“UIDAI Aadhaar Enrolment Analysis: Spatial, Demographic & Temporal Insights”

- UIDAI Data Hackathon 2026

UIDAI_2260

19-01-2026

Tools Used: Python, Pandas, Matplotlib, Seaborn

Platform Used: Google Collab

Problem Statement and Approach:

Problem Statement:

India Aadhaar registration process is not homogeneous with regard to various regions, age categories and time. The enrolment levels will also differ because of variations of population, age distribution, administrative facilities and seasonal demand. There are those regions where enrolment is very active and those regions where the participation is low. Without an appropriate evidence-based view of such differences the resources including enrolment centers and staff can be distributed inefficiently. This might lead to the long queues, congested centers and unequal provision of services.

This analysis aims at analyzing the spatial, time-based, and demographic patterns within the Aadhaar enrolment data. The study will identify the significant trends in enrolment and enrolment hot spots by examining patterns in enrolment in states, districts, age, and months. These lessons can be used in order to plan better, distribute resources more effectively and efficiently, and make Aadhaar services more efficient and accessible.

Approach:

It used a systematic approach of data analysis by investigating the Aadhaar enrolment patterns, by region, age group, and time. The UIDAI Aadhaar database of enrolment was initially cleansed and tidied in order to establish data consistency and reliability. The analysis centered on the functionality of determining relevant patterns and trends that might support the enrolment activity variations as well as make practical decisions.

The principal procedures involved when conducting the analysis include:

- Cleaning and standardization of state, district and pincode data to eliminate inconsistencies.
- Dealing with invalid and duplicate recordings in order to enhance the quality of the data.
- Casting date information and being able to extract time information (month year).
- Carrying out univariate analysis to understand general trends of enrolments.
- The analysis of relationships between enrolment and the region and age group using the bivariate analysis.
- Trivariate analysis to determine the enrolment hotspots and seasonal trends.
- Presenting insights clearly using visualization tools like bar charts and stacked bars, line charts and heatmaps.

Datasets Used:

This discussion is based on the Aadhaar enrolment and update statistics of Unique Identification Authority- India (UIDAI). The data is daily enrolments in Aadhaar in various regions of the country. It encompasses data linked to place, age categories, and time which enable one to familiarize themselves with the difference in Aadhaar enrolment rates by regions and time.

The dataset can be analyzed at various levels including state, district and pincode. It also gives age specific details of enrolment and the data is helpful in the analysis of demographic trends in Aadhaar registration.

Important columns used in the analysis include:

- date — event date
- state — state name
- district — district name
- pincode — geographic pincode
- age_0_5 — enrolment counts for age 0–5
- age_5_17 — enrolment counts for age 5–17
- age_18_greater — enrolment counts for age 18+
- total enrollment — engineered feature equal to age_0_5 + age_5_17 + age_18_greater
- month / year — derived temporal fields for aggregation

Methodology:

Data Cleaning, Standardization, and Preprocessing

To ensure the accuracy and reliability of the analysis, several data cleaning and preprocessing steps were carried out. These steps focused on standardizing geographic fields, handling invalid placeholder values, preparing date-related attributes, and verifying data quality. Proper preprocessing was essential to avoid misleading patterns and to ensure meaningful spatial and temporal analysis.

The following standardization and preprocessing steps were applied:

- Performed case normalization by converting state and district names to a consistent format to avoid duplication caused by case differences.
- Applied whitespace trimming to remove leading and trailing spaces in state and district fields.
- Implemented alias mapping to correct commonly used alternative names and spelling variations (for example, Orissa → Odisha, WestBengal → West Bengal).

- Used the same normalization and alias mapping approach for district names to maintain consistency across geographic levels.
- Converted date strings into datetime format to enable accurate time-based analysis.
- Converted date strings into datetime format to enable accurate time-based analysis.
- Extracted month and year from the date field to support monthly and seasonal trend analysis.
- Created a derived feature, total_enrollment, by summing enrolments across all age groups (age_0_5 + age_5_17 + age_18_greater).

Handling of Placeholder and Invalid Values:

During data exploration, certain records contained placeholder values that did not represent valid geographic information. These records required careful handling to prevent distortion of spatial analysis results while still preserving overall enrolment counts.

The following steps were taken to handle placeholder values:

- Identified records where state = 100000, district = 100000, or pincode = 100000 as invalid or placeholder geographic entries.
- Flagged these records to ensure they did not influence geographic-level analysis.
- Excluded placeholder records from state-level, district-level, and pincode-level analyses to avoid misleading spatial conclusions.
- Retained these records for national-level aggregate analysis, where geographic detail was not required, ensuring overall enrolment totals remained accurate.
- Clearly documented this approach to maintain transparency and analytical validity.

Duplicate Record Handling:

To verify the integrity of the dataset, checks were performed to identify any duplicate records that could affect the analysis.

Duplicate handling steps included:

- Conducted duplicate checks across key fields in the dataset.
- Found no duplicate records during the verification process.
- Confirmed that the dataset was clean and suitable for further analysis without the need for duplicate removal.

Data Analysis and Visualizations:

Univariate Analysis:

Univariate analysis was conducted in order to be aware of the general distribution and primitive features of Aadhaar enrolment data. In this analysis, one variable is looked into each time to come up with the general trends, predominant categories and regional, age and time variations.

It assists in developing a background knowledge before delving into the association between two or more variables.

The important univariate analyses undertaken are:

- Total Aadhaar enrolment by state - Analysis to identify the states with best enrolment volumes and the worst enrolment volumes.
- Analysis of the district-wise distribution of enrolment to identify the hotspots within enrolment.
- Analysis of age-group-based enrolments (0-5, 5-17 and 18+) in order to learn about demographics enrolment participation.
- Evaluation of pincode-level enrolment in order to realise the distribution and concentrations of enrolments at a granular level.

1. Aadhaar Enrolment by Age Group

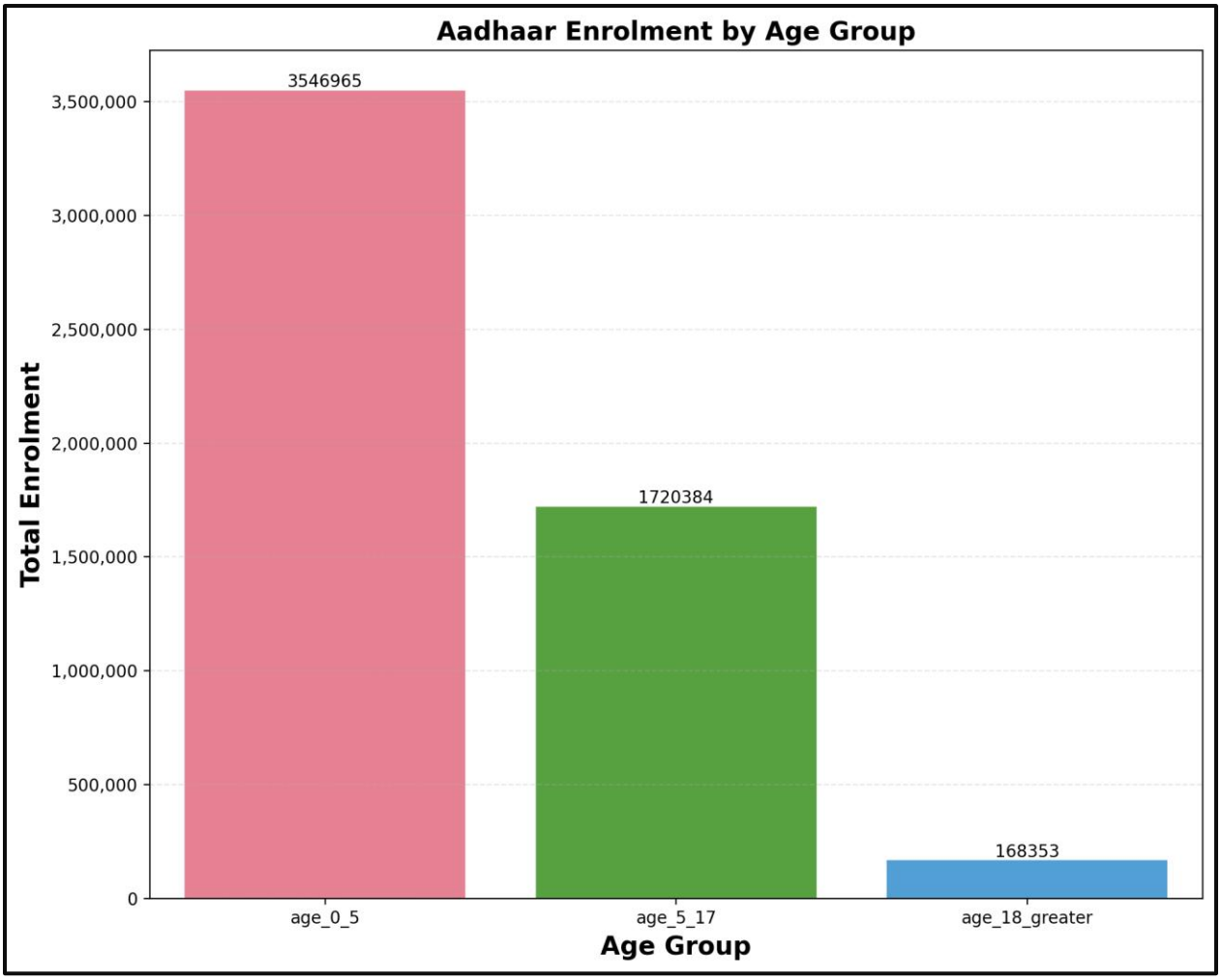

Chart Description:

The bar chart illustrates total Aadhaar enrollment for each group: 0-5, 5-17, 18+ years.

Insights:

- The highest number of enrollments is always among 0-5 age group only to indicate that the young children are being hooked on by the newly forced drives and requirements.
- The 5-17 age bracket is high but lower than 0-5 and the bracket 18+ is the least; this may be an indication of an already existing population of adults who already have Aadhaar and that the current single category of people who have not yet received Aadhaar (on age basis or immigration) form a minor proportion.

Administrative Importance:

- Maintenance of the Early Childhood Enrollments: Since the enrollment rate in the 0-5 age category is high, the administration should maintain the same and even extend the coverage, which is to provide every born child and the young children with the Aadhaar.
- Targeted Adult Engagement: In the case of adults (18+) this should be changed. Rather than general enrolment campaigns, the administrative policies should focus on: **Refreshing Existing Data**
- Marginalized Populations: Reaching out to specific marginalized communities or remote regions where Aadhaar penetration might still be low.
- Migrant Populations: Addressing the unique challenges of enrolling internal and external migrant populations.

Code:

```
plt.figure(figsize=(10, 8))
age_cols = ['age_0_5','age_5_17','age_18_greater']
total_age = total[age_cols].sum()
ax = sns.barplot(x=total_age.index,y=total_age.values,palette='husl')
plt.title('Aadhaar Enrolment by Age Group',fontsize=15,weight='bold')
plt.xlabel('Age Group',fontsize=15,weight='bold')
plt.ylabel('Total Enrolment',fontsize=15,weight='bold')
plt.grid(axis='y',linestyle='--',alpha=0.3)
plt.ticklabel_format(style='plain', axis='y')
import matplotlib.ticker as ticker
plt.gca().yaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
for p in ax.patches:
            ax.annotate(format(p.get_height(), '.0f'),
                        (p.get_x() + p.get_width() / 2., p.get_height()),
                        ha = 'center', va = 'center',
                        xytext = (0, 5),
                        textcoords = 'offset points')
plt.tight_layout()
plt.savefig("UV_analysis_AA_EN_by_age_group",dpi=200, bbox_inches='tight')
plt.show()
```

2. Top 20 States by Aadhaar Enrolment/Update:

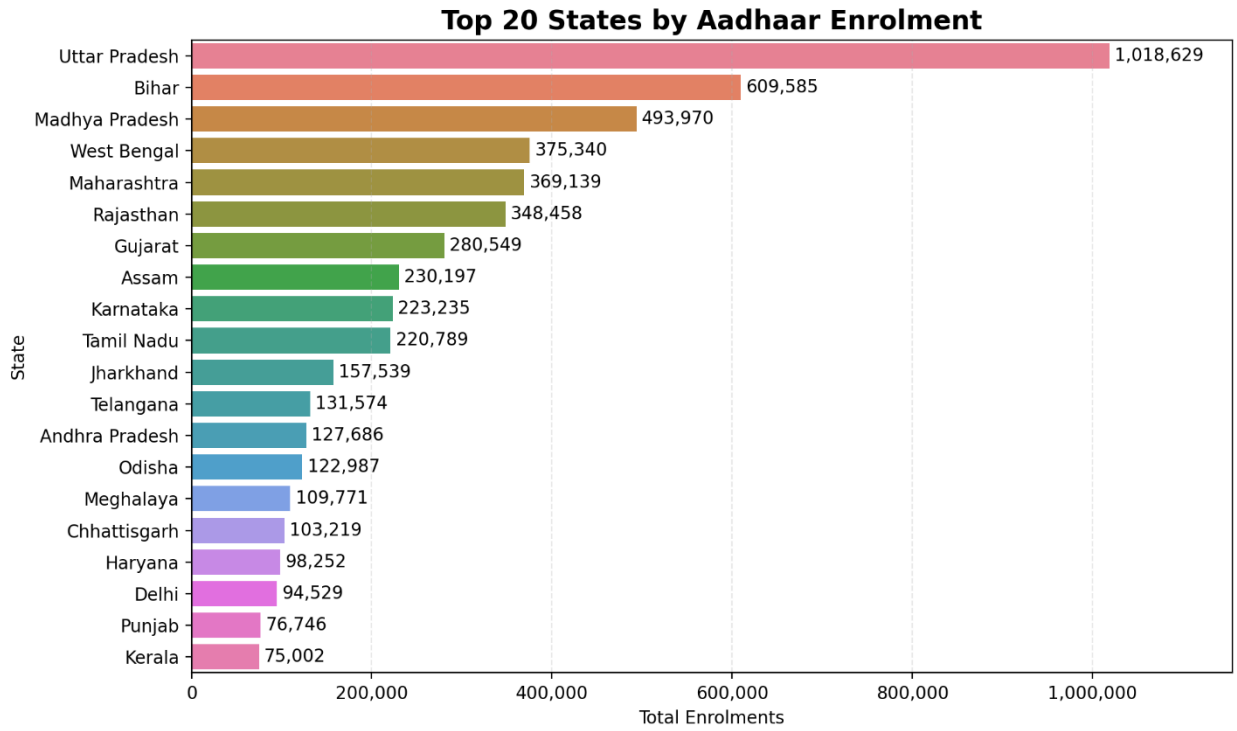

Chart Description:

- The bar chart shows the top 20 states in terms of Aadhaar enrolments

Insights:

- Most bulk of the national volume of enrolment goes to the 5 states, which means that the demand in Aadhaar services will experience strong regional imbalance.
- States that are economically developed like Maharashtra, Tamil Nadu and Karnataka experience relatively low levels of enrolment implying that they have higher Aadhaar saturation as opposed to new registrations.
- The states with a heavy migration population (Uttar Pradesh, Bihar, Madhya Pradesh, Rajasthan, and West Bengal) have a much greater enrolment activity indicating a regular update of address and demographics.
- Enrolment volumes decline abruptly beyond the top 10 states indicating two definite zones of Aadhaar demand namely high-demand and stable-demand regions.
- The active enrolment persists in smaller and Northeastern states where the population size is smaller, that is, less volume does not imply less need of the service.

Administrative Importance:

- Provision distribution of Aadhaar enrolment kits, operators and backend infrastructure according to demand per state rather than randomly.
- Additional capacity in the high volume states such as Uttar Pradesh and Bihar in times of peak enrolment to ease the load on the system and reduce waiting duration.
- Acquire better efficiency in updating services (address and biometric and mobile number updates) in the high-saturation states (Maharashtra, Tamil Nadu, Karnataka).
- Use mobile Aadhaar enrolment unit in the districts with large levels of migration to serve migrant and seasonal populations.
- Adopt state specific Aadhaar service strategies rather than having one national operational model.

Code:

```
plt.figure(figsize=(10,6), dpi=200)
state_group =
total.groupby('state')['total_enrollment'].sum().sort_values(ascending=Fal
se).head(20)
ax1 = sns.barplot(y=state_group.index, x=state_group.values,
palette='husl')
plt.title('Top 20 States by Aadhaar Enrolment', fontsize=15,
weight='bold')
plt.xlabel('Total Enrolments')
plt.ylabel('State')
plt.grid(axis='x', linestyle='--', alpha=0.3)
plt.ticklabel_format(style='plain', axis='x')
import matplotlib.ticker as ticker
plt.gca().xaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
for container in ax1.containers:
    ax1.bar_label(
        container,
        fmt='{:,.0f}',
        label_type='edge',
        padding=3
    )
ax1.set_xlim(0, ax1.get_xlim()[1] * 1.08)
plt.tight_layout()
plt.savefig("UV_analysis_top20_states_by_AA_EN", dpi=200,
bbox_inches='tight')
plt.show()
```

3. Distribution of Aadhaar Enrolment per Pincode

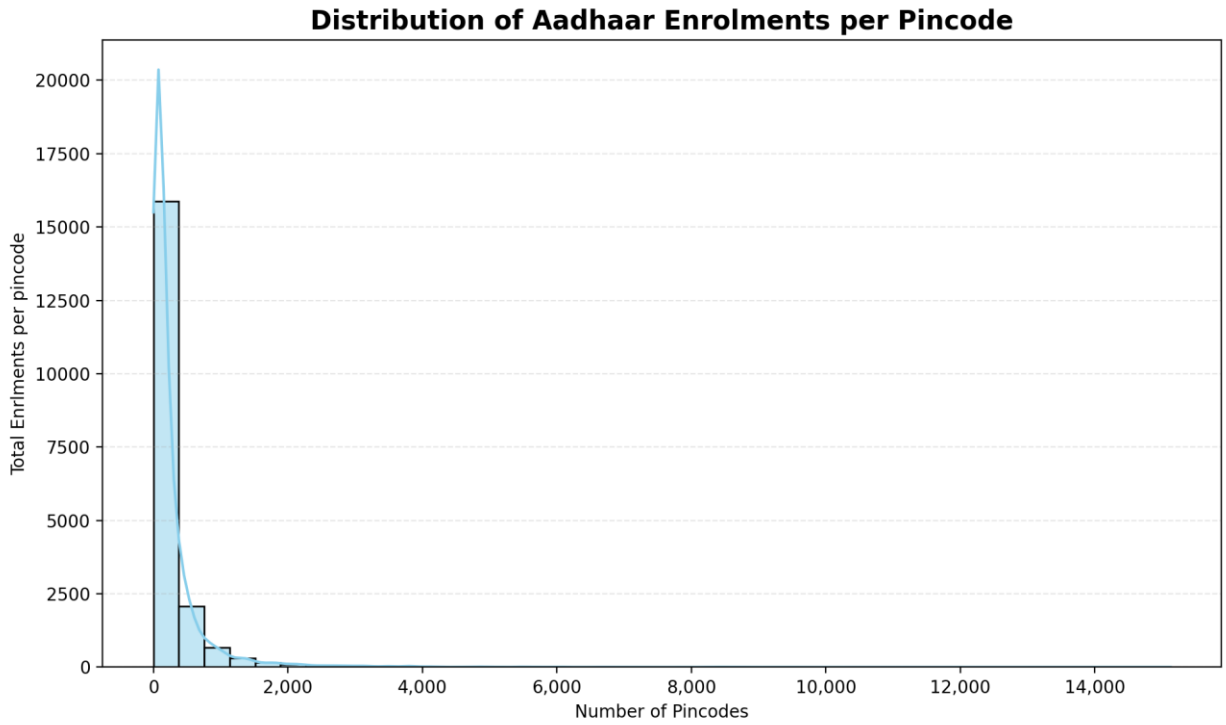

Chart Description:

- This is a histogram showcasing the number of enrollments per pincode.

Insights:

- Aadhaar enrollment across different pincode shows a strong right-skewedness.
- This shows that there are only a few pincode with higher enrollment, and all other pincode have low enrollment.
- Those high enrollment pincode may be due to the following reasons
    - Highly populated cities
    - Migration zones
    - Places where Aadhaar camps or update drives
    - Government Service Centers

Administrative Importance:

- The places with higher Aadhaar activity need to be considered, and more infrastructure needs to be provided, and staffing needs to be increased as well.
- Use mobile Aadhaar centers where the volume is medium.

- Avoid wasting resources where the demand is least, and try to allocate resources based on the requirement.

Code:

```
plt.figure(figsize=(10,6), dpi=200)
pincode_group =
total.groupby('pincode')['total_enrollment'].sum().sort_values(ascending=F
alse)
ax2 = sns.histplot(pincode_group,bins=40,kde=True,color='skyblue')
plt.title('Distribution of Aadhaar Enrolments per Pincode', fontsize=15,
weight='bold')
plt.xlabel('Number of Pincodes')
plt.ylabel('Total Enrlments per pincode')
plt.grid(axis='y', linestyle='--', alpha=0.3)
plt.ticklabel_format(style='plain', axis='x')
import matplotlib.ticker as ticker
plt.gca().xaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
plt.tight_layout()
plt.savefig('UV_analysis_Distribution_AA_EN_per_pincode')
plt.show()
```

4. District-Wise Total Enrollment Distribution:

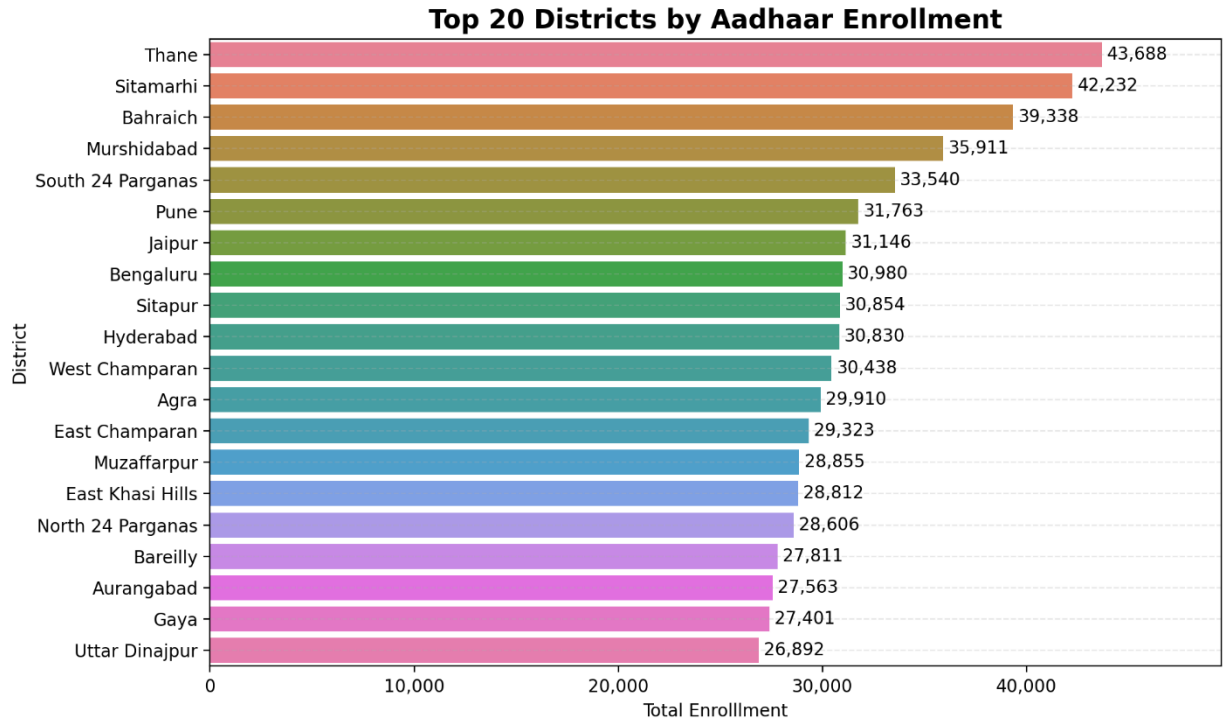

Chart Description:

- This chart shows the top 20 districts with the highest Aadhaar enrolments, highlighting districts that act as major service hotspots.

INSIGHTS:

- The level of enrollment in Aadhaar at the district level is extremely high with the highest concentration in such districts as Thane, Sitamarhi, Bahraich, Murshidabad and others.
- The best of these districts are in high-population areas or ones receiving high migrations like Maharashtra, Bihar, Uttar Pradesh, and West Bengal.
- Best enrollments are reported in urban and semi-urban cities (e.g., Thane, Pune, Bengaluru, and Hyderabad) as a result of the population density and regular updates.
- Other rural districts (e.g., Sitamarhi, Bahraich, and Champaran districts) are also very high, meaning that Aadhaar is largely relied on in the welfare schemes.
- The numbers of enrolled decrease in a gradual manner following the front runners with the result that there is an unbalanced distribution of workload amongst the districts.

ADMINISTRATIVE IMPORTANCE/ACTIONS:

- Distribution of more enrollment operators and biometric kits to high-enrollment districts should be considered Aadhaar service hotspots areas.
- Frequent Aadhaar updates and corrections should be the priority of urban hotspots districts as it involves rapid and timely dissemination of change.
- Rural high-volume districts must have the additional permanent, enrollment centers to cater the welfare beneficiaries.
- To equalize the workload, the mobile Aadhaar vans can be distributed in the nearby low-capacity districts.

Code:

```
top_districts_total =
total.groupby('district')['total_enrollment'].sum().sort_values(ascending=
False)
top_districts =
total.groupby('district')['total_enrollment'].sum().sort_values(ascending=
False).head(20)
plt.figure(figsize=(10,6), dpi=200)
ax3 = sns.barplot(y=top_districts.index, x=top_districts.values,
palette='husl')
plt.title("Top 20 Districts by Aadhaar
Enrollment",fontsize=15,fontweight='bold')
```

```
plt.xlabel('Total Enrolllment')
plt.ylabel('District')
plt.grid(axis='y', linestyle='--', alpha=0.3)
plt.ticklabel_format(style='plain', axis='x')
import matplotlib.ticker as ticker
plt.gca().xaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
for container in ax3.containers:
    ax3.bar_label(
        container,
        fmt='{:,.0f}',
        label_type='edge',
        padding=3
    )
ax3.set_xlim(0, ax3.get_xlim()[1] * 1.08)
plt.tight_layout()
plt.savefig('UV_analysis_Top20_districts_by_AA_EN')
plt.show()
```

Bivariate Analysis:

Bivariate analysis was conducted to look at the relationship that exist between 2 variables at the same time and the relationship of how the various factors affect the pattern of Aadhaar enrolment. The analysis such as the said one will reveal a dependency and comparison across regions, age groups, and time periods, with deeper insights as compared to individual variable distributions. The following are the important bivariate analyses conducted:

- Analysis of enrolment on the basis of age group in the states in order to gain insight on demographic disparities of Aadhaar enrolment.
- Review of monthly enrolment patterns across the states with the aim of establishing when the enrolment has been active or inactive.
- Comparison of the average daily enrolment in different states to determine continuous demand on operation.

1. State x Age Group Enrolment Pattern:

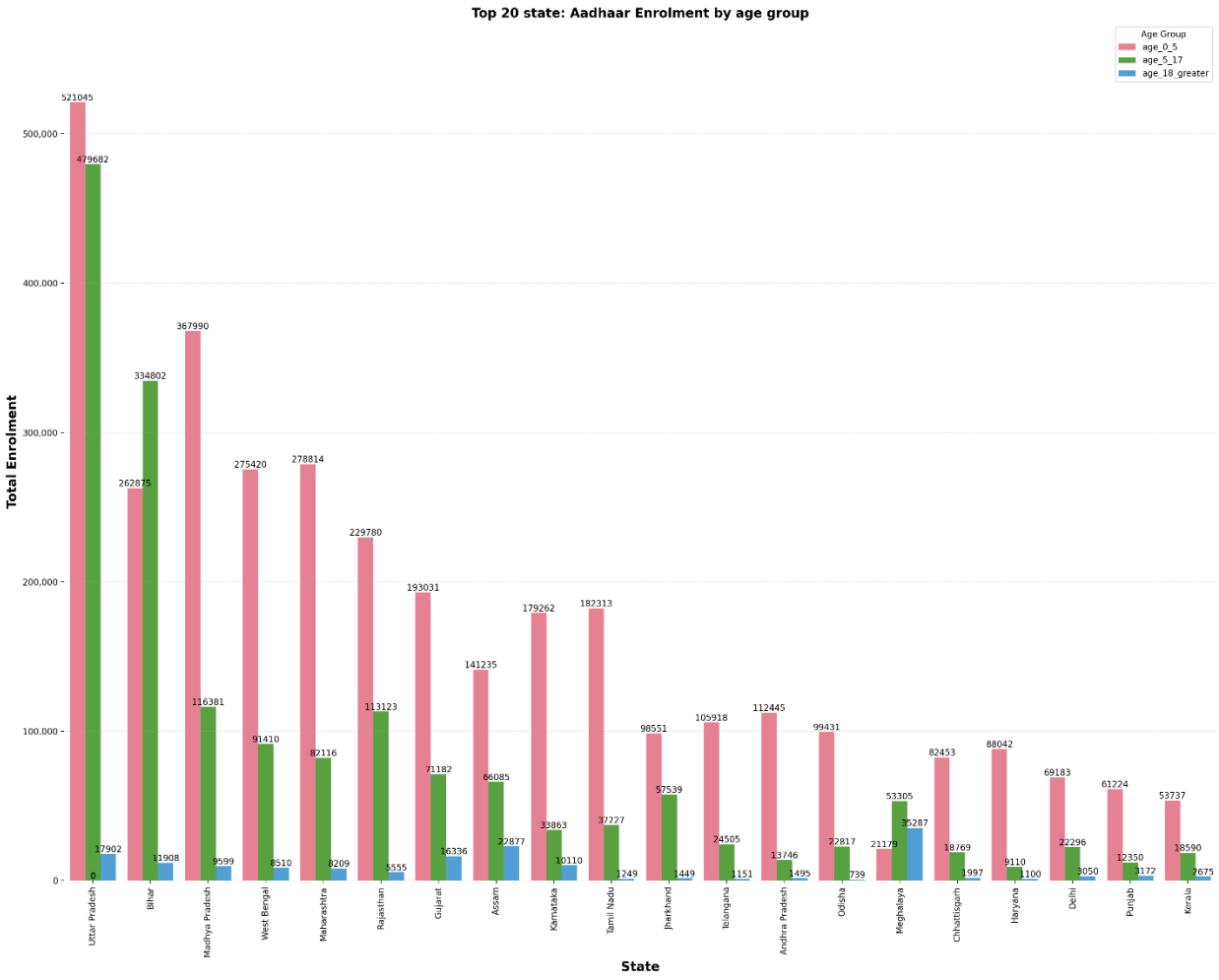

Chart Description:

The stacked bar chart shows how each age group in states differ and the top 20 states in the enrollment

Insights:

- Uttar Pradesh and Bihar have the greatest Aadhaar enrolments by age group as it indicates that these are the two states with immense population and demand of Aadhaar services.
- The number of children at 0-5 age bracket is very high in most states and particularly in UP, Bihar, Madhya Pradesh and Rajasthan, this implies that a large number of children are now enrolled.
- These states have a similar high figure of 5-17 years and this indicates that children of school-going age are constantly updating their biometrics.

- The number of people over the age of 18+ is low in the majority of states, and it means that these adults already possess Aadhaar and visit to update it.
- The Southern states such as Tamil Nadu, Karnataka, Telangana and Kerala have fewer child enrolments but adult-updates which are consistent, i.e. Aadhaar is already fully established in them.

Administrative Importance:

- More Aadhaar enrolment centers in schools and in villages should be introduced in states where there are high numbers of child and youth enrolments.
- The high population states such as UP and Bihar would be provided with additional staff and biometric machines to cope with the huge workload.
- States with more adult enrolments would need to concentrate on more expeditious update services like address and biometric repair.
- The mobile Aadhaar vans must be dispatched to the busy areas and regions of migration so that the people do not have to go far.
- The government is advised to strategize Aadhaar services according to the state requirements and according to the age groups rather than just design it as a system-wide system.

Code:

```
total_agg_raw =
total.groupby('state')[['age_0_5','age_5_17','age_18_greater']].sum()
total_agg_raw['overall_enrollment'] = total_agg_raw['age_0_5'] +
total_agg_raw['age_5_17'] + total_agg_raw['age_18_greater']
total_agg = total_agg_raw.sort_values(by='overall_enrollment',
ascending=False).head(20).reset_index()
total_agg_melt =
total_agg.melt(id_vars='state',value_vars=['age_0_5','age_5_17','age_18_gr
eater'],var_name='Age Group',value_name='Enrolment')
# State x Age Group Enrollment Pattern
plt.figure(figsize=(20,16), dpi=200)
ax4 = sns.barplot(data=total_agg_melt,x='state',y='Enrolment',hue='Age
Group',palette='husl')
ax4.set_title('Top 20 state: Aadhaar Enrolment by age
group',fontsize=15,weight='bold')
ax4.set_xlabel('State',fontsize=15,weight='bold')
ax4.set_ylabel('Total Enrolment',fontsize=15,weight='bold')
ax4.grid(axis='y',linestyle='--',alpha=0.3)
sns.despine(left=True, bottom=True)
plt.xticks(rotation=90)
plt.ticklabel_format(style='plain', axis='y')
import matplotlib.ticker as ticker
```

```
plt.gca().yaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
for p in ax4.patches:
            ax4.annotate(format(p.get_height(), '.0f'),
                         (p.get_x() + p.get_width() / 2., p.get_height()),
                         ha = 'center', va = 'center',
                         xytext = (0, 5),
                         textcoords = 'offset points')
ax4.set_ylim(0, ax4.get_ylim()[1] * 1.05)
plt.tight_layout()
plt.savefig('BV_analysis_Top20_states_AA_EN_AG_statexage_group')
plt.show()
```

2. Monthly Total Aadhaar Enrollment across Year:

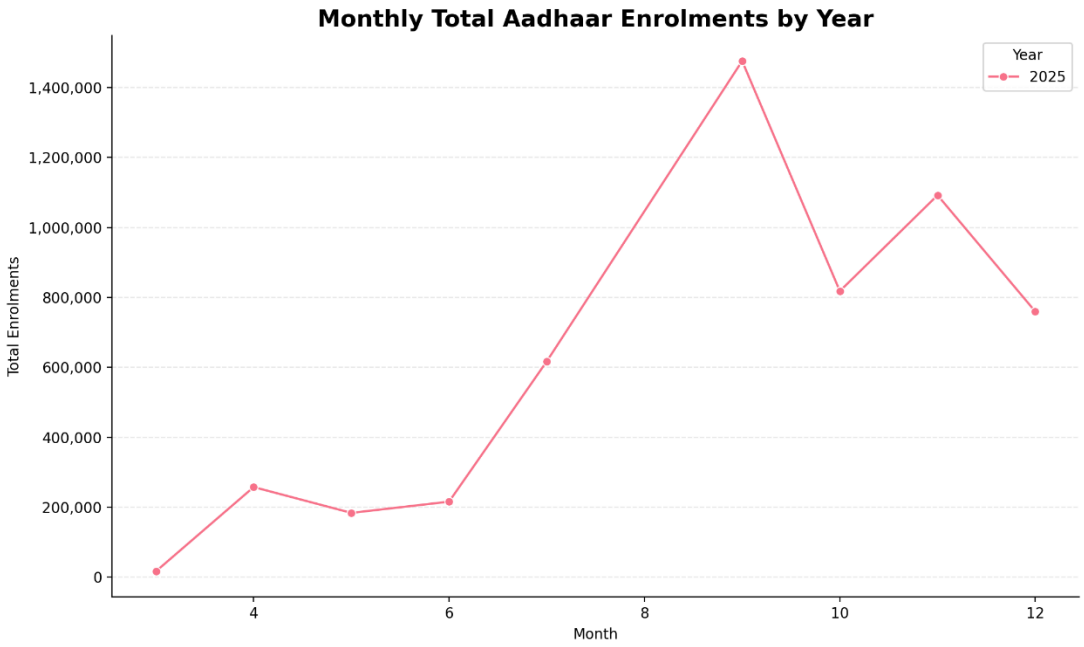

Chart Description:

- This chart presents the month-wise trend of Aadhaar enrolments, highlighting significant variability in enrolment activity across the year.

INSIGHTS:

- The enrolment in Aadhaar is very seasonal with high peaks towards the later months of the year showing demand induced by external factors, than by natural, steady growth.
- Peak-month enrollments are several times greater than the base months, which demonstrates that there is an extreme level of volatility in system-loading.
- This type of spike in demand will probably be the result of policy deadlines, cycles on welfare scheme enrolment, yearly academic updates, or changes in address due to migration.
- The fact that there has been a similar sustained but less enrollment volumes between peak months implies that core Aadhaar demand is more or less stable and surges are of an event-driven nature.
- With no pre-thinking, these peaks are highly dangerous of bottlenecks in service, latency in the system, as well as a quality of user experience.

Administrative (Strategic and Operational) Action:

- Replacing the unchanging yearly planning with an anticipation model based on the calendar to establish alignment of the personnel and structures with anticipated peak of activities.
- Pre-scale enrolment infrastructure (operators, biometric devices, backend capacity) 4-6 weeks to expected spikes.
- Additional surging capacity by altering holiday months: use temporary surge capacity, mobile addition of new enrollment capacity and extending operating hours.
- System upgrades, data verification, operator training and process audits should be done during low-demand months to enhance the quality of service in the long term.
- Make demand control dashboards monthly to enable timely response to the approach of an enrollment burst.

Code:

```
total['date'] = pd.to_datetime(total['date'])
if 'month' not in total.columns:
    total['month'] = total['date'].dt.month
if 'year' not in total.columns:
    total['year'] = total['date'].dt.year
month_group = total.groupby(['year', 'month'])[['age_0_5', 'age_5_17',
'age_18_greater']].sum()
month_group['total_enrollment'] = month_group.sum(axis=1)
month_group = month_group.reset_index()
# Monthly Total Aadhaar Enrollments
plt.figure(figsize=(12, 7), dpi=200)
```

```
ax5 = sns.lineplot(data=month_group, x="month", y="total_enrollment",
hue="year", marker='o', palette='husl')
ax5.set_title('Monthly Total Aadhaar Enrolments by Year', fontsize=16,
weight='bold')
ax5.set_xlabel('Month')
ax5.set_ylabel('Total Enrolments')
ax5.legend(title='Year')
sns.despine()
plt.grid(axis='y', linestyle='--', alpha=0.3)
import matplotlib.ticker as ticker
plt.gca().yaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
plt.savefig("BV_analysis_Monthly_Total_AA_EN_by_year")
plt.show()
```

3. Average Daily Aadhaar Enrolments in States

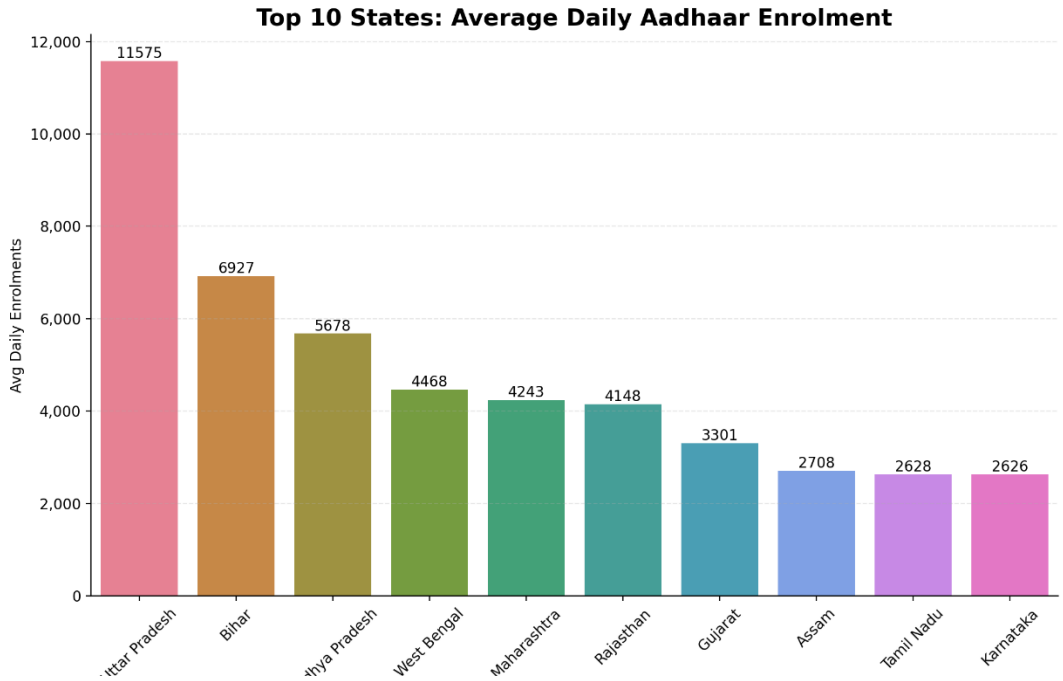

Chart Description:

- This chart presents the top 10 states based on average daily Aadhaar enrolments, capturing sustained operational demand rather than one-time spikes.

INSIGHTS:

- The average daily enrolment in Uttar Pradesh is approximately twice that of the next state signifying chronic operational pressure rather than demand.

- The state of Bihar has shown to have high daily figures implying that Aadhaar services have been incorporated in daily administrative procedures as well as welfare procedures.
- The sharp drop following the top two states indicates that it is mostly the concentration of Aadhaar demand that is highly centralized with most states having much lower daily loads.
- The demand in middle-tier states (Madhya Pradesh, West Bengal, Maharashtra, Rajasthan) is stable and manageable, the best fit in optimized staffing.
- Southern state enrolments that are lower daily also reflect the mature Aadhaar penetrations, with the demand lowly based on updates and not on new-enrolments.
- Average daily enrolment exposes some work load realities that cannot be depicted by total enrolment figures alone.
- States that have low numbers coming in daily might have idle infrastructure, whereas states that have high demand rates may end up dealing with staff burnouts and delays in services.

ACTIONS:

- Making average daily enrolment a key KPI in Aadhaar infrastructure planning and budgeting.
- Install increased operator strength and biometric equipment permanently in a state with a continuous high demand of day.
- Present the levels of operation models:
    - The states with high demand - increased working hours and counters.
    - In the medium demand category - optimal staffing.
    - The conditions that are low-demand, such as shared/mobile resources.
- Facilitate resource redeployment between the states and so that trained operators can be temporarily moved to high-demand times.
- Use rolling daily averages to predict points of trouble by maintaining predictive demand by monitoring.
- Optimize low-demand states with centers consolidation and concentrating on appointment-based update services.
- Credential day-to-day demand information so as to minimize waiting, enhance service quality and operational inefficiencies.

Code:

```
state_daily = total.groupby(['state', 'date'])[['age_0_5', 'age_5_17',
'age_18_greater']].sum()
state_daily['total_enrollment'] = state_daily.sum(axis=1)
```

```
avg_daily =
state_daily.groupby('state')['total_enrollment'].mean().sort_values(ascend
ing=False).head(10).reset_index()
# Average Daily Aadhaar Enrollment in States
plt.figure(figsize=(12, 7), dpi=200)
ax6 = sns.barplot(data=avg_daily, x='state', y='total_enrollment',
palette='husl')
ax6.set_title('Top 10 States: Average Daily Aadhaar Enrolment',
fontsize=16, weight='bold')
ax6.set_xlabel('State')
ax6.set_ylabel('Avg Daily Enrolments')
sns.despine()
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.3)
plt.ticklabel_format(style='plain', axis='y')
import matplotlib.ticker as ticker
plt.gca().yaxis.set_major_formatter(ticker.StrMethodFormatter('{x:,.0f}'))
for p in ax6.patches:
            ax6.annotate(format(p.get_height(), '.0f'),
                         (p.get_x() + p.get_width() / 2., p.get_height()),
                         ha = 'center', va = 'center',
                         xytext = (0, 5),
                         textcoords = 'offset points')
ax4.set_ylim(0, ax6.get_ylim()[1] * 1.05)
plt.savefig("BV_analysis_Top10_states_AA_EN_avg_daily")
plt.show()
```

Trivariate Analysis:

The Trivariate analysis method was used to examine the joint impact of three variables at the same time, which made it possible to understand the trends of Aadhaar enrolment in more detail. This analysis can be used to determine the areas of enrolment, spring and autumn demand fluctuations, and the trends that cannot be determined by using univariate analysis and bivariate analysis only: it includes geographic, demographic and temporal dimensions of its evaluation.

The most important Trivariate analyses were conducted which are:

- Comparison of state enrolment patterns by month to determine seasonal enrolment patterns in the region.
- Research on the enrolment by age in states to determine the influence of the demographic composition on the level of enrolment.
- Analysis of district-specific enrolment by age so as to identify demand due to localized demographics.
- Comparison of monthly enrolment patterns of major states, by in-age-group, to monitor combined effects of time and population age.
- Marking of the states and districts with high demand in peak months to mark centers of enrolment pressure.

1. State x Month x Age Group (Heatmap):

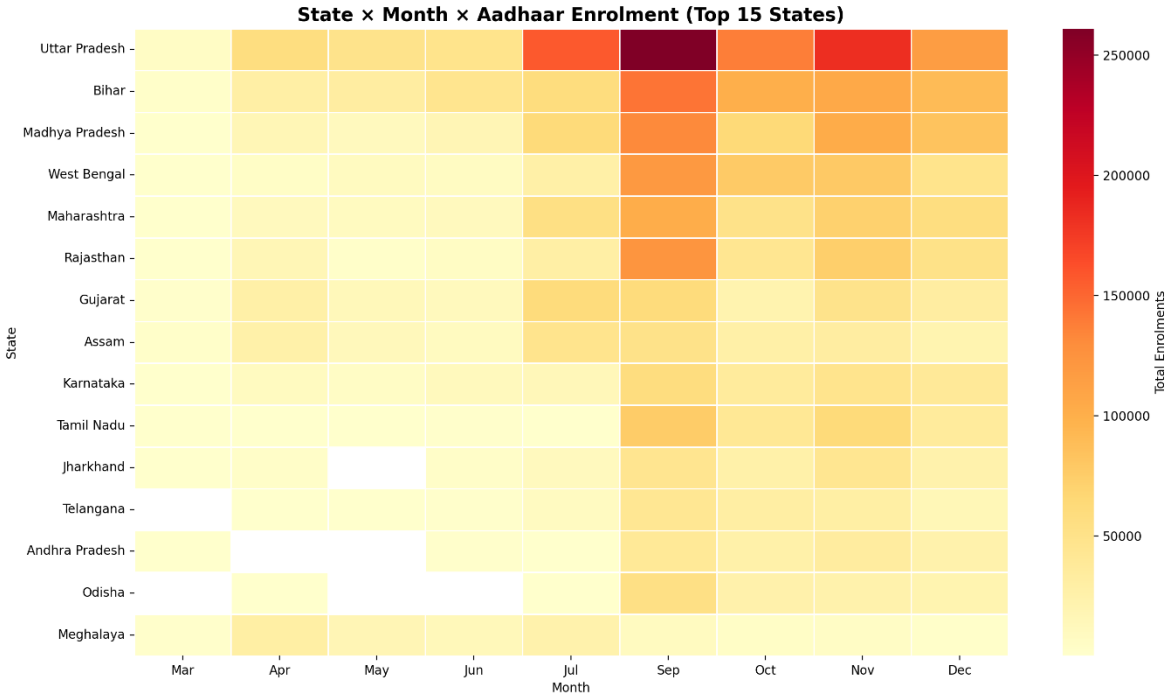

Chart Description:

- This heat map indicates the number of enrolments with Aadhaar in the 15 leading states over the months indicating that the level of enrolment is varying with the states and the months.
- Darker shades indicate greater enrolments, so seasonal and regional hotspots can be easily determined.

Insights:

- There is also no equal distribution of Aadhaar enrolments across months, with majority of states depicting a clear growth in the last half year.
- The months of September through November turn out to be the most successful enrolment months in several populated states.
- The highest volumes of enrolment are always registered in Uttar Pradesh particularly during peak months, which implies continued demand.
- Other states such as Bihar, Madhya Pradesh, Maharashtra and West Bengal have high seasonal spikes however with weaker intensity.
- Smaller states and UTs (e.g. Meghalaya, Odisha) have less and less variable enrolment, with little seasonality.
- The trend indicates enrolment spurts to be event-based (welfare plans, deadlines, college cycles) instead of yearly demand.

Actions:

- Enroll more centers, operators and biometric kits in the high demand states in the peak months (September-November).
- Ramp-up must start at least one month before the periods of expected peak to prevent congestion of the services.
- Implement mobile enrolment units in those states that are characterized by steep seasonal spikes to deal with the transient demand.
- Staff training, which seeks to maintain a system and ensure data quality, is done using low-enrolment months.
- Implement state specific enrolment schedules rather than a national model of deployment.
- Watch monthly trends so to be able to plan proactively instead of acting in response.

Code:

```
# State x Month x Age group(Heatmap)
total['month'] = total['date'].dt.month
top_states = (
    total.groupby('state')['total_enrollment']
    .sum()
    .sort_values(ascending=False)
    .head(15)
    .index
)
filtered = total[total['state'].isin(top_states)]
pivot_state_month = (
    filtered
    .groupby(['state','month'])['total_enrollment']
```

```
    .sum()
    .reset_index()
    .pivot(index='state', columns='month', values='total_enrollment')
)
pivot_state_month = pivot_state_month.loc[
    pivot_state_month.sum(axis=1).sort_values(ascending=False).index
]
plt.figure(figsize=(14,8), dpi=200)

ax = sns.heatmap(
    pivot_state_month,
    cmap='YlOrRd',
    linewidths=0.4,
    cbar_kws={'label': 'Total Enrolments'}
)

plt.title('State × Month × Aadhaar Enrolment (Top 15 States)',
fontsize=15, weight='bold')
plt.xlabel('Month')
plt.ylabel('State')
ax.set_xticklabels(
    ['Mar','Apr','May','Jun','Jul','Sep','Oct','Nov','Dec'],
    rotation=0
)

plt.tight_layout()
plt.savefig("TV_analysis_State_Month_AA_EN_top15")
plt.show()
```

2. State x Aadhaar Enrolment x Age Group (Stacked Bar)

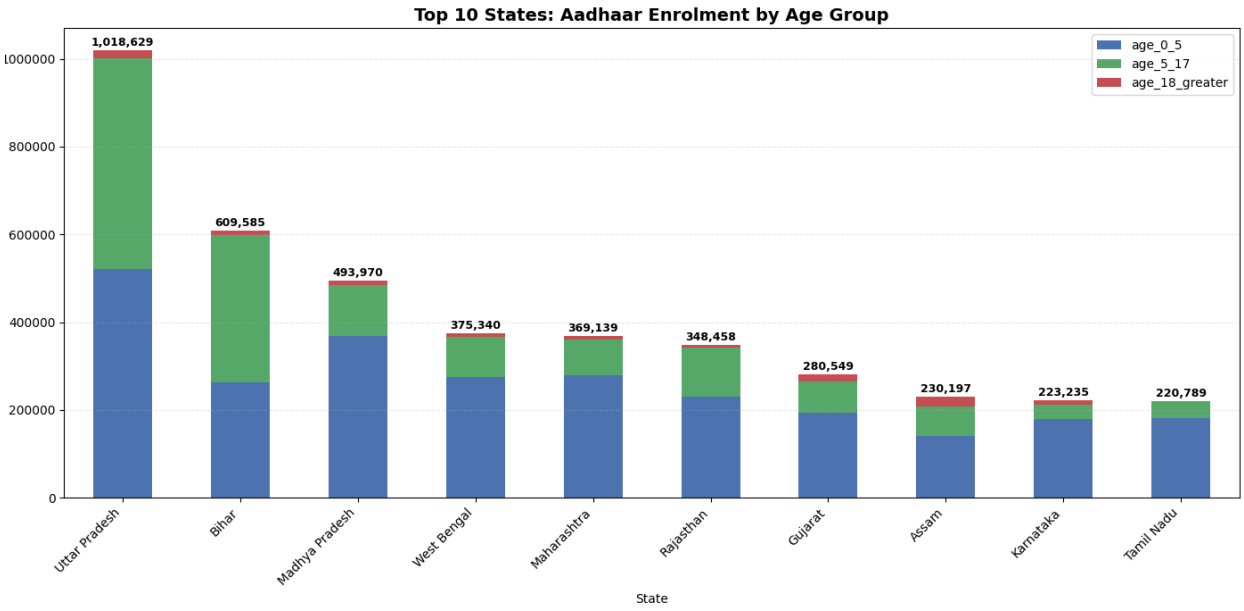

Chart Description:

- The chart figures indicate the total Aadhaar enrolment of the top 10 states in terms of age (0-5, 5-17 and 18+).
- It assists in knowing the age groups that lead to enrolment in high performing states.

Insights:

- The 0-5 year's age is the highest age in enrolment of all the top states, which implies good Aadhaar enrolment among the newborns and children.
- Clearly outstanding is Uttar Pradesh and Bihar which are mainly triggered by high numbers of child and adolescents.
- Adult (18+) enrolments are relatively small, and that is likely Aadhaar covered among adults is mostly saturated.
- Such states as Maharashtra and West Bengal have equal distribution, which is a result of stable enrolment for the different age groups.
- States which are in the South (Tamil Nadu, Karnataka) are at lower totals indicating low fertility and high Aadhaar maturity.

Actions:

- Intensify the Aadhaar enrolment with hospital, anganwadi, and school enrolment in states with high child population.
- Devise age-based plans of enrolment rather than state-wide plans.

- Change the present orientation of the shift operation in developed states toward update services (biometrics, address, mobile).
- Trends in child enrolments can be used to forecast demand in future Aadhaar updates in the next 10 years.

Code:

```
# State x Aadhaar Enrollment x Age group(Stacked Bar Graph)
state_age = (
    total
    .groupby('state')[['age_0_5','age_5_17','age_18_greater']]
    .sum()
)
state_age['total_enrollment'] = state_age['age_0_5'] +
state_age['age_5_17'] + state_age['age_18_greater']

state_age = state_age.sort_values(by='total_enrollment',
ascending=False).head(10)

ax7 = state_age[['age_0_5','age_5_17','age_18_greater']].plot(
    kind='bar',
    stacked=True,
    figsize=(14,7),
    color=['#4C72B0', '#55A868', '#C44E52']
)

plt.title('Top 10 States: Aadhaar Enrolment by Age Group', fontsize=14,
weight='bold')
plt.xlabel('State')
plt.ylabel('Total Enrolment')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.3)
plt.tight_layout()
plt.ticklabel_format(style='plain', axis='y')
for i, total_val in enumerate(state_age['total_enrollment']):
    plt.text(i, total_val + 10000, f'{total_val:,.0f}',
             ha='center', fontsize=9, fontweight='bold')
plt.savefig("TV_analysis_Top10_states_AA_EN_age_group")
plt.show()
```

3. District x Age Group x Enrolment

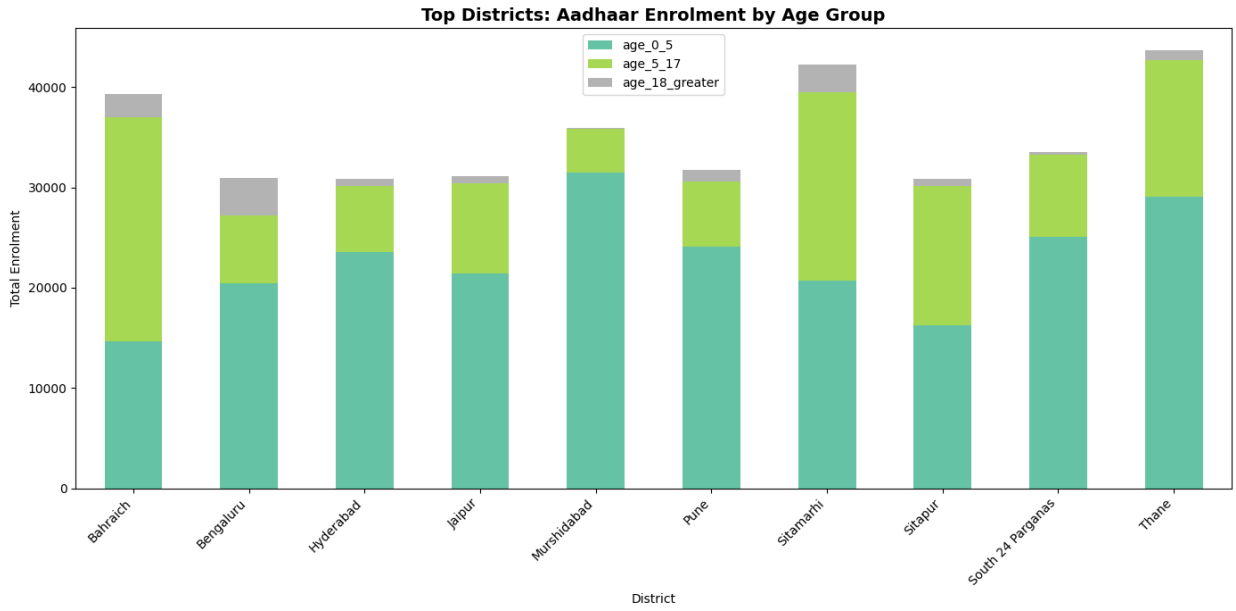

Chart Description:

- In this chart, the enrolments in Aadhaar in the leading districts have been compared with regard to age group.
- It singles out district-level hotspots and Aadhaar demand drivers (demographic).

Insights:

- There are quite high enrolment volumes in districts like Thane, Sitamarhi, Murshidabad, and Bahraich.
- The primary contributor as in most districts is child enrolment (0-5), as is the case in the state level.
- Other places (Bengaluru, Pune, and Hyderabad) have moderate but stable enrolments, primarily facilitated through updates.
- The 5-17 age share is somewhat higher in some districts which suggests that there is a school-related enrolment when compared to other districts.
- The difference at the district level is far more acute than the average at the state level that shows local demand pressure.

Actions:

- Enact district capacity planning, as opposed to implementing state-wide planning.
- Create mobile enrolment units on high demand districts on high seasons.

- Align Aadhaar enrolments with school admissions in enrolment hotspots districts Coordination of Aadhaar with school admissions.
- Give preference to the high-performers districts in terms of pilot use of digital or appointment-based enrolment.

Code:

```
top_districts = (
    total
    .groupby('district')['total_enrollment']
    .sum()
    .sort_values(ascending=False)
    .head(10)
    .index
)

district_age = (
    total[total['district'].isin(top_districts)]
    .groupby('district')[['age_0_5','age_5_17','age_18_greater']]
    .sum()
)

district_age.plot(
    kind='bar',
    stacked=True,
    figsize=(14,7),
    colormap='Set2'
)

plt.title('Top Districts: Aadhaar Enrolment by Age Group', fontsize=14,
weight='bold')
plt.xlabel('District')
plt.ylabel('Total Enrolment')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.savefig("TV_analysis_Top10_districts_AA_EN_age_group")
plt.show()
```

4. State x Month x Age Group (Line Chart):

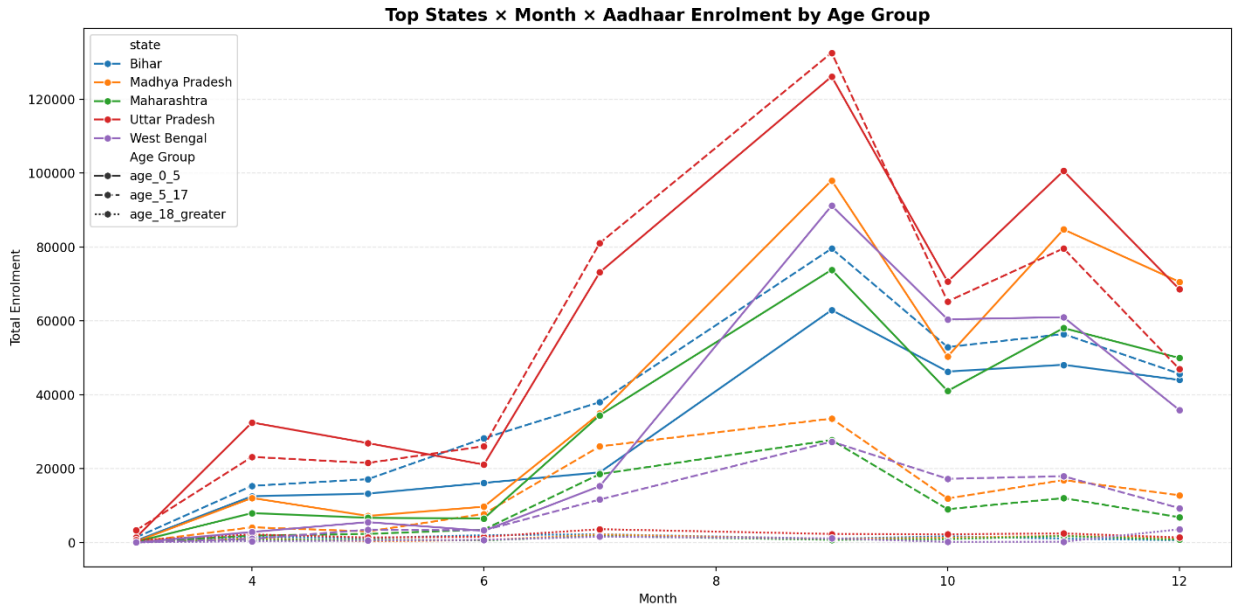

Chart Description:

- This graph monitors the trends of monthly Aadhaar enrolment of the leading states by their age group.
- It displays the time-dependent fluctuation in enrolment demand as well as the variation according to the regions and age.

Insights:

- Spike in enrolment between July and November is observable in all the major states.
- Uttar Pradesh always achieves the greatest number of enrolment monthly, particularly in children.
- The seasonal trend is observed in all the age groups but most in 0-5 and 5-17 groups.
- Adult enrolments are not very high but stable all year round boosting Aadhaar saturation.
- The same peak trends across the states imply triggers of policies or welfare on a national scale.

Actions:

- Pre-plan scale surge anticipating predictable seasonal surge to July before July.
- Additional staffing, working hours, and Biometric kits in July-November.
- Train, audit and upgrade the system at the beginning of the year.

- Bring predictive dashboard with monthly trends to change the reactive planning to proactive planning.

Code:

```
# State x Month x Age Group
monthly_age = (
    total
    .groupby(['state','month'])[['age_0_5','age_5_17','age_18_greater']]
    .sum()
    .reset_index()
)

top_states = (
    total
    .groupby('state')['total_enrollment']
    .sum()
    .sort_values(ascending=False)
    .head(5)
    .index
)

filtered = monthly_age[monthly_age['state'].isin(top_states)]
filtered_melted = filtered.melt(
    id_vars=['state', 'month'],
    value_vars=['age_0_5', 'age_5_17', 'age_18_greater'],
    var_name='Age Group',
    value_name='Enrolment'
)

plt.figure(figsize=(14,7), dpi=200)
sns.lineplot(
    data=filtered_melted,
    x='month',
    y='Enrolment',
    hue='state',
    style='Age Group',
    marker='o'
)

plt.title('Top States × Month × Aadhaar Enrolment by Age Group',
fontsize=14, weight='bold')
plt.xlabel('Month')
plt.ylabel('Total Enrolment') # Update y-axis label
plt.grid(axis='y', linestyle='--', alpha=0.3)
plt.tight_layout()
```

```
plt.savefig("TV_analysis_Top5_states_AA_EN_age_group_month")
plt.show()
```

Overall Insights:

The Aadhaar enrolment data analysis has indicated that the enrolment activity of India is highly uneven in geographic, age and time. The relatively few states and districts take a comparatively high proportion of the total enrolments, which means that demand is concentrated in certain areas and that the operations are always under pressure. At the state level, it is observed that the higher enrolments are always registered in the blockage of populous states, including Uttar Pradesh and Bihar whereas in some states, the activity is relatively low and steady.

Analysis in terms of the age shows that in most areas, children in the 05 age bracket take the highest share of Aadhaar enrolments. This presents good correlation of Aadhaar registration with birth-related and early childhood services. The enrolment rates among adults (18+) are more stable and low; this indicates that the Aadhaar development among adults is already quite saturated, and all progression will be primarily demand-based (updates rather than new enrolments).

Temporal analysis establishes evident season patterns in the activity of enrolment. The majority of states tend to have very high enrolment rates in the second half of the year and especially between July and November. Such time and place coordination of multiple states implies that the demand of enrolment is driven by outside factors like the deadlines of welfare schemes, or academic calendar, or administrative advocacy and not random variation. The pincode- and district-level analysis can further show hotspots which are localized and in which enrolment demand is always high, but which are often obscured by state-level aggregates.

In general, the results indicate that the demand to enroll Aadhaar is associated with the mixture of population characteristics, the geographical specifics, and seasonal phenomena. It is important to understand these multi-dimensional patterns when carrying out effective planning and delivering services fairly.

General Administrative Recommendations:

According to the findings obtained after the analysis, a number of administrative measures might be adopted to enhance the efficiency and access to Aadhaar services. Resource planning ought to cease to be a consistent deployment model but rather a demand-driven and region specific one. The states and districts with the high volumes of enrolment on the regular basis should be offered the permanent basis of staffing, the use of extra devices of biometric character, and the increased time of working.

The trends in enrolment are seasonal implying predictive capacity planning is necessary. The infrastructure of enrolment and manpower must be increased before the peak times such as July to November to alleviate congestions and waiting hours. When the demand is low, the administrative emphasis could be changed to training of staff, system upkeep, and improvement of quality of data.

Since child enrolments prevail, Aadhaar registration activities should be also incorporated with hospitals, anganwadi centers, and schools. High demand districts and remote locations can have mobile enrolment units to address temporary surges and enhancing access. In those areas where Aadhaar has already become a mature system, the priority in operations should be on update services which include biometric refresh and demographic updates.

On a strategic level, real time decision-making according to daily and monthly enrolment indicators should be introduced as continuous monitoring dashboard implements continuous monitoring. This kind of data driven planning can enable to optimize the resources, decrease the bottlenecks of the services as well as make Aadhaar service delivery more equitable in all regions.

Conclusion:

This paper shows that data-driven research has been useful in explaining the trend of Aadhaar enrolment in geographic, demographic, and time contexts. Using the methods of univariate, bivariate, and Trivariate analysis, the paper determines the enrolment hotspots, seasonal demand patterns, and trends influenced by an age group, implying that such patterns cannot be seen with more basic aggregation.

The impressions created have demonstrated the necessity of administrative planning which is flexible, predictive and region-specific instead of one-size-fits-all approach. The suggested measures will lead to efficiency in the operations, minimize enrolment bottlenecks, and increase the accessibility of the Aadhaar services. The analysis, in general, offers a useful system to utilize enrolment data to facilitate informed decision-making and enhance the Aadhaar ecosystem.