

Unlocking Societal Trends in Aadhaar Enrolment and Updates

District-wise Analysis of Madhya Pradesh

Problem Statement : Aadhaar enrolment is a key indicator of access to government services, welfare schemes and digital inclusion. While Aadhaar coverage is widespread, enrolment activity is not evenly distributed across regions or age groups.

Some districts show very high activity while others lag behind. Understanding these patterns is important for improving outreach, planning enrolment infrastructure and ensuring inclusive coverage.

This project studies Aadhaar enrolment data for **Madhya Pradesh** at the district level to identify clear trends, gaps and imbalances in enrolment behaviour.

Key objective : The goal of this project is to convert Aadhaar enrolment data into simple actionable insights that can support better planning and targeted interventions.

Submitted by

Trishansh Sahane

Data Hackathon Participant (Team ID : **UIDAI_3533**)

Code and Dataset Repository

All code files, cleaned datasets and visualizations used in this project are publicly available on GitHub to ensure transparency and reproducibility.

GitHub Repository :

<https://github.com/trish-2610/Aadhaar-enrolment-analysis-UIDAI-Data-Hackathon-2026>

Repository Structure :

```
project/
|
├─ data/
|   ├── aadhaar_dataset.csv
|   ├── aadhaar_cleaned_dataset.csv
|   ├── aadhaar_state_level_summary.csv
|   └─ aadhaar_district_level_summary.csv
|
├─ notebooks/
|   ├── Data_Cleaning_and_Preprocessing.ipynb
|   └─ Data_Analysis_and_Visualizations.ipynb
|
├─ README.md
└─ requirements.txt
```

Dataset Used :

UIDAI Aadhaar Enrolment Dataset
District-level data for Madhya Pradesh

Tools Used :

Python
Pandas
Seaborn and Matplotlib
Plotly for interactive visuals

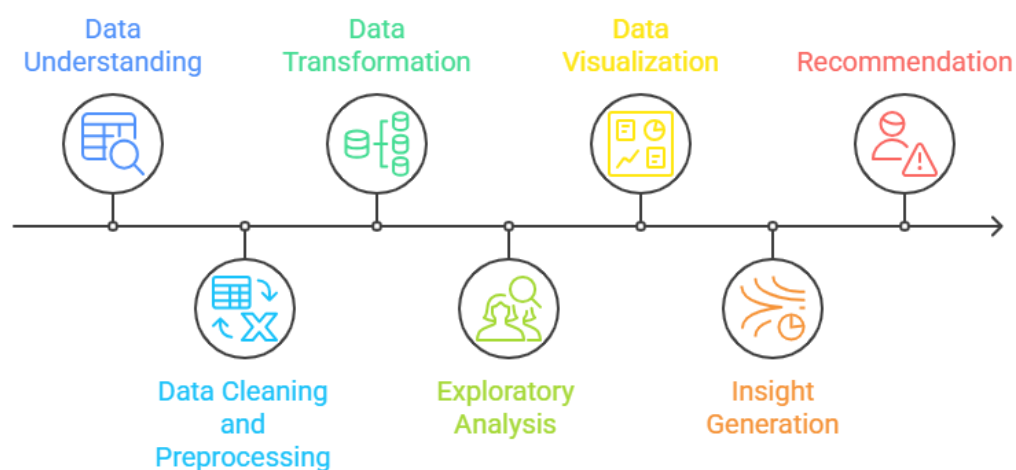
Approach Used :

- Cleaned and structured the raw UIDAI dataset for **district-level analysis**.
- Performed univariate analysis to study individual variables like total enrolments.
- Applied bivariate analysis to compare enrolments across districts and time.
- Used trivariate analysis to examine **district-wise age group patterns**.
- Used clear and simple visualizations to communicate insights.

Output of the Analysis:

- District-wise enrolment comparisons
- Age-group based enrolment patterns
- Time-based enrolment trends
- Identification of high-performing and low-performing districts
- Detection of unusual enrolment behaviour across districts

Methodology



1. **Data Description and understanding** – Reviewed dataset structure and scope for district-level analysis.
2. **Data Cleaning** – Removed unnecessary columns, standardized names, and formatted dates.
3. **Data Transformation** – Aggregated data by district and time and computed total enrolments.
4. **Exploratory Analysis** – Analyzed enrolment patterns using univariate, bivariate, and trivariate views.
5. **Data Visualization** – Used simple static and interactive charts to reveal trends.
6. **Insight Generation** – Identified key enrolment gaps, patterns, and disparities.
7. **Recommendations** – Suggested district-focused actions for better outreach and planning.

Dataset Description

The analysis uses the Aadhaar Enrolment dataset provided by UIDAI. The dataset contains structured records related to Aadhaar enrolment activity across districts in Madhya Pradesh over time.

Geographical Scope :

- **State** : Madhya Pradesh
- **Level of analysis** : District-wise

Time Coverage :

- Date-wise enrolment records used to study trends and variations over time

Key Columns Used :

- **district** : Name of the district where Aadhaar enrolment or update was recorded
- **date** : Date of enrolment activity
- **age_group_0_to_5** : Number of Aadhaar enrolments for children aged 0 to 5 years
- **age_group_5_to_17** : Number of Aadhaar enrolments for individuals aged 5 to 17 years
- **age_group_18_plus** : Number of Aadhaar enrolments for adults aged 18 years and above
- **total_enrolments** : Total number of Aadhaar enrolments recorded for the district on a given date

Note : **PinCode** column is removed during data cleaning as it of no use for the analysis.

Data Cleaning and Transformation

1. Dataset Loading

The raw UIDAI Aadhaar enrolment dataset was loaded from a CSV file into a Pandas DataFrame to examine its structure, columns and overall data quality.

2. Initial Data Inspection

Basic checks were performed to understand column names, data types, missing values and the level of granularity (state, district, date, and age group).

3. Column Renaming and Standardization

Age-related columns were renamed to clear and consistent formats such as `age_group_0_to_5`, `age_group_5_to_17`, and `age_group_18_plus` to improve readability and avoid ambiguity during analysis.

4. Removal of Irrelevant Columns

Columns not required for district-level analysis, such as “**pincode**” and other non-essential identifiers, were removed to keep the dataset focused and reduce noise.

5. Date Formatting

The date column was converted into a proper datetime format to enable time-series analysis and trend-based visualizations.

6. Total Enrolments Calculation

A new column, `total_enrolments` was created by summing enrolments across all age groups for each record, providing a single measure of overall enrolment activity.

7. Data Aggregation

- **State-level aggregation** was performed by grouping records by state and summing total enrolments to understand overall enrolment volume.
- **District-level aggregation** was performed by grouping data by state and district to compute total enrolments at the district level.

8. Creation of Analysis-ready Datasets

Separate structured datasets were prepared for :

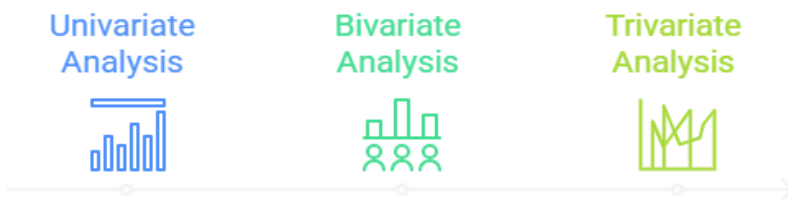
- Cleaned detailed enrolment data
- State-level summary data
- District-level summary data

9. Data Export

All cleaned and aggregated datasets were exported as CSV files to support further analysis and visualization workflows.

These steps ensured the raw UIDAI data was transformed into clean, structured and analysis-ready datasets suitable for district-wise and demographic analysis.

Exploratory Analysis



Univariate Analysis :

- **District-wise Aadhaar Enrolments (Bar Chart)**
Shows total enrolments by district.
Insight : Indore has the highest enrolments, and Aadhaar activity is concentrated in a few major districts.
- **Distribution of District-wise Enrolments (Histogram)**
Shows how enrolments are spread across districts.
Insight : Most districts have low to moderate enrolments, while only a few districts contribute very high numbers.
- **Age-group Distribution of Aadhaar Enrolments (Pie Chart)**
Shows share of enrolments by age group.
Insight : Enrolments are largely driven by the 0–5 age group, with very low adult enrolments.

Bivariate Analysis :

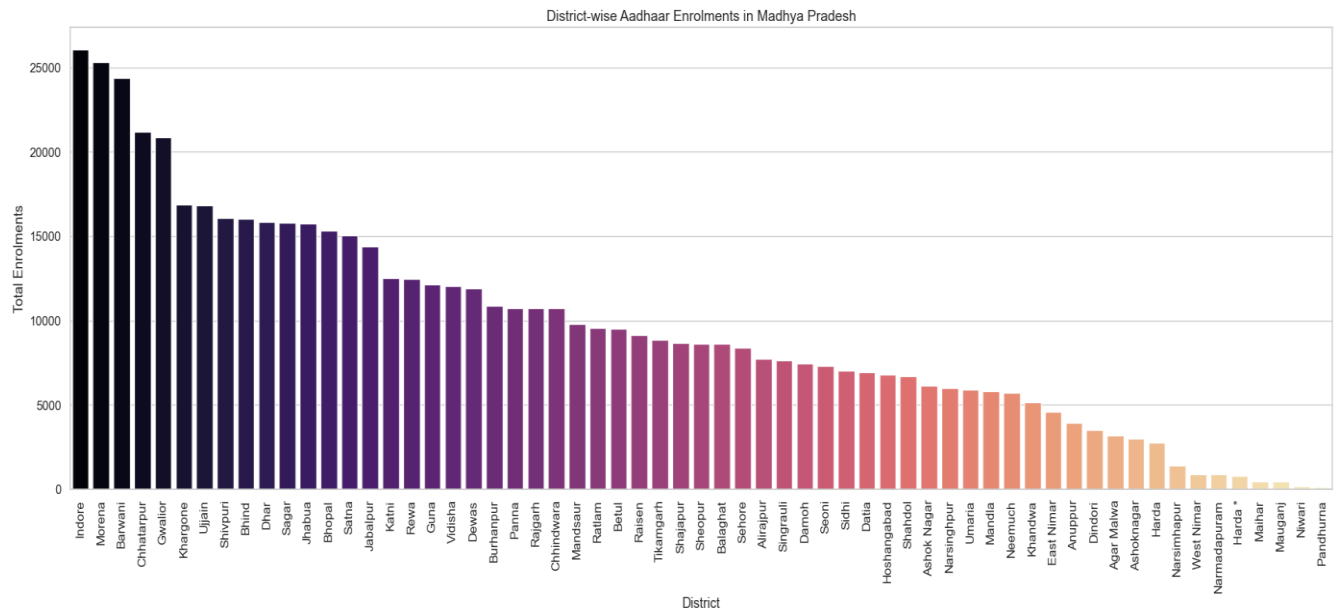
- **District-wise Age-group Enrolments (Grouped/Stacked Bar Chart)**
Compares age-group enrolments across districts.
Insight : High-enrolment districts achieve their totals mainly through child enrolments, while adult enrolments remain low everywhere.
- **Aadhaar Enrolments Over Time (Line Chart)**
Shows total enrolments across time.
Insight : Sharp enrolment spikes indicate time-bound drives rather than steady enrolment activity.

Trivariate Analysis :

- **Top Districts Enrolment Trends Over Time (Multi-line Chart)**
Compares time trends for major districts like Indore and others.
Insight : All major districts show enrolment spikes, but the intensity varies by district.
- **District-level Enrolment Anomalies (Scatter Plot)**
Shows districts with unusual enrolment behaviour.
- **Insight** : Some districts deviate from normal patterns, indicating the need for closer administrative review.

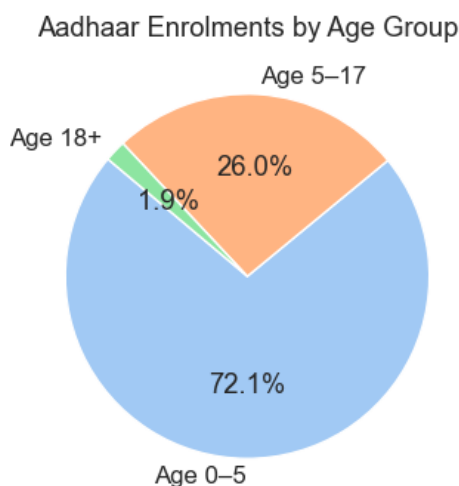
Data Analysis and Visual Insights

1. District-wise Aadhaar Enrolments in Madhya Pradesh.



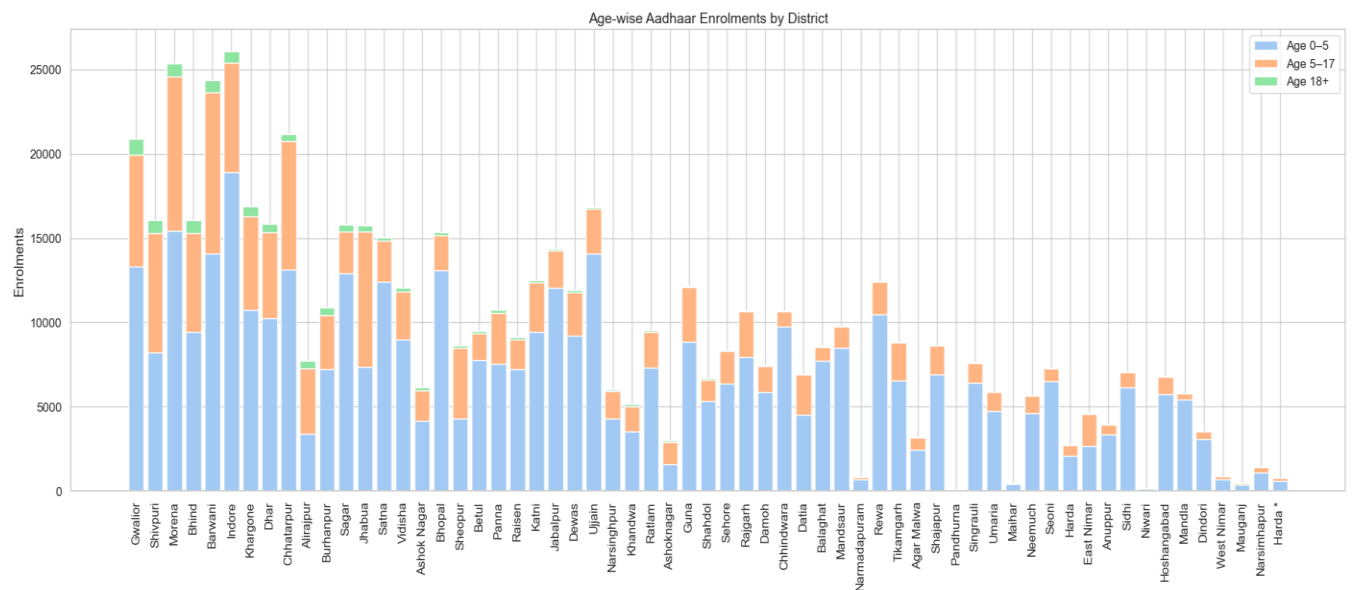
Interpretation : Aadhaar enrolments are **highly uneven** across districts in Madhya Pradesh. **Indore** records the **highest number of enrolments**, clearly standing out from other districts. A small group of districts such as **Bhopal** and **Jabalpur** also show high activity, while many districts report **significantly lower enrolments**. This indicates that Aadhaar enrolment efforts and population coverage are concentrated in a few major districts.

2. Age-group Distribution of Aadhaar Enrolments



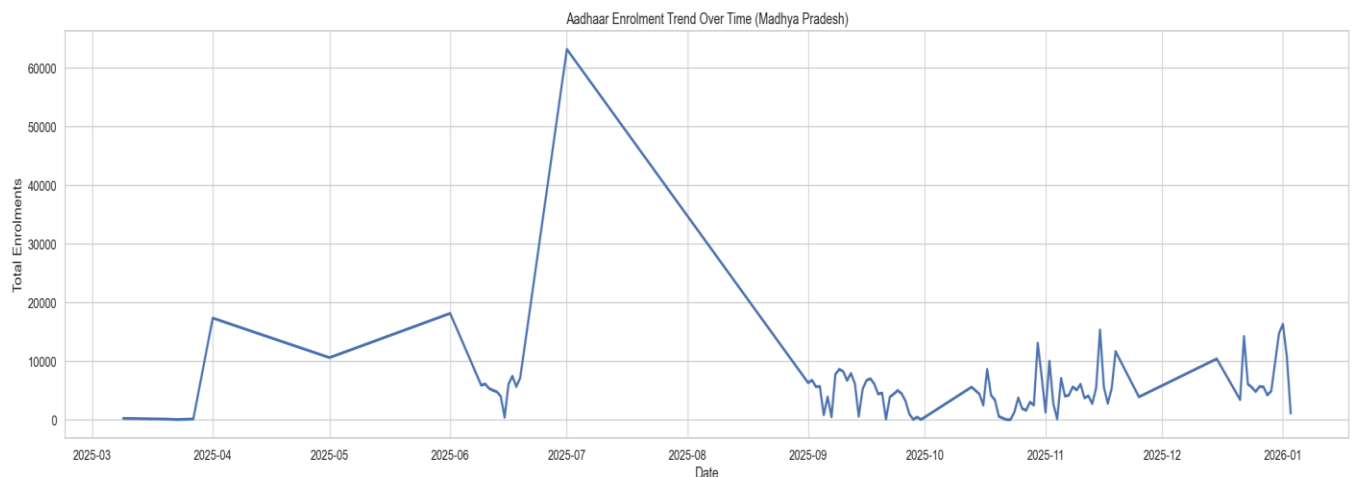
Interpretation : The Aadhaar enrolment data is **heavily dominated by the 0-5 age group**, which accounts for more than seventy percent of total enrolments. Enrolments in the 5-17 age group form a smaller share, while **adult enrolments are minimal**. This suggests that recent Aadhaar activity is primarily driven by **early-age registrations** rather than adult enrolments.

3. District-wise Age-group Distribution of Aadhaar Enrolments



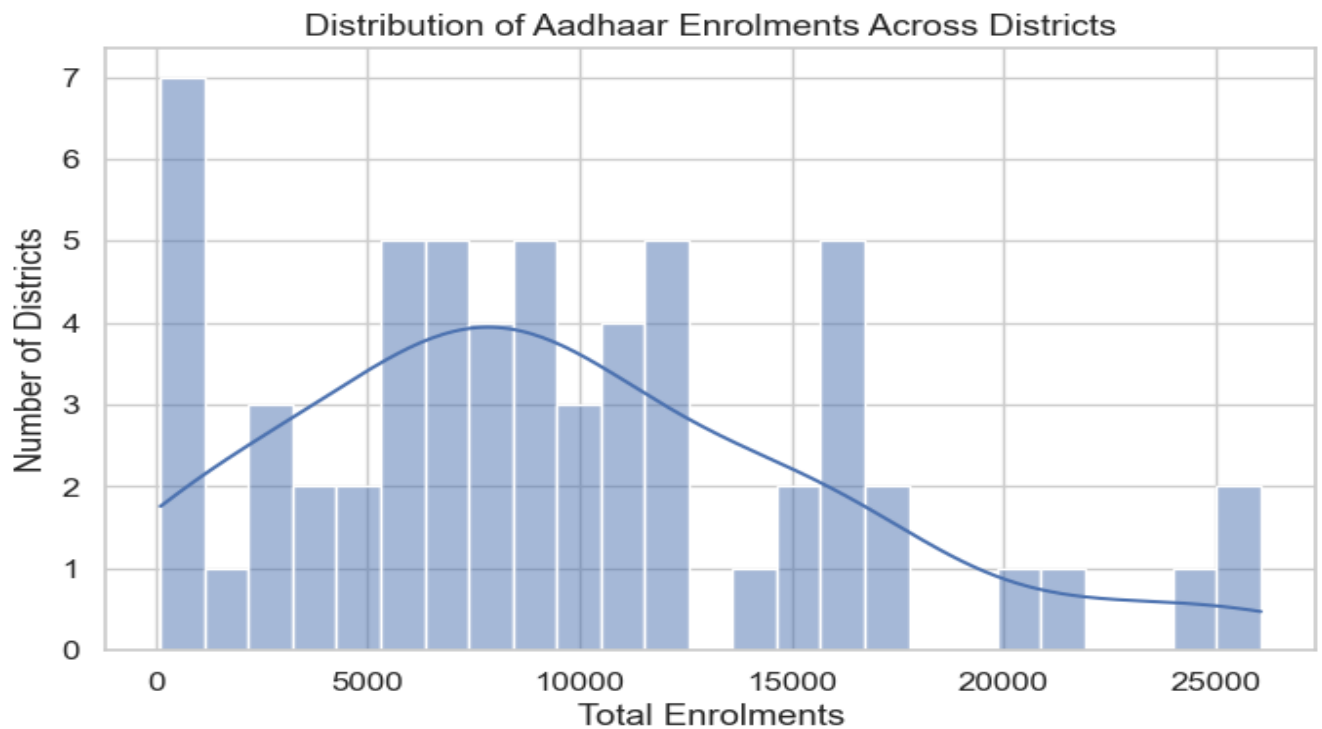
Interpretation : The district-wise age-group analysis shows that Aadhaar enrolments are dominated by the **0–5 age group across all districts**. Districts such as **Indore and Bhopal** record high enrolments mainly due to **strong child registration**, while enrolments in the **18 plus age group remain very low in every district**. This indicates that recent Aadhaar activity is largely driven by child enrolment initiatives rather than adult registrations.

4. Time-based Trend of Aadhaar Enrolments in Madhya Pradesh



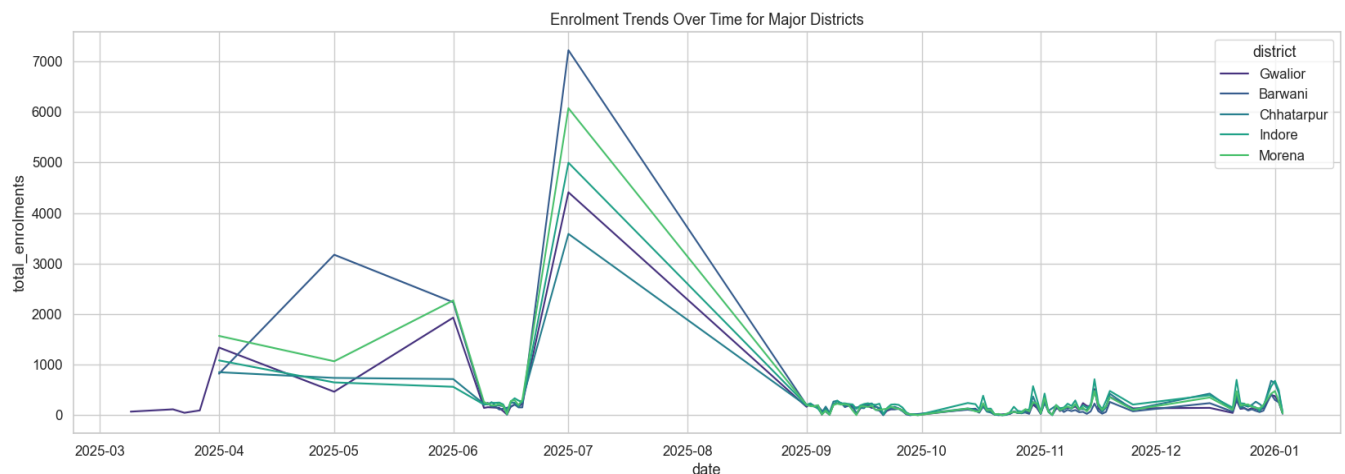
Interpretation : Aadhaar enrolments **do not follow a steady pattern over time**. A sharp spike is visible around mid-year, indicating a major enrolment drive or time-bound initiative. After this peak, enrolments decline and continue with smaller fluctuations, suggesting that Aadhaar activity is **largely driven by specific campaigns rather than continuous enrolment**.

5. Distribution of Aadhaar Enrolments Across Districts



Interpretation : The distribution is **skewed towards lower enrolment values**, meaning most districts have **moderate Aadhaar activity**. Only a small number of districts record very high enrolments, showing that Aadhaar-related services are **concentrated in select regions**.

6. Enrolment Trends Over Time for Major Districts (time × district × total enrolments)

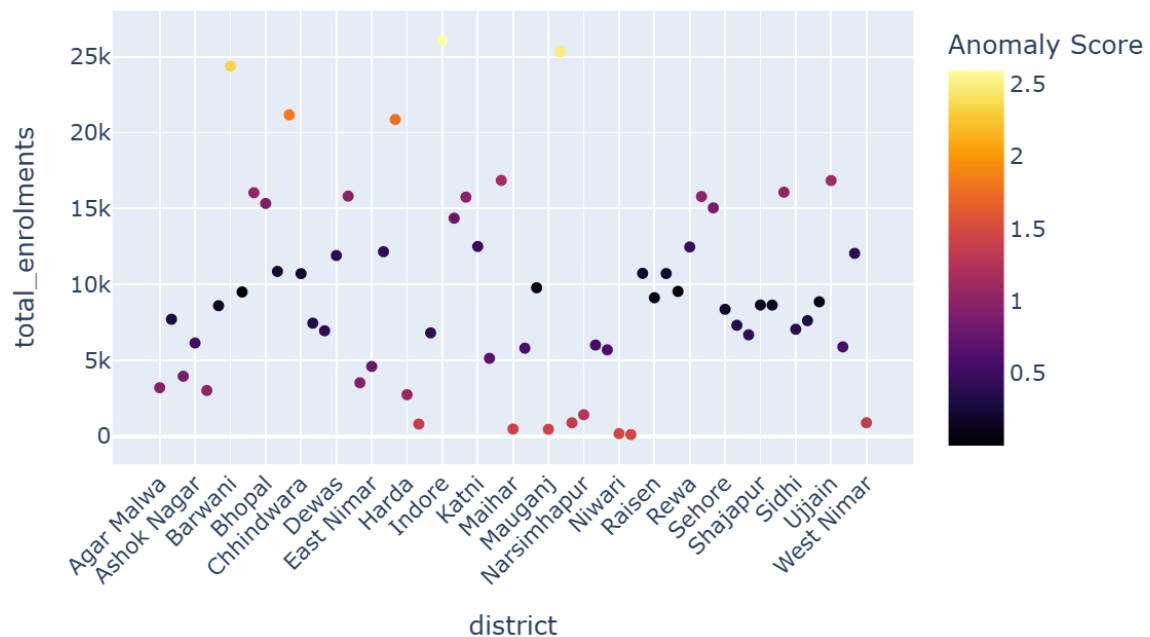


Interpretation : The enrolment trends for major districts show a **strong time-based pattern**. All districts experience a sharp and simultaneous spike in enrolments around mid-year, indicating a large-scale enrolment drive affecting the entire state. **Indore and Morena** record higher peaks compared to other districts, suggesting stronger participation or capacity during this period. After the peak, enrolments drop sharply and remain low, with only small fluctuations in later months.

Anomaly Detection :

7. District-level Aadhaar Enrolment Anomalies (district × total enrolments × anomaly score)

District-level Aadhaar Enrolment Anomalies



Interpretation : The anomaly analysis highlights districts with **unusual Aadhaar enrolment behaviour**. While districts such as **Indore and Bhopal** show high enrolments as expected, some districts exhibit very low enrolments with high anomaly scores. These deviations suggest **irregular enrolment patterns** that may be linked to localized enrolment drives, data reporting variations, or **operational challenges and require closer administrative attention**.

Recommendations

- Prioritise Aadhaar outreach in districts with consistently low enrolments by deploying **mobile units and organising village-level enrolment camps** to improve access.
- Since recent enrolments are largely driven by the 0–5 age group, integrate Aadhaar registration more closely with anganwadi centres, primary health facilities and birth registration systems to ensure early and complete coverage.
- Use **historical enrolment trends** to plan Aadhaar drives in advance, ensuring adequate staffing and technical resources during peak periods to avoid system overload.

- Closely monitor districts that show abnormal enrolment behaviour or sudden spikes, as these may indicate reporting gaps, temporary drives or operational challenges that need targeted administrative attention.
- Develop **district-specific enrolment strategies** based on **local enrolment patterns** rather than applying a uniform approach across the state, allowing resources to be allocated where they are most needed.

These recommendations aim to support ground-level implementation, improve enrolment efficiency and strengthen Aadhaar coverage across all districts.

Future Scope

- **Machine learning** models can be used to forecast district-wise Aadhaar enrolment demand using historical trends, helping plan centres, staff and infrastructure in advance.
- **District clustering** can be applied to group regions with similar enrolment behaviour, enabling tailored and more effective enrolment strategies.
- **Generative AI** can be used to automatically generate simple policy summaries and district-level insights from analytical outputs for decision-makers.

Conclusion

This project demonstrates how Aadhaar enrolment data can be transformed into clear and actionable insights through structured analysis and visualization.

The **district-level study of Madhya Pradesh** highlights uneven enrolment distribution, strong age-group imbalances and time-bound enrolment behaviour driven by targeted initiatives. These findings underline the importance of district-specific planning and focused outreach to ensure inclusive Aadhaar coverage.

The insights and recommendations presented can support data-driven decision-making and improve the effectiveness of Aadhaar enrolment strategies at the **ground level**.
