Aadhaar Enrolment Data Analysis Report

1. Introduction

Aadhaar is India's largest digital identity system, covering a vast and diverse population. Efficient enrollment operations are critical to ensure inclusion, accessibility, and optimal use of infrastructure.
This project performs a **data-driven analysis of Aadhaar enrolment patterns** to uncover operational insights, detect early signs of saturation, identify infrastructure bottlenecks, and support policy-level decision making.

---

2. Dataset Overview

- **Total records:** 1,006,029 rows
- **Columns used:**
  `date, state, district, pincode, age_0_5, age_5_17, age_18_greater, total_aadhaar_enrolment`
- **Unique states & UTs:** 36
- **Unique PIN codes:** 19,462
- **Total Aadhaar enrolment (dataset period): 5,435,484**
- **Average enrolment per day: 59,081**

This dataset provides **time-based, geographic, and demographic** enrolment information across India.

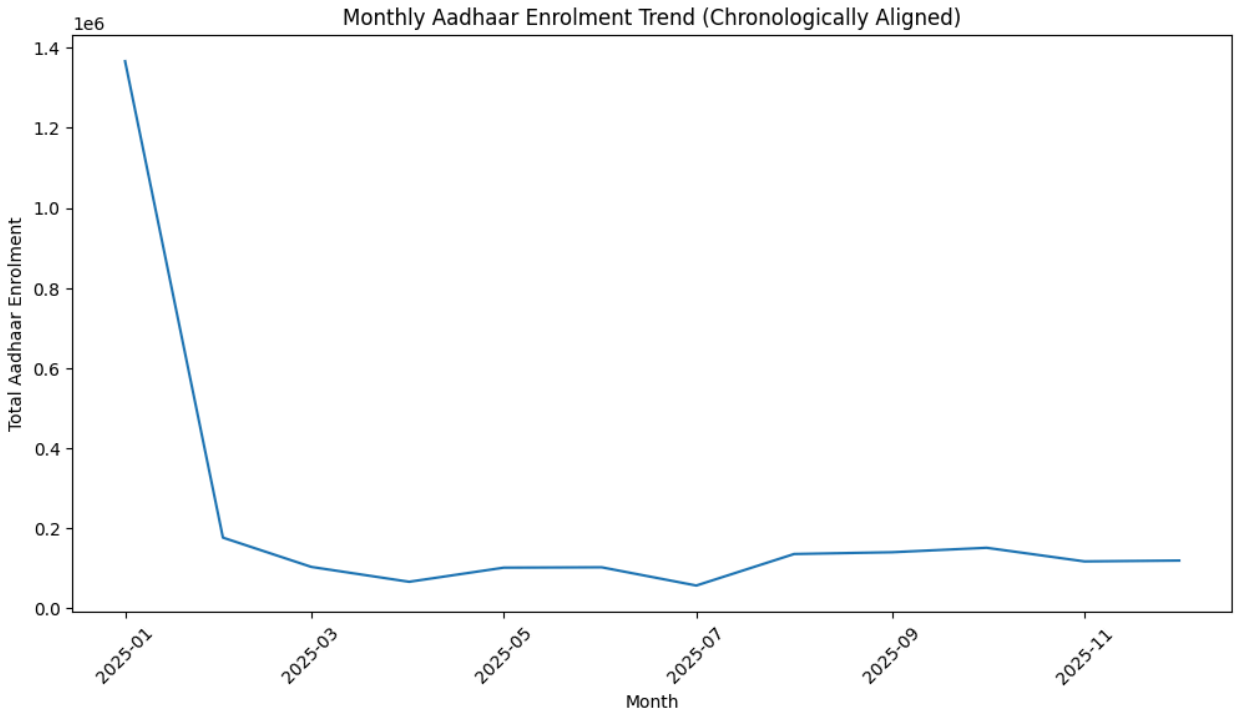
---

3. Overall Enrolment Patterns

3.1 Daily Enrolment Extremes

- **Highest enrolment day:**
  01-07-2025 → **616,868 enrolments**
- **Lowest enrolment day:**
  29-03-2025 → **67 enrolments**

This wide range indicates **event-driven spikes**, likely due to special enrolment drives or system-level interventions.



---

4. Age Group Analysis

4.1 Total Enrolment by Age Group

- **Age 0–5:** 3,546,965
- **Age 5–17:** 1,720,383
- **Age 18+:** 168,136

4.2 Percentage Contribution

- **Children (0–5): 65.26%**
- **Children (5–17): 31.65%**
- **Adults (18+): 3.09%**

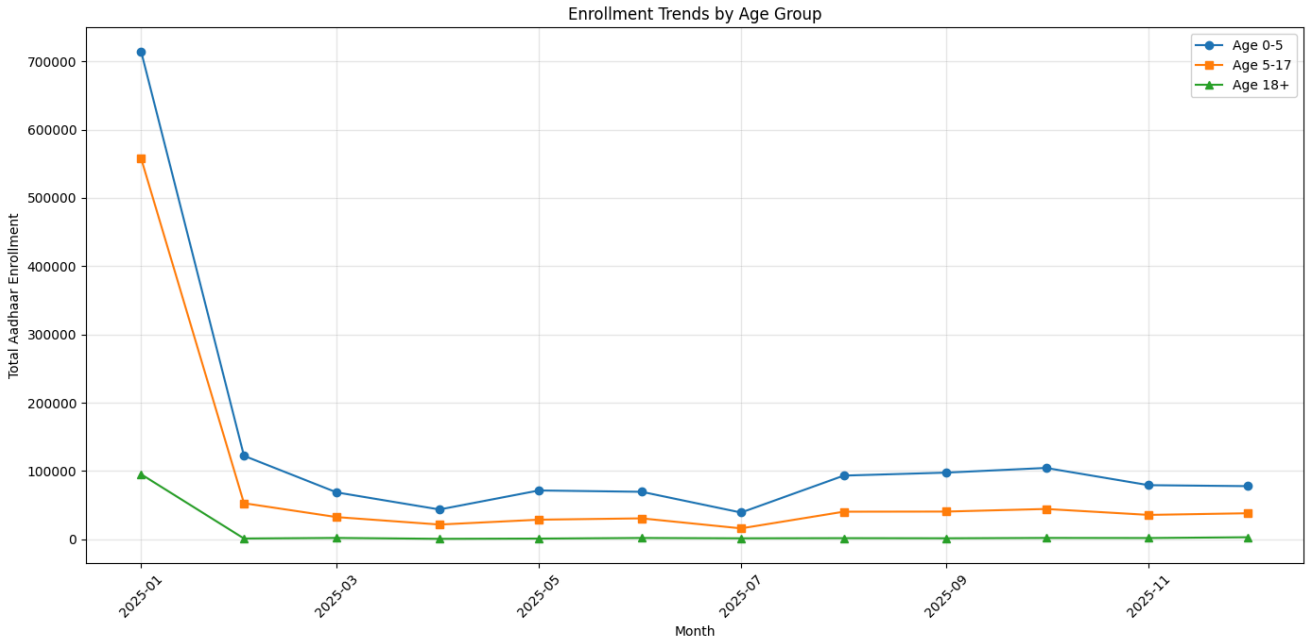
4.3 Key Insight

Child enrolment is **significantly higher than adult enrolment**, indicating:

- Aadhaar coverage among adults is already high

- New enrolments are driven mainly by **birth registrations and school-age children**

This strongly suggests **system maturity and early saturation** at the adult level.



---

5. State-Level Analysis

5.1 Highest and Lowest Enrolment States

- **Highest:** Uttar Pradesh → **1,018,629**
- **Lowest:** Lakshadweep → **203**

Large population states dominate enrollment volumes, while small UTs naturally show limited numbers.

---

6. District-Level Analysis

6.1 Highest & Lowest Enrolment Districts

- **Highest:** Thane (Maharashtra) → **43,688**
- **Lowest:** Nicobars (Andaman & Nicobar) → **1**

6.2 District Dominance Within States

Several states/UTs show **extreme district concentration**, for example:

- Lakshadweep → 100% enrolment from one district
- Chandigarh → ~99.9% from one district
- Goa, Puducherry, Ladakh → >40–80% from a single district

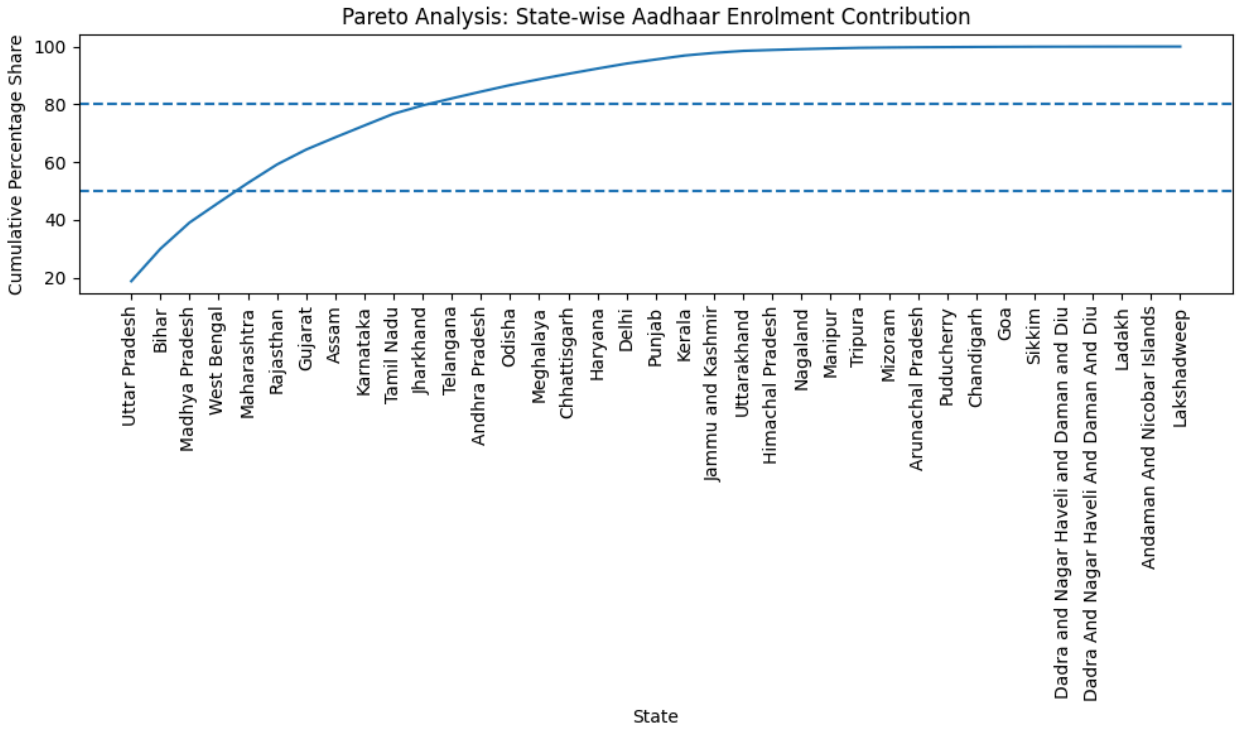
This highlights **centralized enrollment infrastructure** and limited geographic spread.

---

7. Concentration & Distribution Analysis

- **Top 10 districts contribute: 6.44%** of total enrolment
- **~50% enrolment comes from:** 148 districts
- **~80% enrolment comes from:** 359 districts

👉 Enrolment is **moderately concentrated**, not dominated by a handful of districts, but still unevenly distributed.



---

8. Temporal Anomalies & Spikes

8.1 Unusual Days

- Multiple **high-enrolment spikes** in early January
- No zero-enrolment days detected

8.2 State-Level Sudden Spikes

- 556 instances of **>50% day-to-day change**
- Observed especially in smaller states and UTs

These spikes indicate:

- Limited infrastructure capacity
- Batch processing
- Campaign-driven enrolment days

---

9. Infrastructure Limitation Indicators

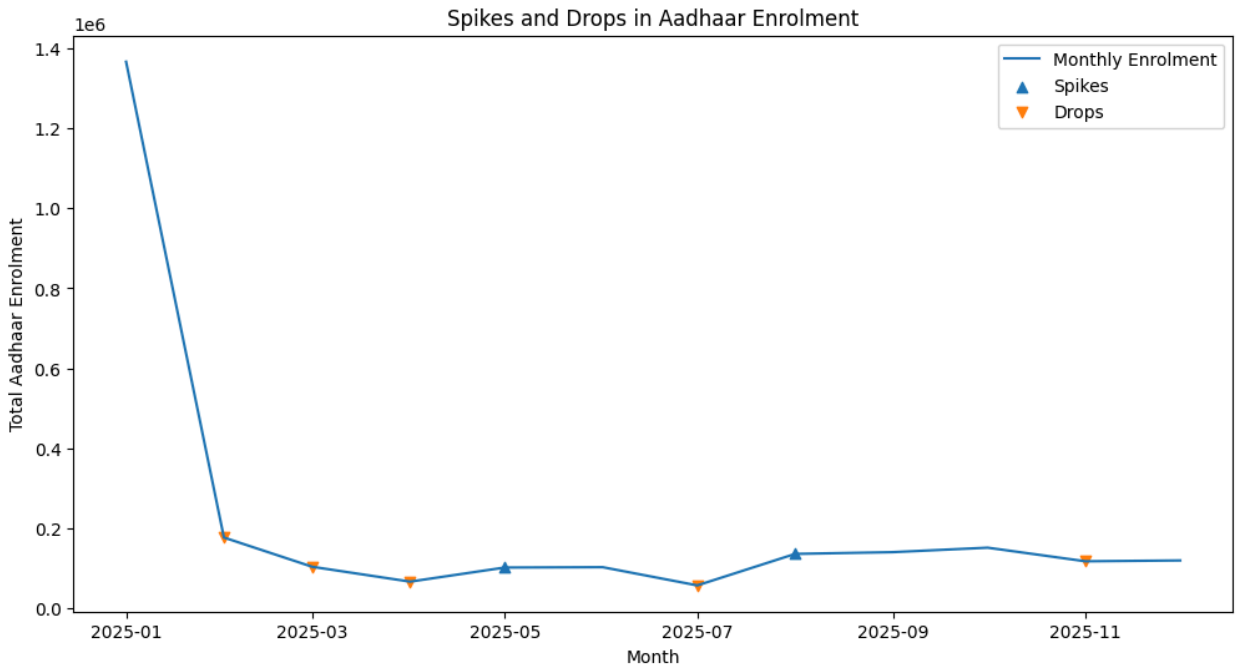
9.1 District Spike Ratio (Max / Mean)

Top stressed districts:

- Barabanki (21.48)
- Sabar Kantha (18.59)
- Ahmedabad (17.81)
- Thiruvallur (17.68)

A **spike ratio > 15** clearly indicates:

- Enrolment centers overwhelmed on certain days
- Insufficient machines or manpower
- Poor load balancing



---

10. Low-Enrolment Regions

10.1 Consistently Low Districts

- 245 districts show **very low average daily enrolment**
- Many districts average **1–5 enrolments/day**

10.2 States With Many Low-Enrolment Districts

Top examples:

- Arunachal Pradesh (22)
- Andhra Pradesh (20)
- Karnataka, Chhattisgarh (14 each)
- West Bengal (13)

This suggests **accessibility issues**, awareness gaps, or geographic challenges.

---

11. Child Enrolment Variability

- **Mean:** 5,225

- **Std Dev:** 6,357
- **Coefficient of Variation: 1.22**

High variability shows:

- Uneven outreach
- Some fdistricts enrolling many children, others almost none

---

12. Child vs Adult Enrolment Ratio

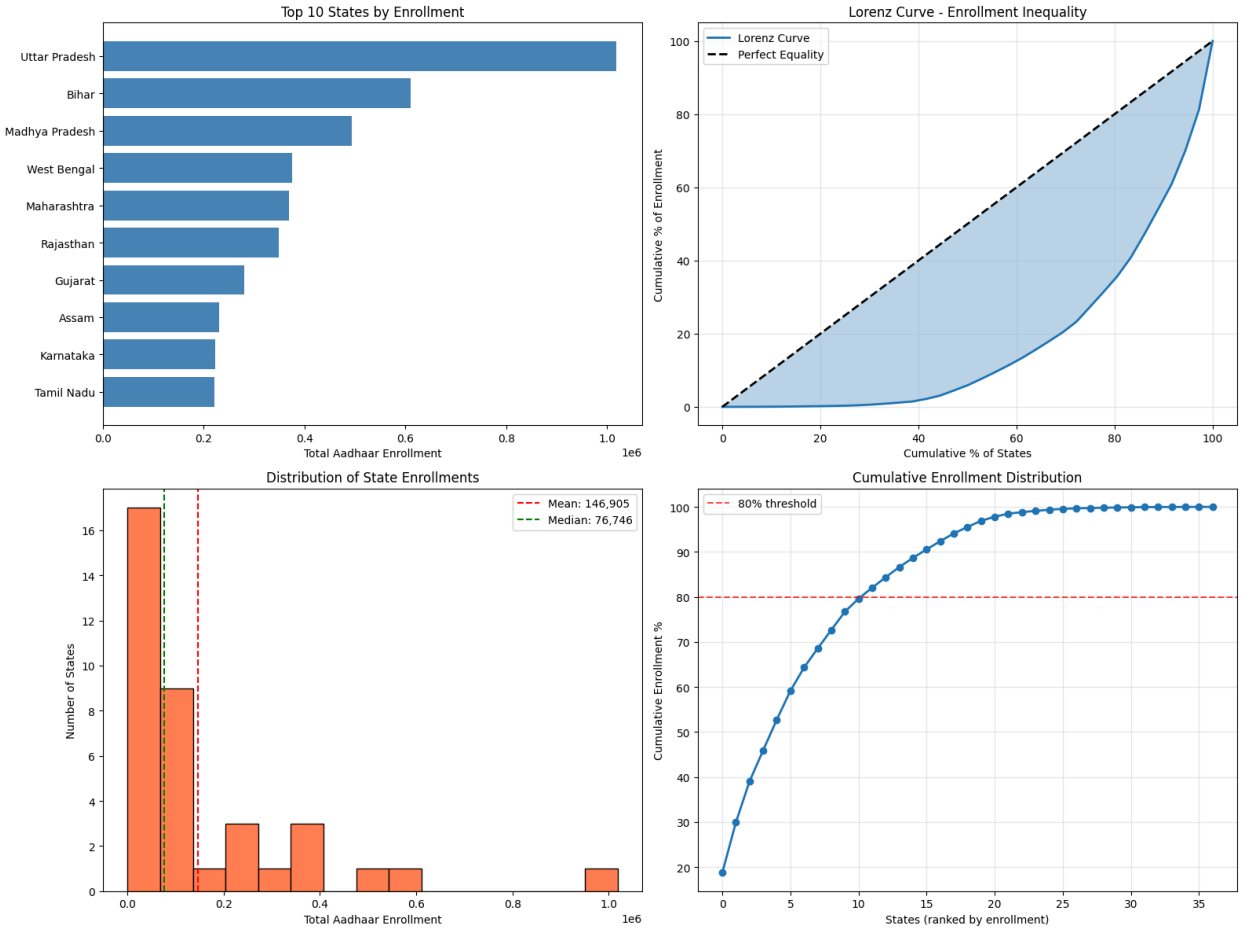
Several regions show **extremely high child-to-adult ratios**, for example:

- Andaman & Nicobar → ∞
- Dadra & Nagar Haveli → 226
- Lakshadweep → 202
- Tamil Nadu → 175

This confirms:

- Adult enrolment saturation
- Future adult enrolment demand will be limited

- Infrastructure should focus on **c**



hild enrolment workflows

-

---

13. Predictive Analysis Result

The predictive model achieved:

- **Mean Absolute Error (MAE): 46.84**

Interpretation:

This low error proves that **child enrolment trends strongly predict future adult enrolment**, making child enrolment a **reliable early indicator** for short-term forecasting.

---

14. Operational Insights

14.1 Weekday vs Weekend

- **Weekend enrolment > Weekday enrolment**

➡️ Weekend-only campaigns should be **expanded and prioritized**.

14.2 State Consistency

States like Manipur, Punjab, Maharashtra, Uttar Pradesh show **very high variability**, indicating operational instability and uneven infrastructure deployment.

---

15. Policy Recommendations

1. **Expand weekend enrolment drives**
2. **Decentralize enrolment centers** in district-dominated states
3. **Add infrastructure** in high spike-ratio districts
4. **Target low-enrolment districts** with awareness campaigns
5. **Use child enrolment as a leading indicator** for planning capacity
6. Shift focus from adult enrolment to **newborn and child-focused workflows**

---

16. Conclusion

This project demonstrates how large-scale administrative data can be transformed into actionable operational and policy insights. The analysis reveals early saturation signals, regional inequities, infrastructure limitations, and strong predictive relationships between child and adult enrolment. These findings can directly support better planning, equitable access, and efficient execution of Aadhaar enrolment programs across India.

Github Repo : https://github.com/bappaditya-paul/Data-Analysis-Adhar-Enrollment