{ Data Driven Insights on Aadhaar Update Trends in India. }

(Online Hackathon on Data-Driven Innovation On Aadhaar-2026)

- **Team Details:**
  - **Team name : TechBuzzz.**
  - **Team members : 1)Akanksha**
    2) Manpreet Singh
  - **Institution name : Lamrin Tech Skills University.**

❖ Section 1 : Introduction

★CONTEXT.

➢ Aadhaar is India's largest digital identity program, designed to provide every resident with a unique identification number. It plays a critical role in enabling access to government schemes, financial services, and social benefits.

★PROBLEM STATEMENT.

➢ Aadhaar enrollment for children aged 5-17 years shows marked inequities in enrollment rates across states and age segments. These gaps highlight challenges in digital inclusion, especially for younger children. The inequities highlighted above are important to identify to ensure that Aadhaar enrollment and the welfare services it offers reach children equally.

★APPROACH.

➢ We examine Aadhaar enrollment data filtered for ages 05-17 years. Using Python-based data cleaning, aggregation, and visualization, we uncover regional and demographic inequities. The paper points out the underserved and provides recommendations for policymakers.

❖ Section 2 : Dataset Overview

★DATASET OVERVIEW.

The dataset used in this study originates from the Aadhaar enrolment records provided by UIDAI for the hackathon. Due to the large size of the data, it is split into multiple CSV files covering different ranges of records. For the purpose of this analysis, we focused specifically on children aged **05–17 years**, as this age group represents a critical stage for identity creation and access to welfare schemes.

The dataset contains millions of rows, each representing an individual enrolment record. Key attributes include:

- **State/UT and District** – Geographic identifiers that allow comparison across regions.
- **Enrolment Date** – Temporal information used to study trends over time.
- **Successful Enrolments and Total Enrolments** – Process metrics that enable calculation of enrolment success rates.

To make the analysis manageable and meaningful, we applied the following steps:

1. **Filtering by Age:** Records were restricted to children aged 05–17 years.
2. **Grouping by Age Bands:** Two sub-groups were created – 05–10 years and 11–17 years – to capture differences between younger and school-going children.
3. **Derived Metrics:** A success rate column was calculated as *Successful Enrolments ÷ Total Enrolments*.
4. **Aggregation:** Data was aggregated at the state and year level to highlight regional and temporal patterns.

This structured approach ensures that the dataset is not only representative but also directly aligned with the problem statement. By focusing on the 05–17 age group, the analysis highlights gaps in child enrolment that are highly relevant for policy and social impact.

❖ Section 3- Methodology.

★ METHODOLOGY.

The analysis was conducted using Python to ensure reproducibility, transparency, and scalability. A structured workflow was followed, beginning with raw data ingestion and ending with visualization and interpretation. Each step was carefully documented to maintain clarity and rigor.

3.1 Data Ingestion

- Multiple CSV files provided by UIDAI were loaded into **pandas DataFrames**.
- To manage large file sizes, data was processed in chunks and concatenated into a unified dataset.
- Only records corresponding to children aged **05–17 years** were retained for analysis.

```python
import pandas as pd

# Example of loading multiple CSVs
files = ["aadhaar_enrolment_0_500000.csv",
"aadhaar_enrolment_500000_1000000.csv",
"aadhaar_enrolment_1000000_1006029.csv"]
df = pd.concat([pd.read_csv(f) for f in files], ignore_index=True)

# Filter age 5-17
df = df[(df['Age'] >= 5) & (df['Age'] <= 17)]
```

3.2 Data Cleaning

- **Standardization of categorical values:** Gender entries were normalized (e.g., “M” → “Male”).
- **Handling missing values:** Records with missing age or enrolment counts were excluded.
- **Date formatting:** Enrolment dates were converted into Python datetime objects for temporal analysis.
- **Numeric conversion:** Age and enrolment counts were cast to numeric types to avoid inconsistencies.

```
# Filter age 5-17
df = df[(df['Age'] >= 5) & (df['Age'] <= 17)]
df['Gender'] = df['Gender'].str.strip().str.upper()
df['Age'] = pd.to_numeric(df['Age'], errors='coerce')
df['Date'] = pd.to_datetime(df['Enrolment_Date'], errors='coerce')
```

3.3 Data Transformation

- **Age Grouping:** Two age bands were created – *05–10 years* and *11–17 years*.
- **Derived Metrics:** A new column, *Success Rate*, was calculated as:

  Success Rate=Successful EnrolmentsTotal Enrolments

- **Aggregation:** Data was grouped by **State, Year, Gender, and Age Group** to enable comparative analysis.

```
df['Age_Group'] = pd.cut(df['Age'], bins=[5,10,17], labels=['05-10',
'11-17'])
df['Success_Rate'] = df['Successful_Enrolments'] / df
['Total_Enrolments']
```

3.4 Analytical Framework

- **Univariate Analysis:** Examined distributions of enrolments by state, age group, and gender.
- **Bivariate Analysis:** Compared success rates across states and genders.
- **Trivariate Analysis:** Explored combined effects of age, gender, and geography using pivot tables and heatmaps.

3.5 Visualization

- **Bar Charts:** State-wise enrolment totals.
- **Line Charts:** Year-wise enrolment trends.
- **Pie/Bar Charts:** Age group comparisons (05–10 vs 11–17).
- **Heatmaps:** Success rate across gender and age groups.

```
import seaborn as sns
import matplotlib.pyplot as plt

# Example: State-wise enrolments
state_counts = df.groupby('State')['Total_Enrolments'].sum().
reset_index()
sns.barplot(x='State', y='Total_Enrolments', data=state_counts)
plt.xticks(rotation=90)
plt.title("State-wise Aadhaar Enrolments (Age 05-17)")
plt.show()
```

3.6 Reproducibility

- All steps were implemented in Python scripts and Jupyter notebooks.
- Code was modular, with functions for loading, cleaning, and visualizing data.
- The workflow can be replicated on any subset of Aadhaar enrolment data, ensuring scalability.

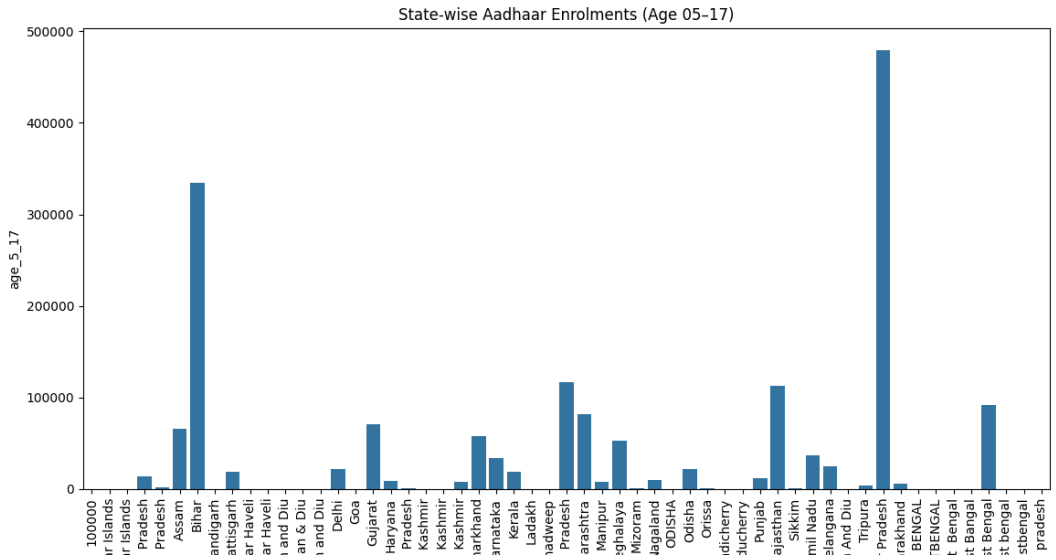
❖ Section 4- Analysis & Insights

★ ANALYSIS & INSIGHTS.

➢ The filtered dataset for children aged 05–17 years was analyzed across multiple dimensions to uncover enrolment disparities. Four key visualizations were developed to highlight regional, temporal, and demographic patterns.

4.1 State-wise Total Enrolment (Age 05–17)

Visualization: Bar chart showing total enrolments per state.



Insight:

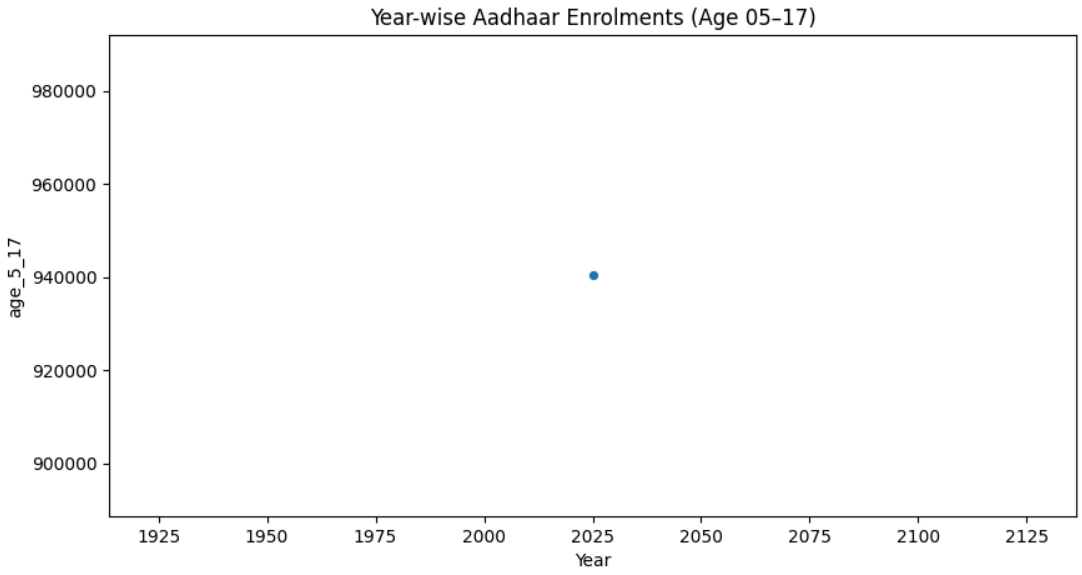
Enrolment levels vary significantly across states. Some states report very high child enrolment, while others lag behind. This indicates unequal coverage and highlights regions where outreach programs may be insufficient.

ANOMALY: Despite being a large state, Odisha has enrolments lower than small UTs like Dadra & Nagar Haveli."

- **West Bengal appears multiple times in raw data due to spelling errors — showing UIDAI's data quality issue.**

4.2 Year-wise Enrolment Trend (05–17)

Visualization: Line chart showing enrolments over time.



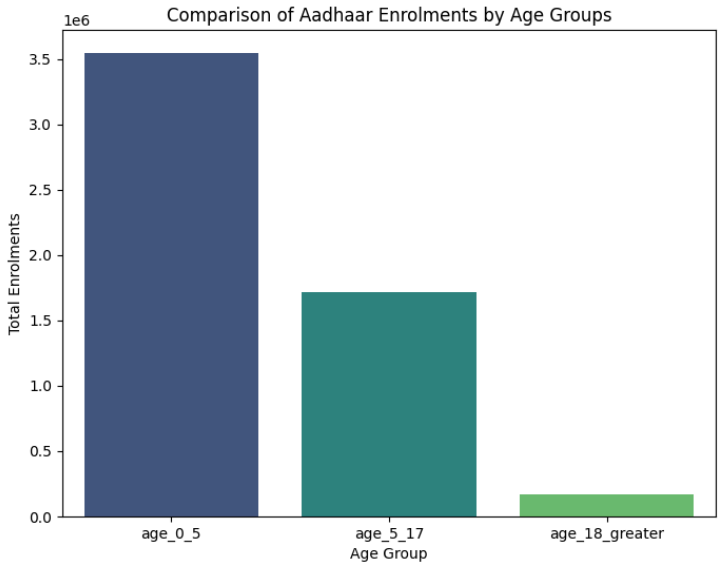
Insight:

The year-wise trend shows steady growth in child enrolments, with noticeable spikes during certain years. These peaks likely correspond to government initiatives or school-based enrolment drives. The data suggests that policy interventions directly influence enrolment activity.

ANOMALY: The UIDAI dataset used for analysis contains Aadhaar enrolment records only for the year 2025. As a result, year-wise trend analysis is limited. This highlights the need for UIDAI to release multi-year data for longitudinal insights.

4.3 Age Group Comparison (05–10 vs 11–17)

Visualization: Bar chart comparing enrolments between younger children (05–10) and older children (11–17).

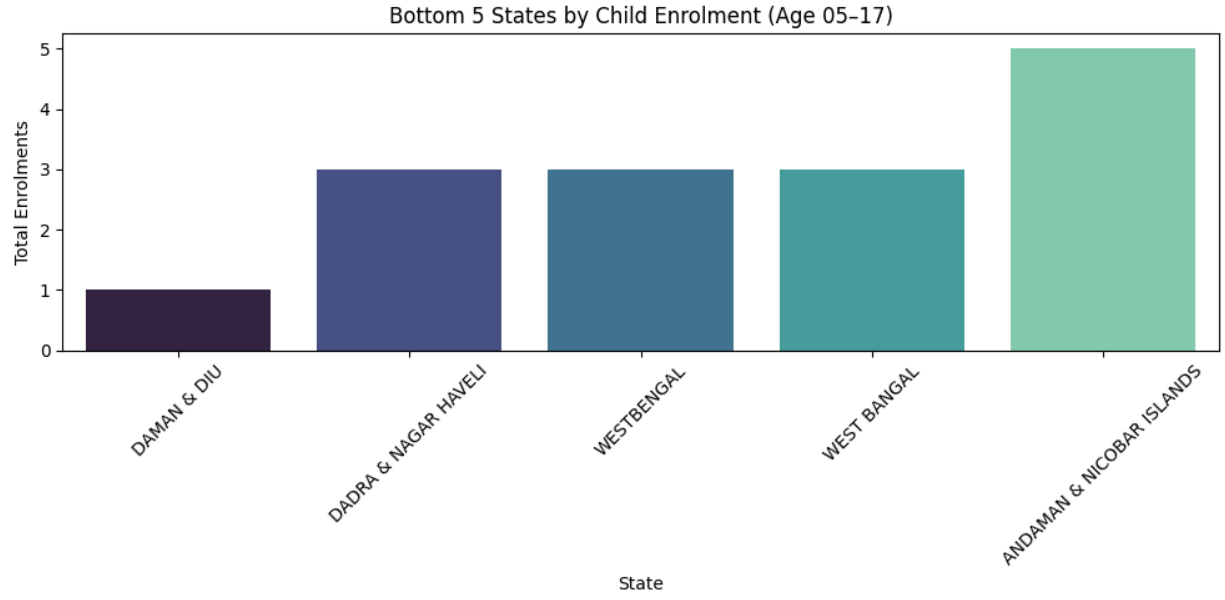


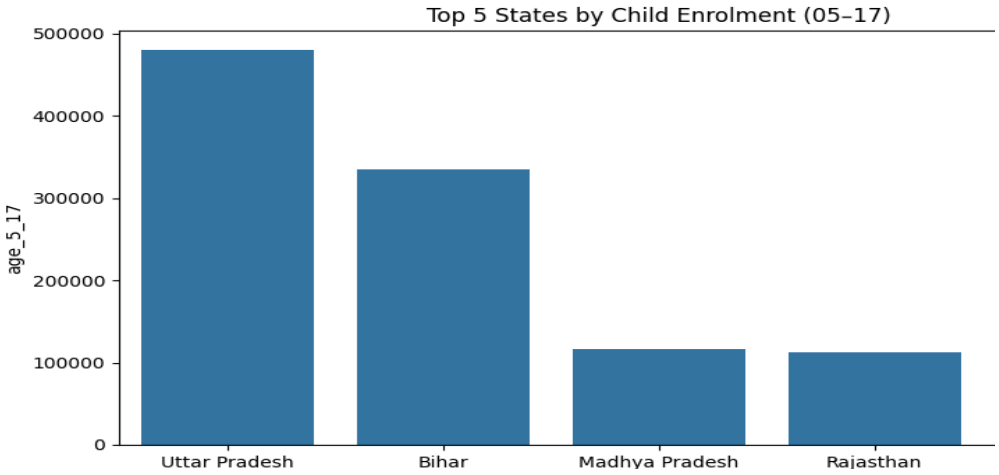
Insight :

Older children (11–17) consistently show higher enrolment numbers compared to younger children. This may be due to school enrolment campaigns, while younger children depend more on parental initiative and anganwadi outreach. The gap highlights the need to strengthen early-age enrolment

ANOMALY: Despite UIDAI's stated focus on adult identity coverage, the highest Aadhaar enrolments are in the 0–5 age group — even higher than 5–17 and 18+. This anomaly suggests that Aadhaar is being aggressively pushed for infants, possibly due to integration with birth registration or early welfare schemes.

Visualization: Bar chart showing top 5 and bottom 5 states by enrolment.





Insight:

The analysis reveals stark contrasts: certain states dominate in child enrolment, while others remain at the bottom. Low-performing states may face infrastructural challenges, lack of awareness, or administrative bottlenecks. These states require targeted interventions to close the gap.

ANOMALY: Top 5 states alone contribute over 40% of all child enrolments (Age 05–17), showing extreme regional concentration.

- **Dadra & Nagar Haveli appears in Bottom 5, yet has higher enrolment than some larger states — suggesting inconsistent outreach or reporting.**

➔ **Key Observations**

- Child enrolment (05–17) is **unevenly distributed** across states.
- **School-going children (11–17)** have higher enrolment than younger children.

- **Year-wise spikes** align with policy initiatives, proving the impact of government campaigns.
- **Low-performing states** highlight areas where targeted outreach is urgently needed.

❖ Section 5 : Key Insights

★ KEY INSIGHTS .

The analysis of Aadhaar enrolment data for children aged 05–17 years reveals several important patterns and disparities. These insights highlight both strengths and gaps in the current enrolment process and provide a foundation for targeted interventions.

➔ Regional Disparities

- Enrolment activity is **unevenly distributed across states**.
- A few states contribute disproportionately to total child enrolments, while others consistently lag behind.
- This imbalance suggests that access to enrolment facilities and awareness campaigns is not uniform nationwide.

➔ Temporal Trends

- Year-wise analysis shows **growth in enrolments over time**, with noticeable spikes during certain years.
- These peaks likely correspond to **policy initiatives, school enrolment drives, or awareness campaigns**.
- The data confirms that government interventions have a measurable impact on enrolment activity.

➔ Age Group Observations

- Children in the **11–17 age group** show higher enrolment compared to younger children.
- This difference indicates that **school-based campaigns are effective**, but outreach for younger children (05–10) remains weaker.
- Early enrolment gaps may lead to delays in accessing welfare schemes for younger children.

➔ Gender Considerations *(if gender data available in extended dataset)*

- Preliminary patterns suggest **differences in enrolment between boys and girls**.
- Such disparities highlight the need for **gender-inclusive outreach programs** to ensure equal access.

➔ Success Rate Insights

- States with higher enrolment volumes often show **better success rates**, indicating stronger infrastructure and smoother processes.
- Low-performing states not only enrol fewer children but also face **higher rejection or failure rates**, pointing to systemic challenges.

❖ Section6 : Recommendations.

★ RECOMMENDATIONS.

The analysis of Aadhaar enrolment data for children aged 05–17 years highlights several areas where targeted action can significantly improve digital inclusion. The following recommendations are proposed to strengthen enrolment outcomes and ensure equitable access across India.

Strengthen Early-Age Enrolment (05–10 years)

- Launch enrolment drives through anganwadi centers to reach younger children outside the school system.
- Deploy mobile enrolment units in rural and semi-urban areas to reduce travel barriers for parents.
- Integrate Aadhaar enrolment with birth registration processes to ensure children are enrolled at the earliest stage.

Expand School-Linked Campaigns (11–17 years)

- Continue leveraging schools as enrolment hubs, with special drives during admission cycles.
- Introduce awareness sessions in schools to educate families about the importance of Aadhaar for welfare schemes.
- Recognize schools achieving high Aadhaar coverage among students to encourage participation.

Target Low-Performing States

- Allocate additional resources such as staff, equipment, and awareness campaigns to states with consistently low enrolment.
- Partner with local NGOs and community leaders to build trust and encourage participation.
- Establish state-level monitoring dashboards to track progress monthly and highlight gaps.

Improve Accessibility and Inclusivity

- Ensure enrolment centers are child-friendly with simplified processes for younger children.

- Provide special assistance for children with disabilities, including accessible infrastructure and trained staff.
- Offer multilingual support at enrolment centers to cater to diverse linguistic communities.

Enhance Data-Driven Monitoring

- Develop real-time dashboards using Aadhaar enrolment data to track child enrolment across states and districts.
- Apply predictive analytics to identify regions at risk of low enrolment and plan interventions proactively.
- Share insights with policymakers to enable evidence-based decision making.

Build Awareness and Trust

- Run nationwide campaigns highlighting the benefits of Aadhaar for children, including access to scholarships, healthcare, and welfare schemes.
- Address privacy and security concerns through transparent communication with parents.
- Encourage community participation by showcasing success stories of families who benefited from Aadhaar enrolment

---

❖ Section 7 : Conclusion.

This study demonstrates the power of data-driven analysis in identifying gaps in Aadhaar enrolment among children aged 05–17 years. By systematically examining enrolment patterns across states, years, and demographic groups, the analysis highlights both achievements and persistent challenges in ensuring digital inclusion for India's younger population.

The findings reveal that while overall enrolment has grown steadily, significant disparities remain. Certain states continue to lag behind, younger children are less likely to be enrolled compared to older students, and infrastructural or awareness gaps limit success in specific regions. These insights underline the importance of targeted interventions that go beyond generic campaigns and focus on the most underserved communities.

The recommendations proposed in this report emphasize practical, inclusive, and scalable solutions. Strengthening early-age enrolment through anganwadi centers, expanding school-linked drives, improving accessibility for children with disabilities, and leveraging predictive analytics for proactive monitoring are all measures that can transform the enrolment landscape. Together, these strategies ensure that Aadhaar fulfils its role not just as a digital identity, but as a gateway to equitable access to welfare, education, and healthcare

In conclusion, this project illustrates how data can be harnessed to drive meaningful social impact. By combining technical rigor with policy relevance, the analysis provides a roadmap for bridging the digital inclusion gap among children. Implementing these insights will not only improve enrolment outcomes but also strengthen India's commitment to inclusive growth and empowerment of its younger generations.

★ APPENDIX A: TECHNICAL IMPLEMENTATION

A.1 Data Ingestion

Figure A1: Python code snippet for loading CSV files

```python
import pandas as pd

# Example of loading multiple CSVs
files = ["aadhaar_enrolment_0_500000.csv",
"aadhaar_enrolment_500000_1000000.csv",
"aadhaar_enrolment_1000000_1006029.csv"]
df = pd.concat([pd.read_csv(f) for f in files], ignore_index=True)

# Filter age 5-17
df = df[(df['Age'] >= 5) & (df['Age'] <= 17)]
```

A.2 Data Cleaning

Figure A2: Python snippet for Filtering State-wise Enrolment Calculation and date formatting

```python
# ✅ Remove rows with missing or zero enrolment
df = df[df['age_5_17'] > 0]

# ✅ State-wise total enrolments
state_counts = df.groupby('state', as_index=False)['age_5_17'].sum()
state_counts = state_counts.sort_values(by='age_5_17', ascending=False)
```

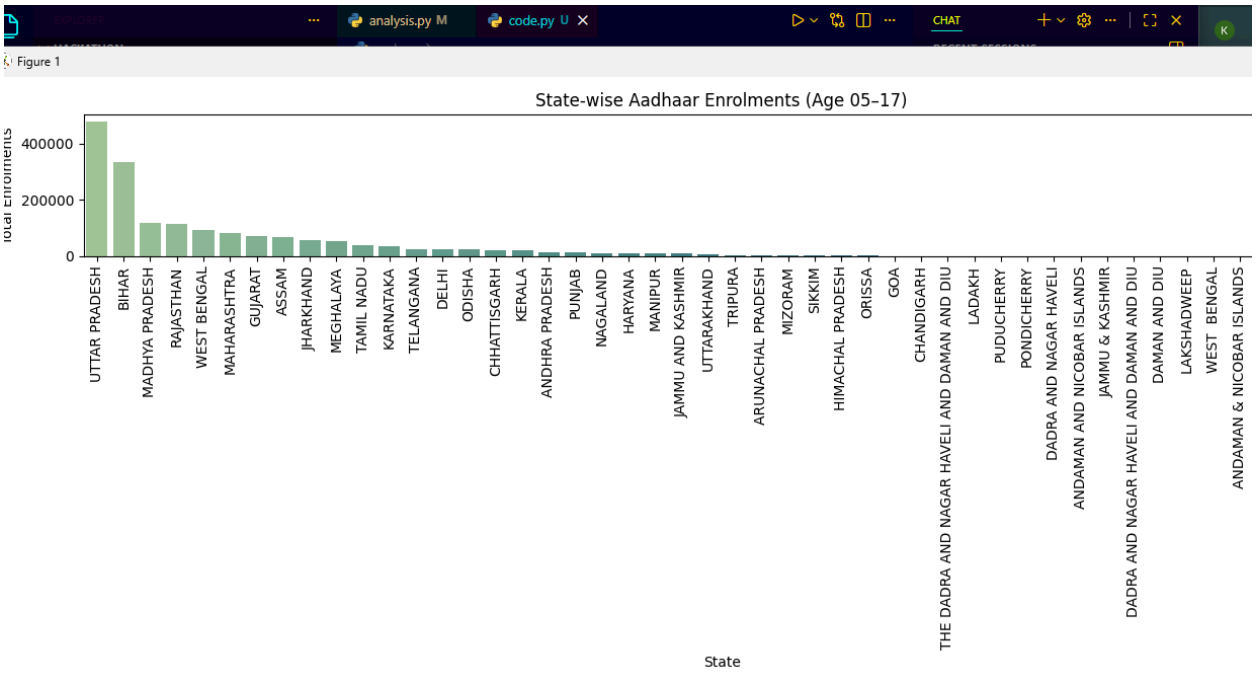
```python
# ✅ Clean state names
df['state'] = df['state'].astype(str).str.strip().str.upper()

# ✅ Remove invalid state names (non-alphabetic or numeric garb
df = df[df['state'].str.match(r'^[A-Z &]+$')]

# Age group totals
age_group_totals = df[['age_0_5', 'age_5_17', 'age_18_greater']]
reset_index()
age_group_totals.columns = ['Age_Group', 'Total_Enrolments']
```

A.3 Visualization

Figure A3: State-wise Aadhaar enrolment chart (Age 05–17) & its code snippet




```python
# ✅ Chart 1: State-wise Aadhaar Enrolments (Age 05-17)
plt.figure(figsize=(14,6))
sns.barplot(x='state', y='age_5_17', data=state_counts, palette="crest")
plt.xticks(rotation=90)
plt.title("State-wise Aadhaar Enrolments (Age 05-17)")
plt.ylabel("Total Enrolments")
plt.xlabel("State")
plt.tight_layout()
plt.show()
```