UIDAI Data Hackathon 2026

Project Title

Age-wise and District-wise Aadhaar Enrolment Analysis in Tamil Nadu

Participant Details

Name: Inbha
Course: MCA
Institution: Easwari Engineering College

Problem Statement

The objective of this project is to analyze age-wise and district-wise Aadhaar enrolment patterns in Tamil Nadu in order to:

- Identify dominant age groups contributing to enrollment.
- Understand regional disparities across districts
- Detect trends, anomalies, and unusual enrolment behavior over time
- Provide data-driven recommendations for improving enrolment strategies

Approach

The analysis aims to uncover patterns across districts, age groups, and time to identify enrolment distribution, growth trends, and potential disparities. A data-driven analytical approach using Python and Power BI was adopted to clean, process, visualize, and interpret the data.

Dataset Used

The dataset used for this project is the Aadhaar enrolment and update dataset released by UIDAI. The analysis focuses on Tamil Nadu state-level data. Each record represents enrollment data for a specific date, district, and pincode.

Key Columns used

- Date: Enrolment record date
- District: District name within Tamil Nadu

- Pincode: Area-level identifier
- Age_0_5: Enrolments for children aged 0–5 years
- Age_5_17: Enrolments for individuals aged 5–17 years
- Age_18_greater: Enrolments for individuals aged 18 years and above

The dataset contains approximately 111,325 records and captures detailed age-wise enrolment information across districts.

Methodology

The methodology followed in this project includes the following steps:

- Data Import and Inspection using Python (Pandas)
- Data type correction and date formatting
- Feature aggregation by district, month, year, and age group
- Exporting cleaned data for visualization in Power BI

Data Analysis and Visualisation

The dataset underwent multiple preprocessing steps to ensure accuracy and consistency:

- The date column was converted to datetime format for time-series analysis
- Missing values were replaced with zero to maintain numerical consistency
- Redundant columns were removed where applicable
- Data was sorted chronologically to enable accurate trend analysis

These steps ensured the dataset was clean, structured, and analysis-ready.

To enhance analytical insights, additional features were created:

- A total_enrolment column was generated by summing enrolments across all age groups
- Year and month were extracted from the date column for temporal analysis

These features enabled more meaningful comparisons across time and districts.

Exploratory Data Analysis

Age-wise Enrolment Analysis

The analysis revealed that Aadhaar enrolments in Tamil Nadu are overwhelmingly driven by children aged 0–5 years. Enrolments in the 5–17 age group were significantly lower, while enrolments for individuals aged 18 and above were minimal.

District-wise Analysis

Urban and densely populated districts recorded the highest total enrolments. Several districts consistently showed low enrolment activity, indicating possible challenges related to accessibility, awareness, or infrastructure.

Time-series Analysis

Monthly enrolment trends showed consistent child enrolments with periodic spikes in total registrations. These spikes are likely associated with special enrolment drives or policy initiatives.

Key Insights

- Aadhaar enrolments in Tamil Nadu are primarily driven by the 0–5 age group
- Adult enrolments (18+) are extremely low, indicating near-complete saturation
- Urban districts contribute significantly higher enrolment numbers than rural districts
- Certain districts consistently underperform compared to the state average
- Periodic enrolment spikes suggest the impact of government-led enrolment campaigns
- Total Aadhaar enrolments analyzed: approximately 274,000
- Chennai, Kancheepuram, and Tiruvallur are top-performing districts
- Noticeable monthly and yearly enrolment trends indicating seasonal variation
- Certain districts consistently show low enrolment, highlighting scope for policy intervention

Anomalies Observed

- Certain months exhibit unusually high enrolment levels compared to the average
- A few months show sharp declines in enrolment, possibly due to operational or external disruption.
- District-level anomalies highlight persistent overperformance and underperformance patterns

Recommendations

- Conduct targeted enrolment drives in consistently low-performing districts
- Deploy mobile enrolment units in rural and underserved regions
- Strengthen awareness programs focused on child Aadhaar enrolment
- Use historical time-based patterns to plan future enrolment campaigns

Conclusion

This project provides a comprehensive analysis of Aadhaar enrolment patterns in Tamil Nadu using age-wise, district-wise, and time-based data. The findings indicate strong child-driven

enrolment trends and near-saturation among adults. The insights derived can assist UIDAI in improving enrolment strategies, enhancing outreach, and optimizing operational planning.

Code and Tools Used

- Python (Pandas, NumPy, Matplotlib)
- Google Colab for analysis and preprocessing
    - UIDAIDataHackathon.ipynb
- Power BI for interactive dashboards and visualizations

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('tamilnadu_uidai.csv')
```

```
df.head()
```

	date	state	district	pincode	age_0_5	age_5_17	age_18_greater
0	01-05-2025	Tamil Nadu	Pudukkottai	614616	143	34	13
1	01-06-2025	Tamil Nadu	Kanchipuram	631502	264	76	11
2	01-06-2025	Tamil Nadu	Salem	636006	90	21	11
3	01-06-2025	Tamil Nadu	Vellore	635810	154	65	16
4	01-06-2025	Tamil Nadu	Kallakurichi	605801	53	12	10

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111325 entries, 0 to 111324
Data columns (total 7 columns):
 #   Column          Non-Null Count   Dtype 
---  ------          --------------   ----- 
 0   date            111325 non-null  object
 1   state           111325 non-null  object
 2   district        111325 non-null  object
 3   pincode         111325 non-null  int64 
 4   age_0_5         111325 non-null  int64 
 5   age_5_17        111325 non-null  int64 
 6   age_18_greater  111325 non-null  int64 
dtypes: int64(4), object(3)
memory usage: 5.9+ MB
```

```
df.describe()
```

	pincode	age_0_5	age_5_17	age_18_greater
count	111325.000000	111325.000000	111325.000000	111325.000000
mean	622097.958203	1.977283	0.472733	0.015684
std	13371.428849	2.703450	1.002640	0.262118
min	517401.000000	0.000000	0.000000	0.000000
25%	608602.000000	1.000000	0.000000	0.000000
50%	625018.000000	1.000000	0.000000	0.000000
75%	635001.000000	2.000000	1.000000	0.000000
max	643253.000000	264.000000	76.000000	19.000000

```
df["date"]=pd.to_datetime(df["date"], format="%d-%m-%Y")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111325 entries, 0 to 111324
Data columns (total 7 columns):
 #   Column          Non-Null Count   Dtype         
---  ------          --------------   -----         
 0   date            111325 non-null  datetime64[ns]
 1   state           111325 non-null  object        
 2   district        111325 non-null  object        
 3   pincode         111325 non-null  int64         
 4   age_0_5         111325 non-null  int64         
 5   age_5_17        111325 non-null  int64         
 6   age_18_greater  111325 non-null  int64         
dtypes: datetime64[ns](1), int64(4), object(2)
memory usage: 5.9+ MB
```

```
df.isnull().sum()
```

	0
date	0
state	0
district	0
pincode	0
age_0_5	0
age_5_17	0
age_18_greater	0

dtype: int64

```
df.drop(columns=['state'],inplace=True)
```

```
df.sort_values(by='date',inplace=True)
```

```
df.head()
```

	date	district	pincode	age_0_5	age_5_17	age_18_greater
66622	2025-03-20	Kanchipuram	600069	12	12	10
66624	2025-04-01	Thanjavur	614602	35	10	11
44617	2025-04-01	Madurai	625017	94	10	11
22606	2025-04-01	Kancheepuram	600100	148	56	10
66623	2025-04-01	Kancheepuram	600045	81	19	14

```
df.tail()
```

	date	district	pincode	age_0_5	age_5_17	age_18_greater
19468	2026-01-03	Dindigul	624618	0	2	0
19469	2026-01-03	Erode	638402	0	1	0
19470	2026-01-03	Erode	638505	0	1	0
19461	2026-01-03	Coimbatore	641697	0	1	0
85799	2026-01-03	Viluppuram	604305	0	1	0

```
df["total_enrolment"]=(df['age_0_5']+df['age_5_17']+df['age_18_greater'])
```

```
df.head()
```

	date	district	pincode	age_0_5	age_5_17	age_18_greater	total_enrolment
66622	2025-03-20	Kanchipuram	600069	12	12	10	34
66624	2025-04-01	Thanjavur	614602	35	10	11	56
44617	2025-04-01	Madurai	625017	94	10	11	115
22606	2025-04-01	Kancheepuram	600100	148	56	10	214
66623	2025-04-01	Kancheepuram	600045	81	19	14	114

```
df['year'] = df['date'].dt.year
```

```
df.head()
```

	date	district	pincode	age_0_5	age_5_17	age_18_greater	total_enrolment	year
66622	2025-03-20	Kanchipuram	600069	12	12	10	34	2025
66624	2025-04-01	Thanjavur	614602	35	10	11	56	2025
44617	2025-04-01	Madurai	625017	94	10	11	115	2025
22606	2025-04-01	Kancheepuram	600100	148	56	10	214	2025
66623	2025-04-01	Kancheepuram	600045	81	19	14	114	2025

```
df['month'] = df['date'].dt.month
```

```
df['month_name'] = df['date'].dt.month_name()
```

```
df.head()
```

	date	district	pincode	age_0_5	age_5_17	age_18_greater	total_enrolment	year	month	month_name
66622	2025-03-20	Kanchipuram	600069	12	12	10	34	2025	3	March
66624	2025-04-01	Thanjavur	614602	35	10	11	56	2025	4	April
44617	2025-04-01	Madurai	625017	94	10	11	115	2025	4	April
22606	2025-04-01	Kancheepuram	600100	148	56	10	214	2025	4	April
66623	2025-04-01	Kancheepuram	600045	81	19	14	114	2025	4	April

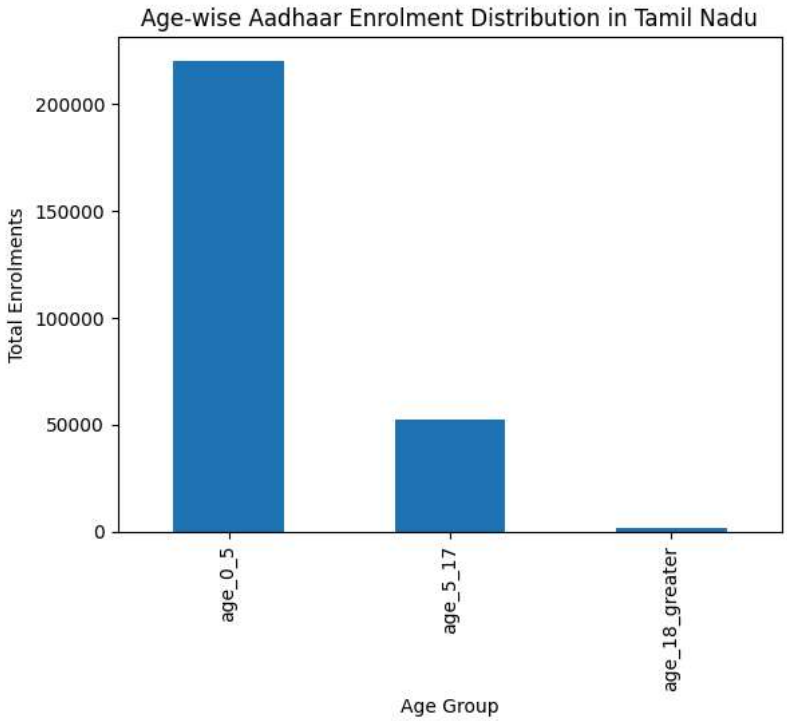
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 111325 entries, 66622 to 85799
Data columns (total 10 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   date             111325 non-null  datetime64[ns]
 1   district         111325 non-null  object
 2   pincode          111325 non-null  int64
 3   age_0_5          111325 non-null  int64
 4   age_5_17         111325 non-null  int64
 5   age_18_greater   111325 non-null  int64
 6   total_enrolment  111325 non-null  int64
 7   year             111325 non-null  int32
 8   month            111325 non-null  int32
 9   month_name       111325 non-null  object
dtypes: datetime64[ns](1), int32(2), int64(5), object(2)
memory usage: 8.5+ MB
```

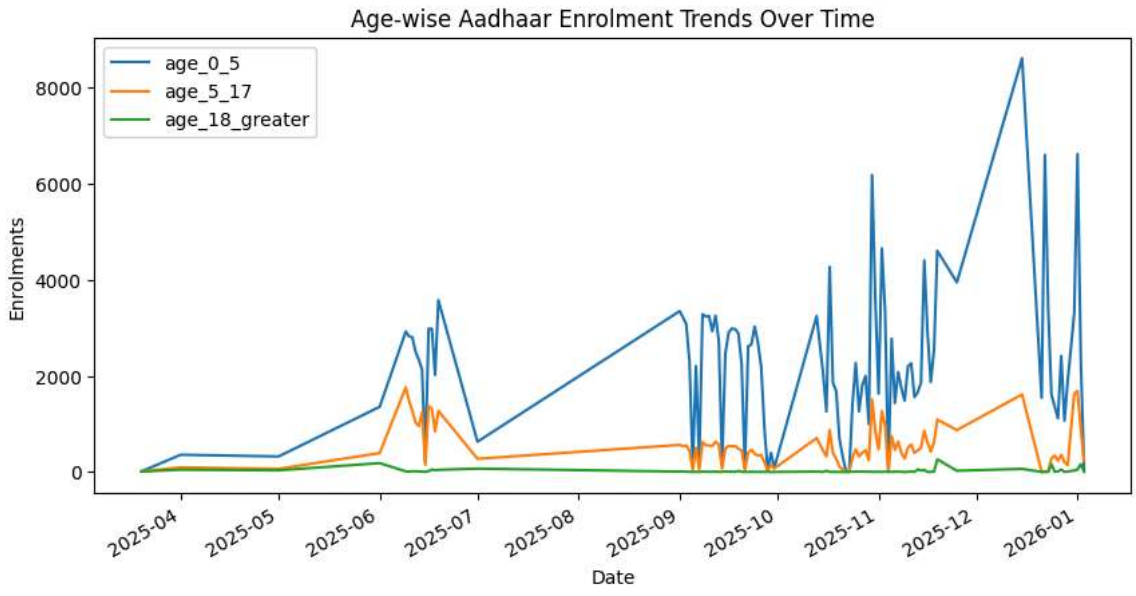
```
df[['age_0_5', 'age_5_17', 'age_18_greater', 'total_enrolment']].describe()
```

	age_0_5	age_5_17	age_18_greater	total_enrolment
count	111325.000000	111325.000000	111325.000000	111325.000000
mean	1.977283	0.472733	0.015684	2.465700
std	2.703450	1.002640	0.262118	3.393247
min	0.000000	0.000000	0.000000	1.000000
25%	1.000000	0.000000	0.000000	1.000000
50%	1.000000	0.000000	0.000000	2.000000
75%	2.000000	1.000000	0.000000	3.000000
max	264.000000	76.000000	19.000000	351.000000

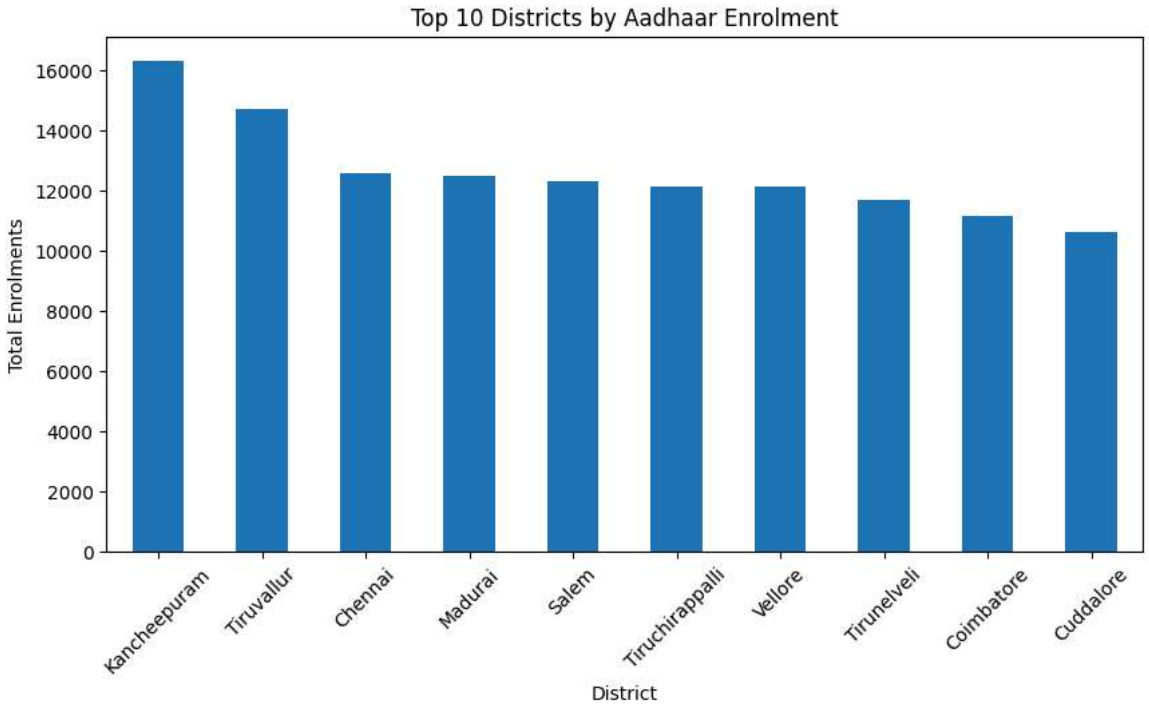
```
age_totals = df[['age_0_5', 'age_5_17', 'age_18_greater']].sum()

age_totals.plot(kind='bar')
plt.title('Age-wise Aadhaar Enrolment Distribution in Tamil Nadu')
plt.xlabel('Age Group')
plt.ylabel('Total Enrolments')
plt.show()
```



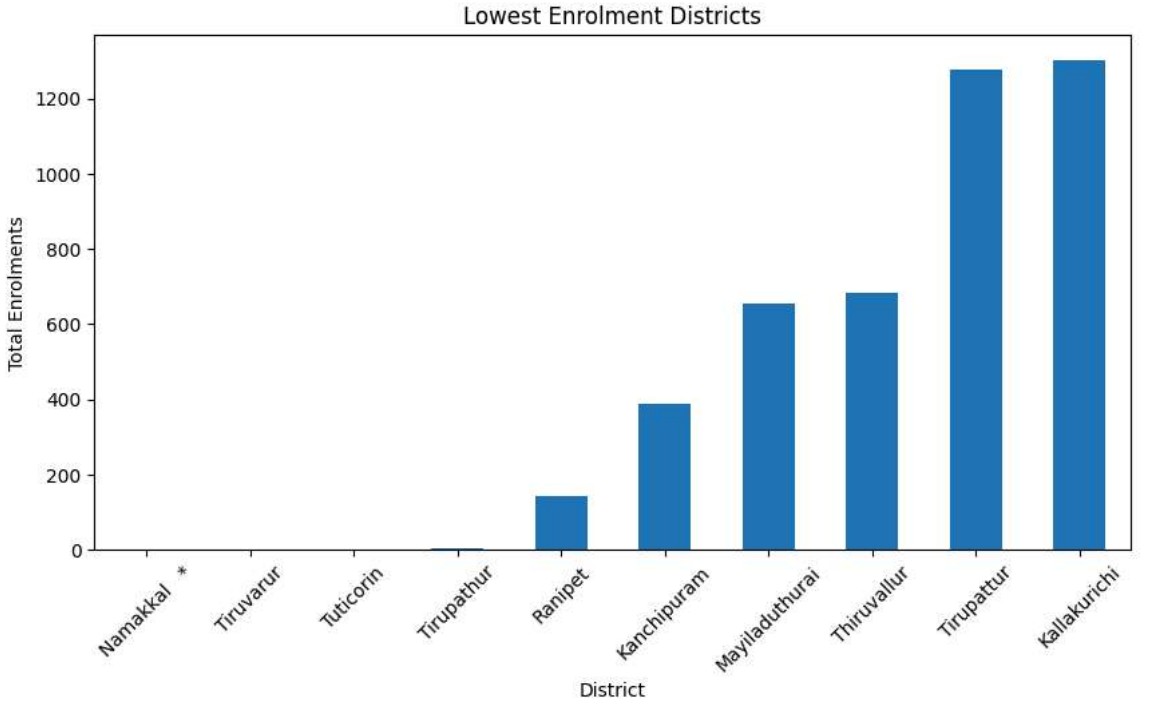

```
monthly_trend = df.groupby('date')[['age_0_5', 'age_5_17', 'age_18_greater']].sum()

monthly_trend.plot(figsize=(10,5))
plt.title('Age-wise Aadhaar Enrolment Trends Over Time')
plt.xlabel('Date')
plt.ylabel('Enrolments')
plt.show()
```



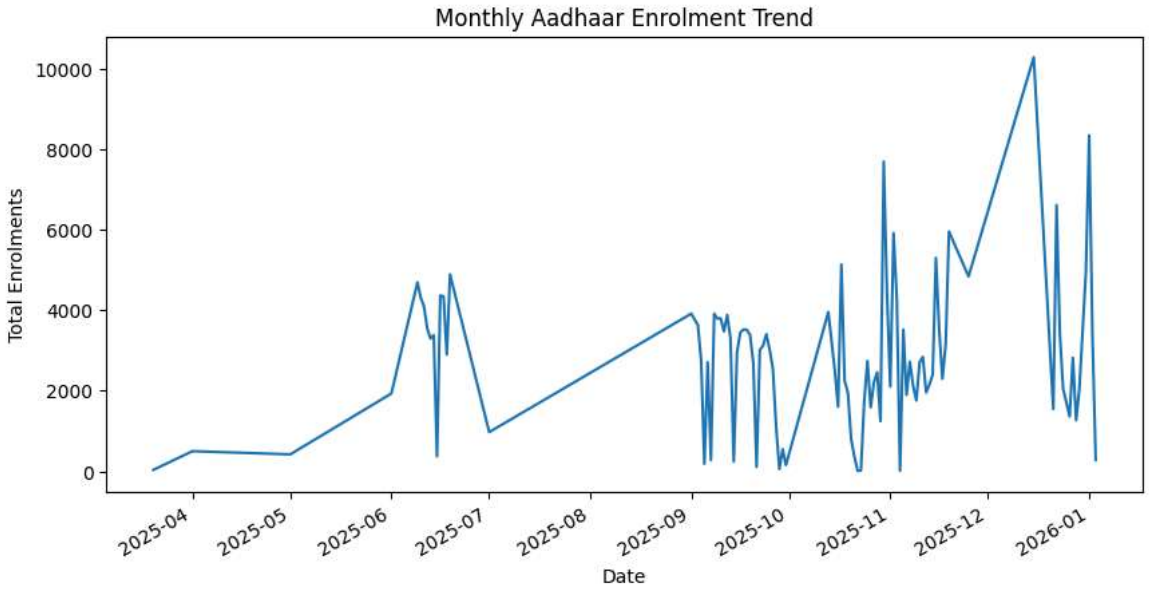

```
district_totals = df.groupby('district')['total_enrolment'].sum().sort_values(ascending=False).head(10)

district_totals.plot(kind='bar', figsize=(10,5))
plt.title('Top 10 Districts by Aadhaar Enrolment')
plt.xlabel('District')
plt.ylabel('Total Enrolments')
plt.xticks(rotation=45)
plt.show()
```




```
low_districts = df.groupby('district')['total_enrolment'].sum().sort_values().head(10)

low_districts.plot(kind='bar', figsize=(10,5))
plt.title('Lowest Enrolment Districts')
plt.xlabel('District')
plt.ylabel('Total Enrolments')
plt.xticks(rotation=45)
plt.show()
```




```
monthly_total = df.groupby('date')['total_enrolment'].sum()

monthly_total.plot(figsize=(10,5))
plt.title('Monthly Aadhaar Enrolment Trend')
plt.xlabel('Date')
plt.ylabel('Total Enrolments')
plt.show()
```

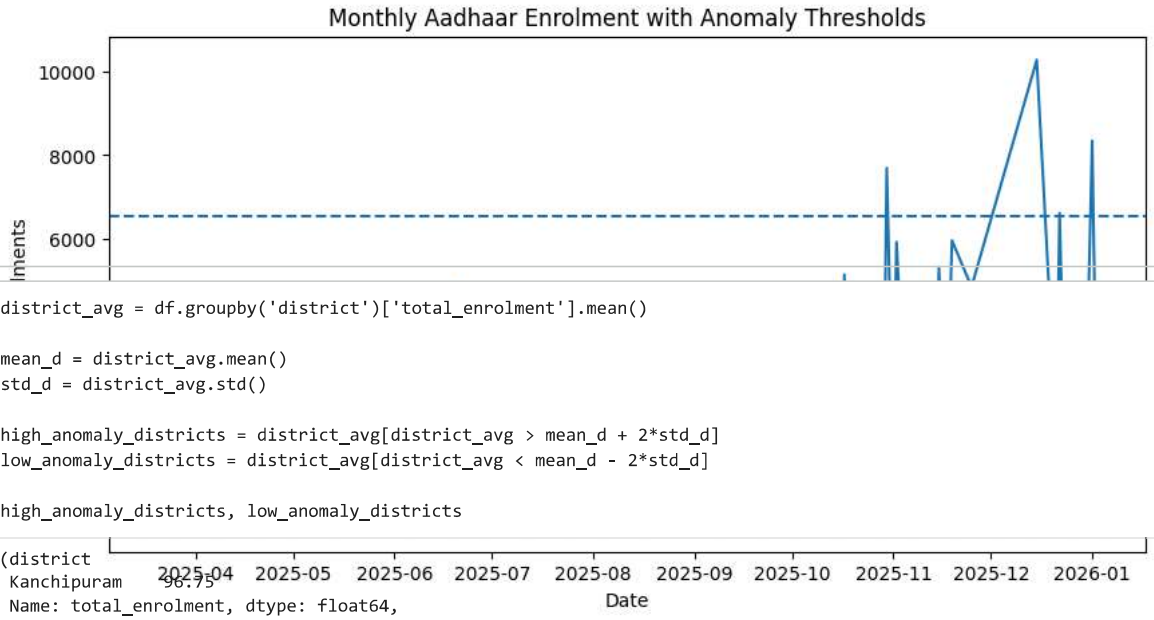



```
monthly_total = df.groupby('date')['total_enrolment'].sum()

mean_val = monthly_total.mean()
std_val = monthly_total.std()

upper_threshold = mean_val + 2 * std_val
lower_threshold = mean_val - 2 * std_val

anomalies = monthly_total[
    (monthly_total > upper_threshold) |
    (monthly_total < lower_threshold)
]

anomalies
```

date	total_enrolment
2025-10-30	7690
2025-12-15	10283
2025-12-22	6611
2026-01-01	8342

dtype: int64

```
plt.figure(figsize=(10,5))
plt.plot(monthly_total)
plt.axhline(upper_threshold, linestyle='--')
plt.axhline(lower_threshold, linestyle='--')
plt.title('Monthly Aadhaar Enrolment with Anomaly Thresholds')
plt.xlabel('Date')
plt.ylabel('Total Enrolments')
plt.show()
```




```
district_avg = df.groupby('district')['total_enrolment'].mean()

mean_d = district_avg.mean()
std_d = district_avg.std()

high_anomaly_districts = district_avg[district_avg > mean_d + 2*std_d]
low_anomaly_districts = district_avg[district_avg < mean_d - 2*std_d]

high_anomaly_districts, low_anomaly_districts
```

```
(district
 Kanchipuram    96.75
 Name: total_enrolment, dtype: float64,
 Series([], Name: total_enrolment, dtype: float64))
```

```
age_means = df[['age_0_5', 'age_5_17', 'age_18_greater']].mean()

age_means
```

	0
age_0_5	1.977283
age_5_17	0.472733
age_18_greater	0.015684

dtype: float64

```
district_median = df.groupby('district')['total_enrolment'].median()

low_activity_districts = district_median[district_median < 1]

low_activity_districts
```

district	total_enrolment

dtype: float64

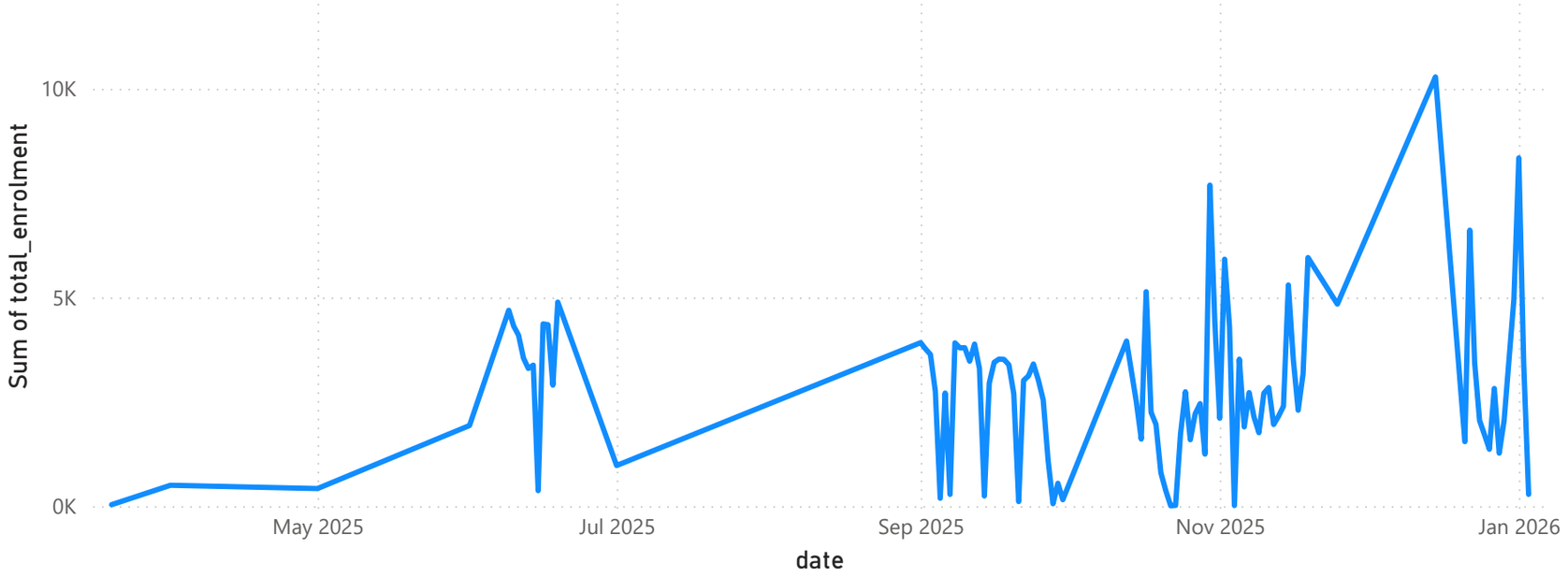
```
df.to_csv('tamilnadu_uidai_cleaned.csv', index=False)
```

Total aadhar enrolment

274K

Sum of total_enrolment

Aadhar enrolment trend over time

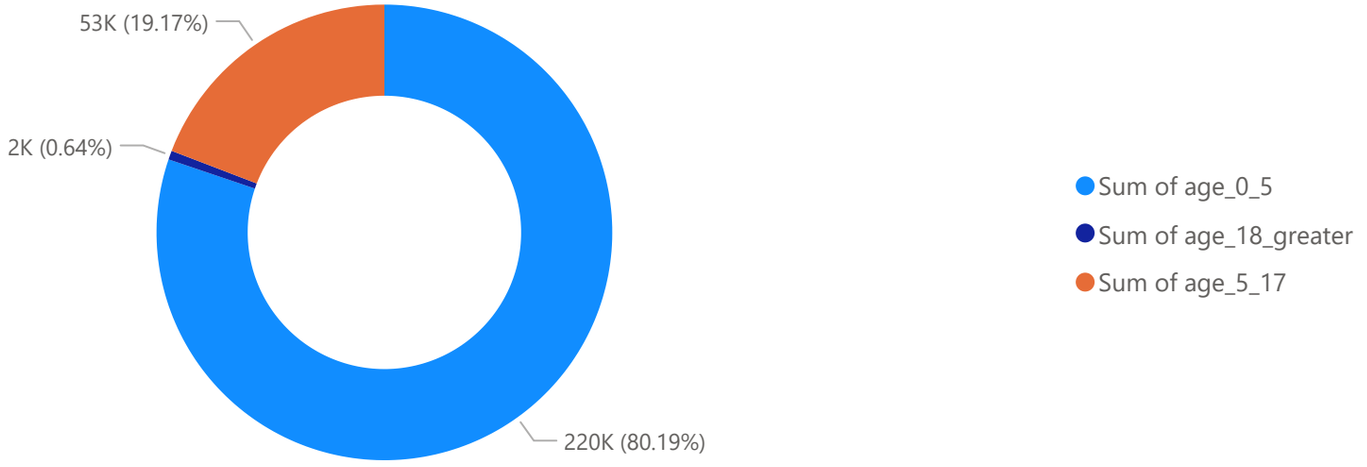


Number of districts

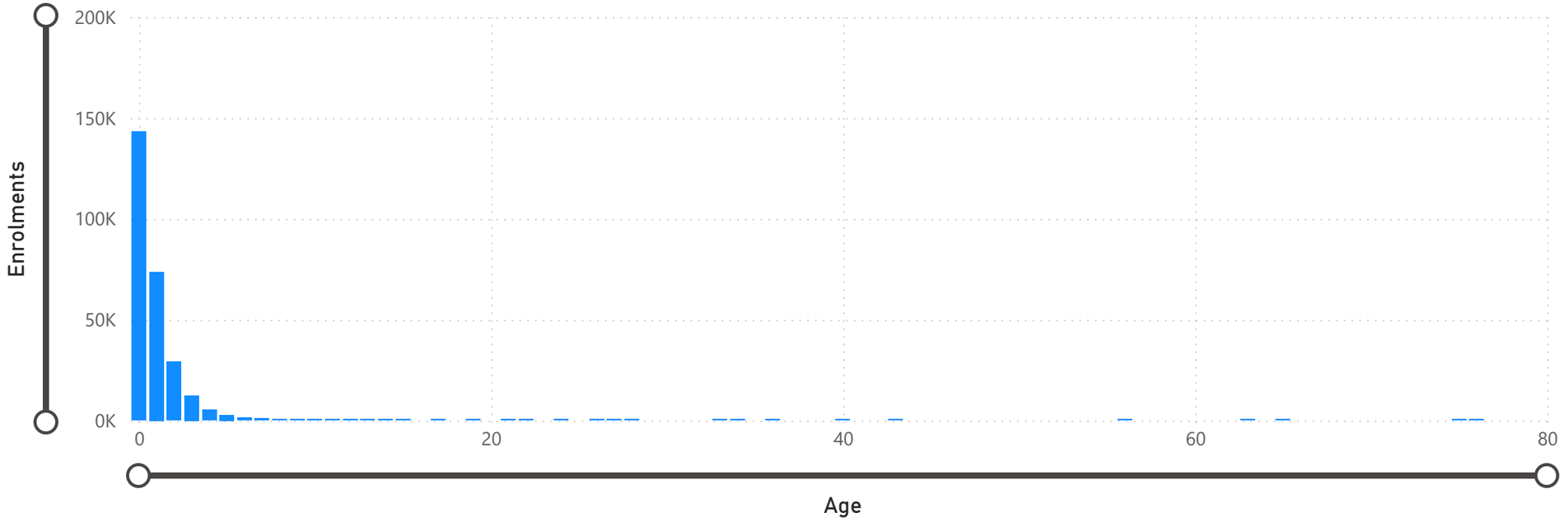
46

Count of district

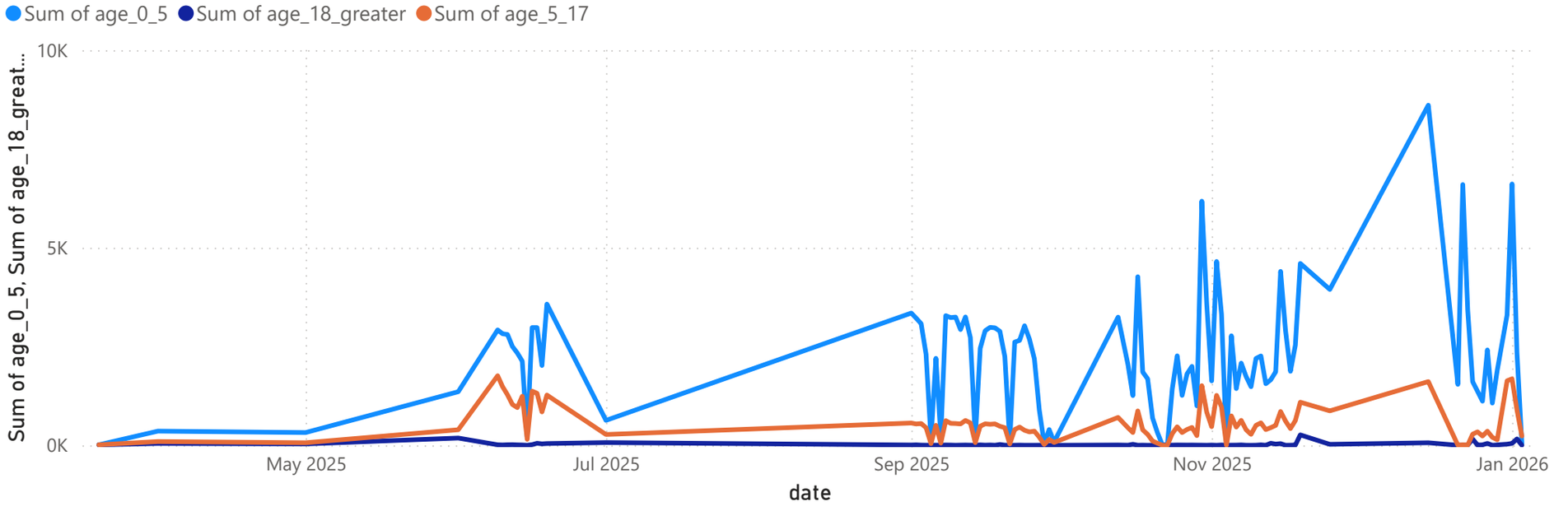
Age wise enrolment distribution



Enrolment by age group



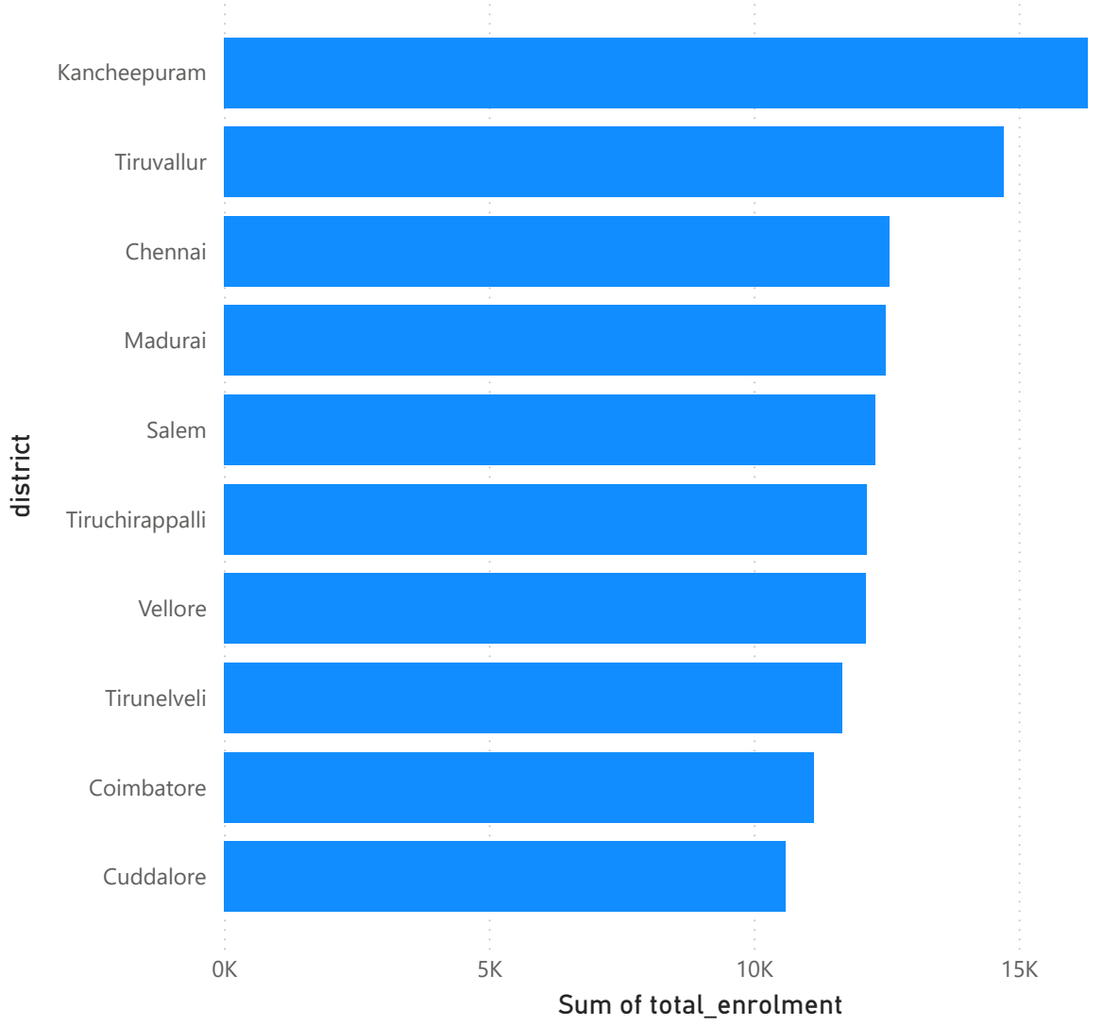
Age wise enrolment trend



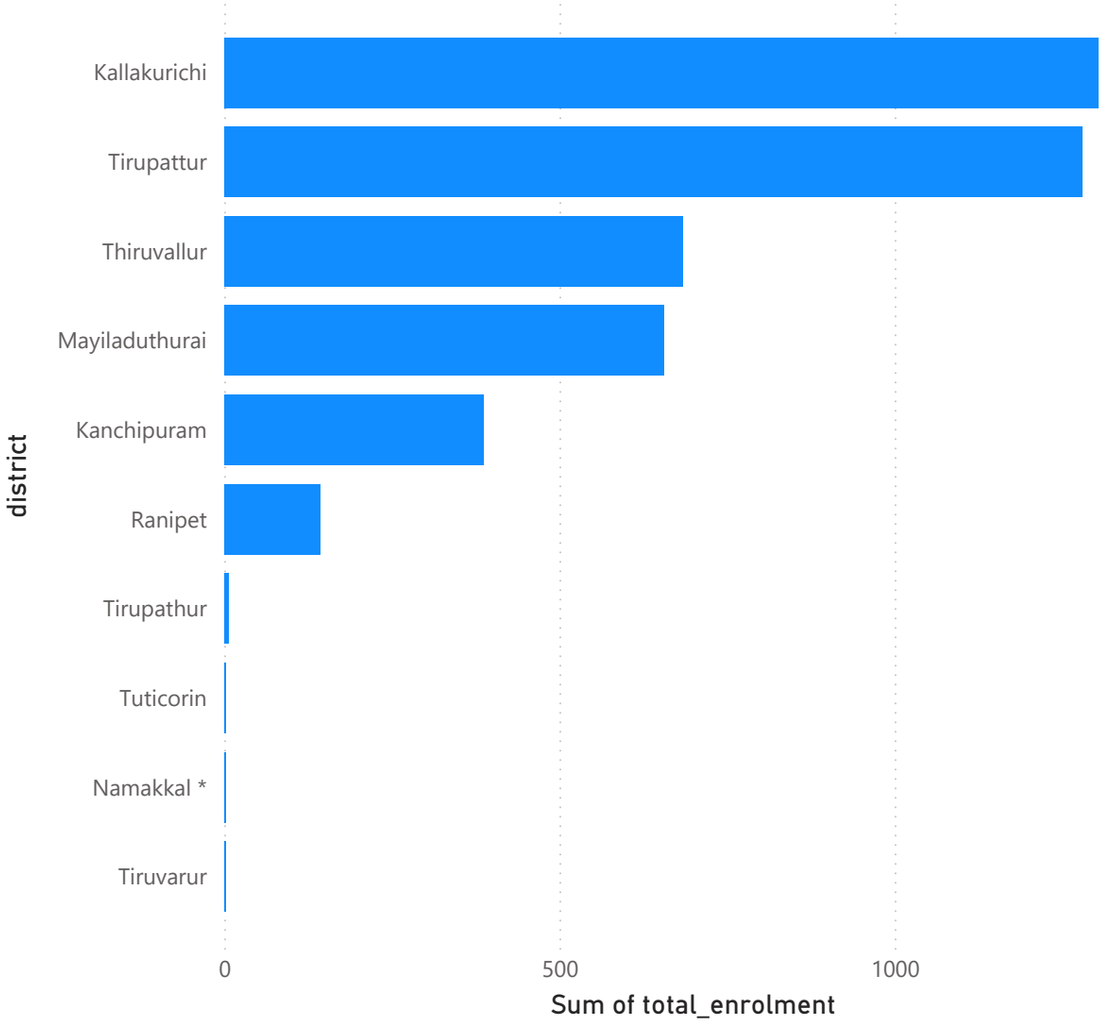
Month

January	February
March	April
May	June
July	August
September	October
November	December

Top 10 districts by enrolment



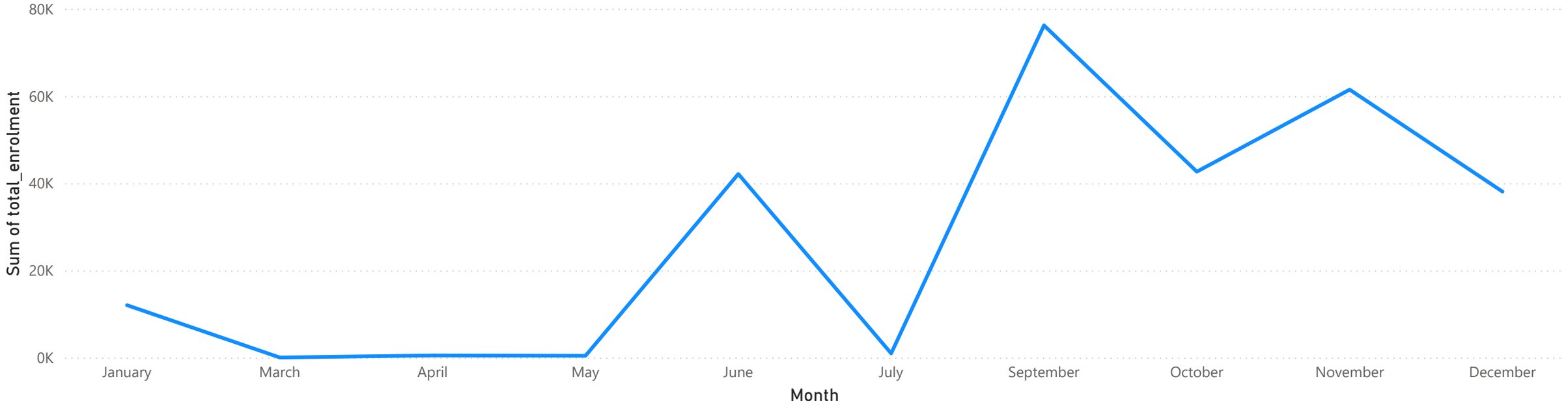
Bottom 10 districts by enrolment



district

Ariyalur	Chengalpattu	Chennai	Coimbatore	Cuddalore	Dharmapuri	Dindigul	Erode	Kallakurichi
Kancheepuram	Kanchipuram	Kanniyakumari	Kanyakumari	Karur	Krishnagiri	Madurai	Mayiladuthurai	Nagapattinam

Monthly aadhar enrolment



Year wise enrolment comparison