Aadhaar Enrolment Patterns in India (March-December 2025)

UIDAI Aadhaar Data Hackathon - Project Report

Project Overview

This project explores Aadhaar enrollment patterns across India using UIDAI data, focusing on monthly trends, geographic distribution, and age-group composition. The findings are summarized through statistical analysis and an interactive Power BI dashboard.

1. Problem Statement and Approach

Problem Statement

Aadhaar enrolment is a critical public digital infrastructure process in India. Understanding **where**, **when**, and **for which age groups** enrolments occur can help identify regional demand patterns, seasonal variations, and demographic trends. However, raw enrolment data is large, inconsistent, and geographically fragmented, making meaningful insights difficult without structured analysis.

Approach

This project analyses Aadhaar enrolment data at **district and state levels** using a **data-driven, multi-level analytical approach**. The analysis focuses on:

- Temporal patterns (monthly trends)
- Geographic distribution (districts and states)
- Demographic composition (age groups)

Python was used for **data cleaning, preprocessing, aggregation,** and **exploratory analysis**, while **Power BI** was used to build an **interactive, map-based dashboard** for visual storytelling.

2. Datasets Used

Dataset

- UIDAI Aadhaar Enrolment Dataset (All-India)
- Time period: **March 2025 - December 2025**

Key Columns Used

Column Name	Description
date	Enrolment date
state	State name
district	District name
pincode	Enrolment location pincode
age_0_5	Enrolments for age group 0-5
age_5_17	Enrolments for age group 5-17
age_18_greater	Enrolments for age group 18+

Derived Columns

- month - Extracted from date
- total_enrollments - Sum of all age groups
- age_group_share_% – Percentage contribution of each age group

3. Methodology

3.1 Data Cleaning

- Removed invalid records (rows with numeric values in state/district fields).
- Standardised **state** and **district** names to resolve spelling variations, casing issues, and legacy names.
- Retained duplicate records where enrolments occurred on the same date across different **pincodes**, as these represent valid enrolment activity.

3.2 Data Preprocessing

- Converted date fields to proper datetime format.
- Aggregated data at: District-Month level, State-Month level
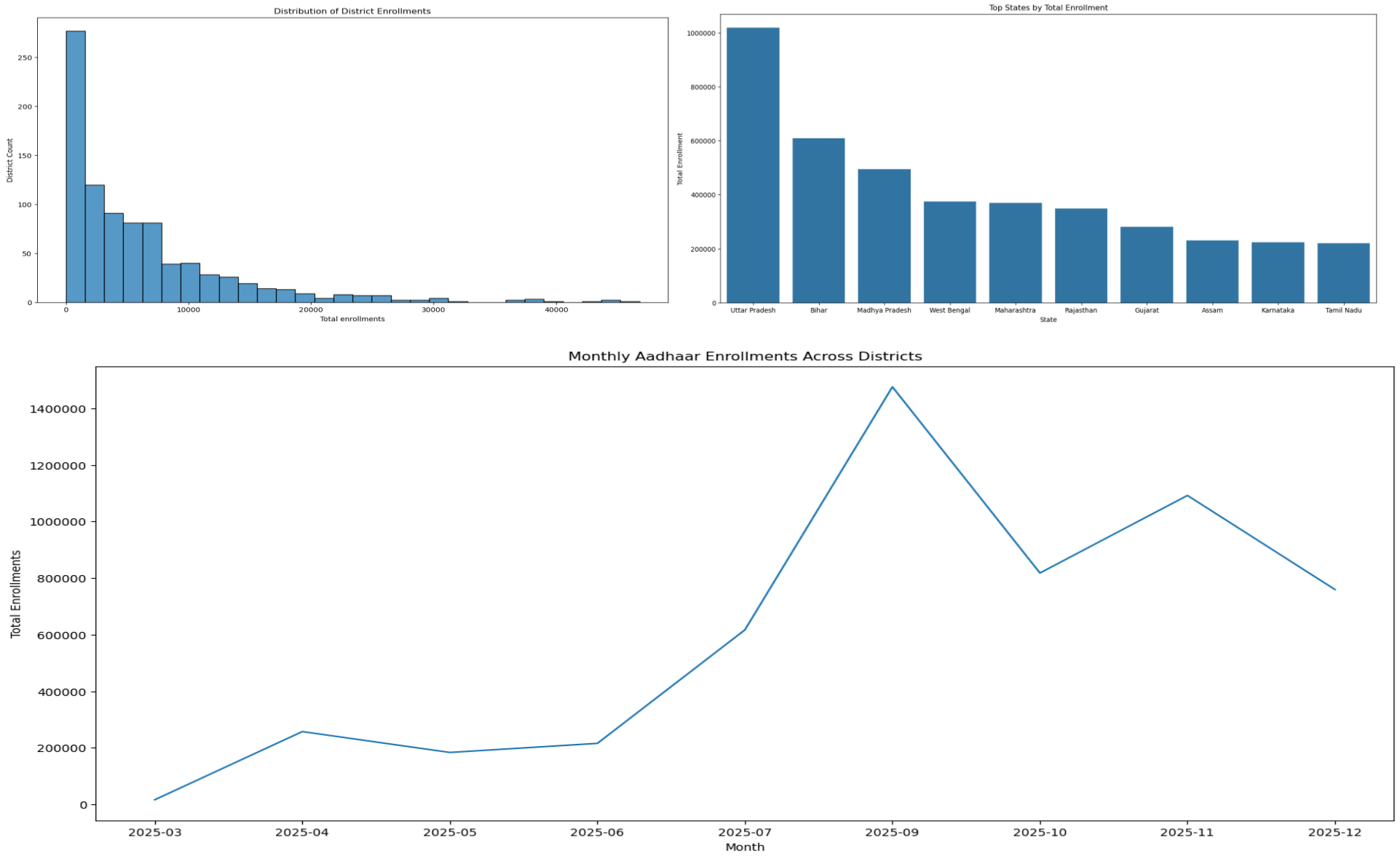- Created separate aggregation tables to preserve the integrity of the original dataset.

3.3 Feature Engineering

- Computed total enrolments per record.
- Calculated age group percentage distribution.
- Prepared long-format and aggregated datasets to support univariate, bivariate, and trivariate analysis.
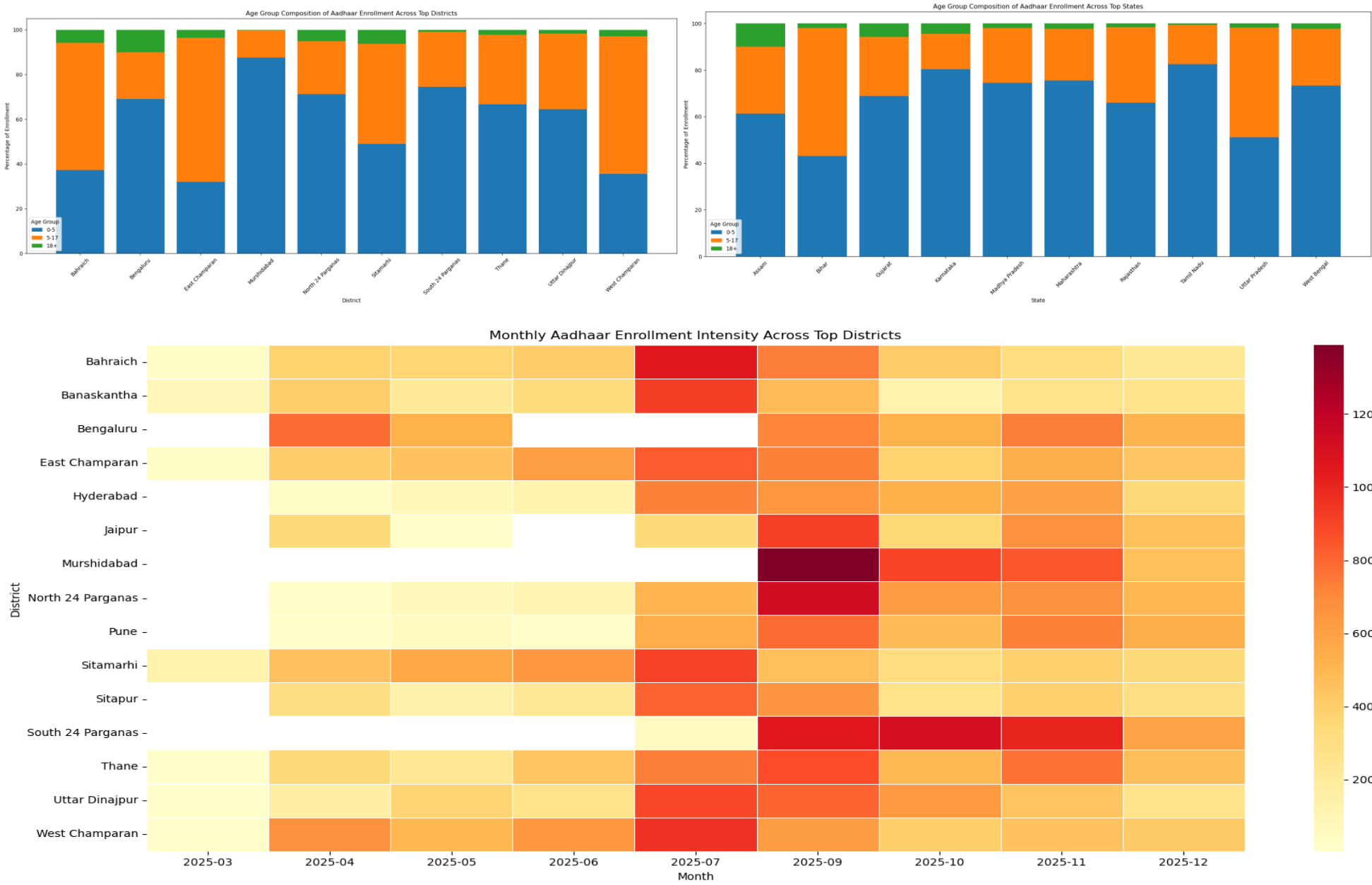
4. Data Analysis and Visualisation

4.1 Univariate Analysis

- **District-level enrolments** show a right-skewed distribution, with most districts below 10,000 enrolments and a few high-volume outliers.
- **Monthly trends** indicate a sharp rise in enrolments from July, peaking in September, followed by a decline towards December.
- **State-level analysis** shows Uttar Pradesh and Bihar as the highest enrolment contributors.
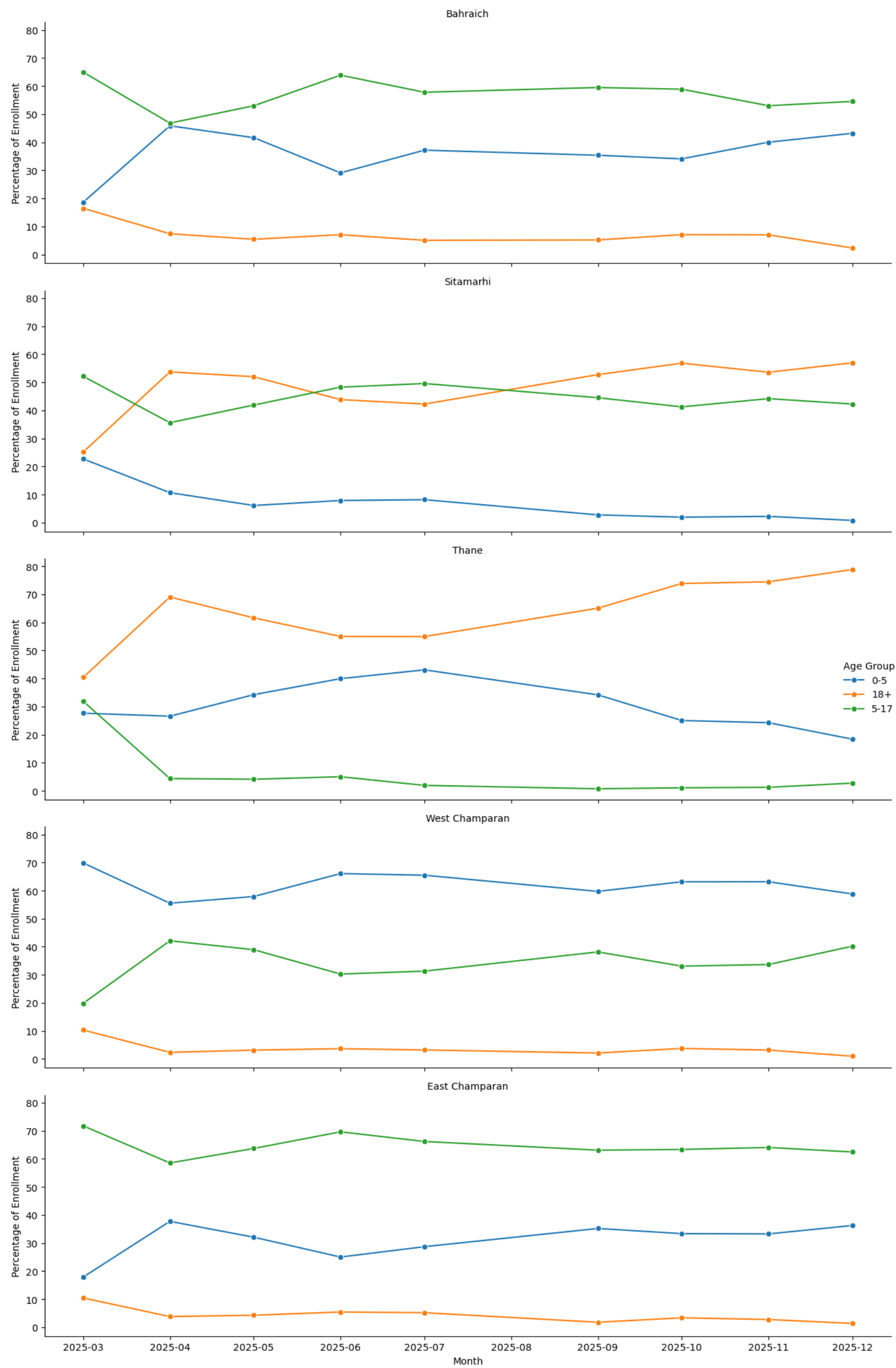

4.2 Bivariate Analysis

- District-wise monthly trends reveal uneven growth patterns, with some districts showing sharp spikes and others remaining stable.
- Age group composition varies significantly by district and state, highlighting regional demographic differences.

4.3 Trivariate Analysis

- **District × Month × Age Group** analysis reveals:
    - 0–5 age group dominates enrolments across most regions.
    - 18+ enrolments remain consistently lower.
    - Certain districts show temporary spikes driven primarily by child enrolments.

4.4 Dashboard Summary and Insights

A single-page interactive Power BI dashboard was created to summarise and visualise the results of the analysis. It combines temporal trends, geographic distribution, and age-group composition of Aadhaar enrollment across India at state and district levels

Key Indicators

- **Total Enrollment (5M):** Represents cumulative Aadhaar enrollments across all records in the dataset.
- **Average Monthly Enrollment (603.94K):** Captures the typical monthly enrollment volume, highlighting deviations and seasonal intensity.
- **Dominant Age Group (0–5 Years):** Indicates that early-age registrations account for the largest share of enrollments across regions.

Analytical Visuals and Findings

Total Enrollment by Month

A time-series line chart reveals a steady rise in enrollments from March, a sharp peak in **September**, followed by a decline towards year-end. This non-linear trend suggests seasonality or administrative drives rather than uniform enrollment behaviour.

Top 10 States by Total Enrollment

A ranked bar chart highlights strong geographic concentration. **Uttar Pradesh** leads with over **1 million enrollments**, followed by Bihar and Madhya Pradesh, indicating enrollment volume closely aligns with population density and regional scale.

Age Group Composition of Aadhaar Enrollment

A 100% stacked bar chart shows that the **0–5 age group consistently dominates enrollment shares across states**, while enrollments in the **18+ category remain minimal**, reinforcing the observation that Aadhaar enrollment is driven primarily by early-life registration.

Monthly Enrollment Trends Across Top Districts

Small multiple line charts capture district-level variability over time. While national trends are consistent, districts such as **East Champaran and Sitamarhi** exhibit sharper mid-year spikes, highlighting localized enrollment dynamics.

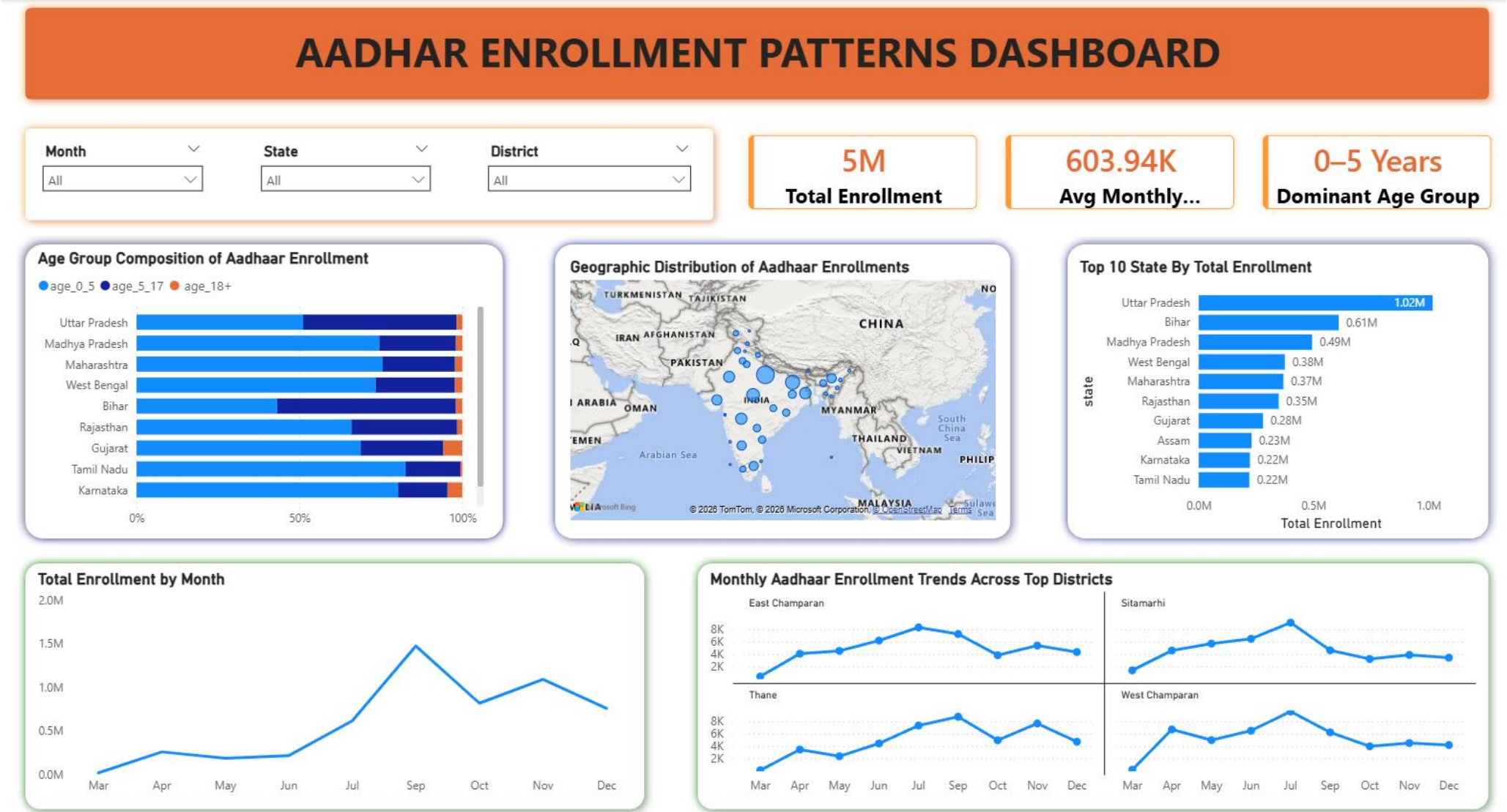#### Geographic Distribution of Aadhaar Enrollments

An interactive map visualizes enrollment intensity at the state level. Bubble size reflects total enrollments, and the map dynamically responds to filters and selections from other visuals, enabling geographic drill-down analysis.

Interactivity and Design

- Slicers for **Month, State, and District** allow flexible exploration of the data.
- All visuals are cross-filtered, ensuring coherent interactions across time, geography, and demographics.
- A clean, minimal design prioritizes readability while maintaining analytical depth.

Overall Insight

The dashboard reveals seasonal enrollment patterns, high geographic concentration, and a dominant contribution from the 0–5 age group. Significant district-level variation is observed across months, indicating that national trends mask localized enrollment dynamics.

5. Code and Tools Used

Tools

- **Python**: Pandas, NumPy, Matplotlib, Seaborn
- **Power BI**: Data modelling, DAX measures, interactive visuals

Code Artefacts

Jupyter Notebook containing:

- Data loading
- Cleaning and preprocessing
- Feature engineering
- Aggregations

6. Conclusion

This analysis reveals clear patterns in Aadhaar enrollment across India, highlighting strong monthly seasonality, uneven geographic distribution, and a dominant contribution from early-age registrations. District-level variation plays a significant role in overall enrollment behaviour, underscoring the importance of localized analysis. The integration of statistical analysis with an interactive Power BI dashboard enables both high-level monitoring and granular exploration of enrollment trends.