administrative reporting␣
↪cycle.

# September Peak: September stands out as the most critical month, accounting␣
↪for the highest volume of total Aadhaar enrolments in the period shown.
```

```
[80]: ## CHECKING THE YEAR TO WHICH THE DATASET BELONGS.

      df['year'].value_counts()
```

```
[80]: year
      2025    1006029
      Name: count, dtype: int64
```

```
[81]: ### NOTE - DATA LIMITATION
      ## The available Aadhaar enrolment dataset covers a limited temporal window␣
       ↪(year 2025 only).
      ## As a result, long-term trend analysis and forecasting are constrained.
      ## The analysis will now focus on cross-sectional demographic and geographic␣
       ↪patterns to derive actionable insights.
```

```
[82]: ## STAGE 5 - TRIVARIATE ANALYSIS
      ### Q1. Which states exhibit significant imbalance in age-wise Aadhaar␣
       ↪enrolments?


      state_age_prop = (
          df.groupby('state')[['age_0_5', 'age_5_17', 'age_18_greater']]
          .sum()
      )

      state_age_prop = state_age_prop.div(state_age_prop.sum(axis=1), axis=0)

      top_states = state_age_prop.sort_values('age_18_greater', ascending=False).
       ↪head(10)

      plt.figure()
      top_states.plot(kind='bar', stacked=True)
      plt.xlabel('State')
      plt.ylabel('Proportion of Enrolments')
      plt.title('Age-wise Enrolment Proportion across States')
      plt.show()
```
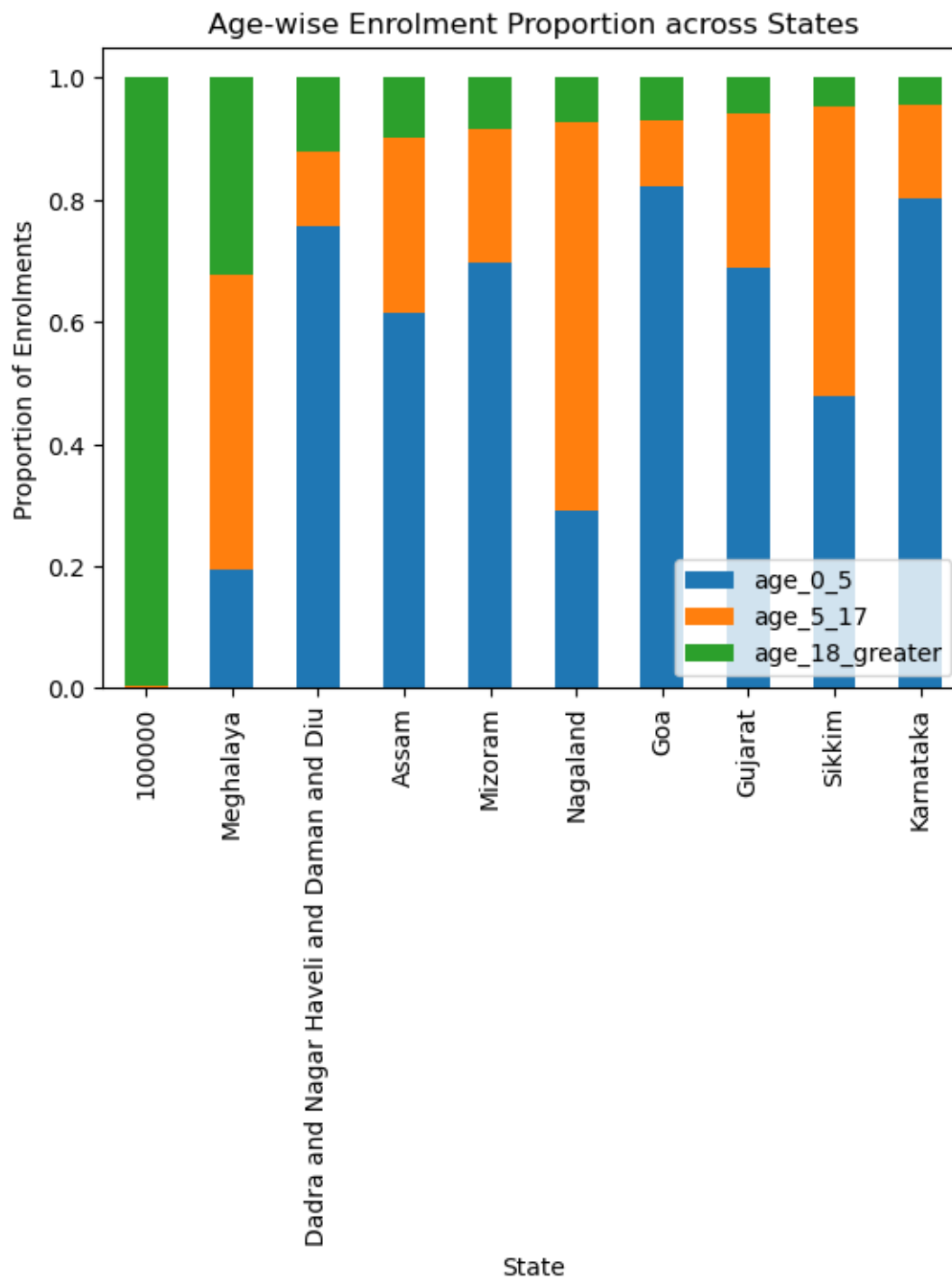
<Figure size 640x480 with 0 Axes>

Age-wise Enrolment Proportion across States

[83]: ### Insight:

Demographic Imbalance: Most states have moved into a "maintenance" phase where␣
↪enrolments are 80-90% children, as adult coverage is already fully saturated.

```
# Data Anomaly: The entry for "100000" is a significant outlier, showing almost␣
 ↪exclusively adult enrolments, which contradicts the national trend of␣
 ↪child-led growth.

# Regional Variation: Meghalaya shows the most significant ongoing adult␣
 ↪enrolment activity, suggesting it may still be working toward universal␣
 ↪coverage for its mature population.

# Infant Focus: In high-performing states like Goa and Karnataka, the focus has␣
 ↪shifted almost entirely to newborns, with the 0-5 age group forming the core␣
 ↪of all new registration activity.
```

[84]:
```
### Q2. Are there districts with unusually high Aadhaar enrolment activity?

district_enrolment = df.groupby('district')['total_enrolments'].sum()

threshold = district_enrolment.mean() + 2 * district_enrolment.std()
outlier_districts = district_enrolment[district_enrolment > threshold]

outlier_districts.sort_values(ascending=False).head(10)
```

[84]:
```
district
Thane               43688
Sitamarhi           42232
Bahraich            39338
Murshidabad         35911
South 24 Parganas   33540
Pune                31763
Jaipur              31146
Bengaluru           30980
Sitapur             30854
Hyderabad           30830
Name: total_enrolments, dtype: int64
```

[85]:
```
### Insight:

# Urban-Rural Hotspots: High activity is split between metropolitan centers like␣
 ↪Thane or Pune and populous rural districts in Bihar and West Bengal,␣
 ↪indicating broad-based registration drives.

# Regional Leaders: Thane and Sitamarhi stand out as the primary drivers of␣
 ↪enrolment volume, significantly outperforming other major districts.

# Sustained Demand: All top 10 districts maintain a high baseline of 30,000+␣
 ↪enrolments, reflecting robust administrative capacity and consistent demand␣
 ↪for new registrations.
```

```
[86]: ### Q3. How does child enrolment vary across districts within high-enrolment␣
      ↪states?

      top_states_list = state_age_prop.sum(axis=1).sort_values(ascending=False).
      ↪head(3).index

      filtered = df[df['state'].isin(top_states_list)]

      district_child = (
          filtered.groupby(['state', 'district'])['age_0_5']
          .sum()
          .reset_index()
      )

      district_child.head()
```

```
[86]:          state      district  age_0_5
      0        100000        100000        0
      1   Lakshadweep   Lakshadweep      192
      2   Maharashtra    Ahilyanagar       12
      3   Maharashtra     Ahmadnagar     9369
      4   Maharashtra    Ahmed Nagar      277
```

```
[87]: ### Insights:

      # Intra-State Imbalance: High-performing states still exhibit internal "blind␣
      ↪spots" where child enrolment is significantly lower in certain districts␣
      ↪despite high overall state averages.

      # Localized Enrollment Drives: Districts with high counts (like Ahmadnagar)␣
      ↪likely benefit from permanent enrolment centers or targeted camps in schools␣
      ↪and anganwadis.

      # Saturation Gaps: The low numbers in districts like Ahilyanagar suggest either␣
      ↪a smaller target population or a need for improved accessibility to enrolment␣
      ↪services.

      # Administrative Focus: Disparities highlight where state governments may need␣
      ↪to redeploy staff to ensure universal coverage for children entering␣
      ↪school-going age.
```

```
[ ]:
```