

UIDAI DATA HACKATHON 2026

Team Id - UIDAI_1336

1. Problem Statement :

The challenge lies in extracting meaningful patterns, trends, anomalies, and predictive indicators from the UIDAI dataset and translating these findings into actionable insights that enable data-driven decision-making and support systematic improvements in service delivery and operational efficiency.

2. Datasets Used:

Source: The analysis utilizes the Aadhaar Enrolment and Update dataset provided by the Unique Identification Authority of India (UIDAI).

Data Structure: The merged dataset schema:

Dataset Name	Columns in dataset	Total rows in dataset
Enrolment	Date, State, District, Pincode, age_0_5, age_5_17, age_18_greater	1,006,029
Demographic	Date, State, District, Pincode, demo_age_5_17, demo_age_17_	1,861,108
Biometric	Date, State, District, Pincode, bio_age_5_17, bio_age_17_	2,071,698

3. Methodology:

3.1 Tools Used:

- Data preprocessing, Exploratory Data Analysis (EDA), and visualizations were performed using **Python** and its libraries

within **Google Colab**. For finding insights, **MySQL** was utilized via **MySQL Workbench**.

3.2 Preprocessing:

- **Concatenating data:** The raw data was split into multiple batches (in .csv). These were loaded individually and concatenated along the axis (vertical) to create a consolidated dataframe.
- **Normalization:** The **state** and **district** column was normalized by converting all values to string format, removing extra whitespaces, consolidating multiple spaces into single spaces, standardizing various dash types to hyphens, removing special characters (except spaces, hyphens, and parentheses) and standardizing the text to title case. The cleaned values were stored in **state_clean** and **district_clean** to maintain data integrity.
- **Standardization:** A standardization dictionary was applied to correct naming inconsistencies in the **state_clean** and **district_clean** column, including replacing ampersands with "And", updating deprecated names (Orissa to Odisha, Pondicherry to Puducherry), and fixing spelling errors (West Bangal/Westbengal to West Bengal). This ensures uniform state nomenclature across the dataset for accurate analysis. "100000" refers to Lakshadweep, a correlation readily found through an internet search.
- **Data Type conversion:** The date column, which was initially represented as a string, was successfully converted to the date data type.
- **Duplicates :** Upon the identification of duplicate entries, the initial occurrence of a row is designated as **is_duplicate = 0**, signifying its status as a unique record. Subsequent rows that are identical (duplicates) are consequently marked as **is_duplicate = 1**. *For further analysis we took the rows where is_duplicate value is 0(unique rows).*

- **Regional Office(RO)** : We have added a new column titled Region, which maps the States and Union Territories covered by each RO. Using the reference [link](https://uidai.gov.in/en/contact-support/regional-offices.html) (source: <https://uidai.gov.in/en/contact-support/regional-offices.html>).

3.3 Exploratory Data Analysis(EDA):

- **Null Value Exploration:** We checked the three datasets and found no missing (null) values in all the columns.
- **Data Type Exploration:** We also reviewed the data types for each column to ensure they were correct for our analysis.
- **Year/Month Analysis:** We checked the dataset to determine the total number of years and months covered by the datasets. This data covers year 2025 and 9 months. Out of 12 months, 3 months data is missing(Jan, Feb, Aug).
- **Duplicate Rows Analysis:** We used the `is_duplicate` column to count the number of duplicate rows in all the datasets.

Dataset Name	Number of Duplicate Rows	Percentage Duplicity rate
Enrolment	22,957	2.28%
Biometric	4,73,601	22.86%
Demographic	94,896	5.09%

- **State & District Analysis:** We analyzed the dataset to identify the number of unique states and districts. The data covers **38 states** and **910 districts**.
- **Pincode Analysis:** Similarly, we verified the number of unique pincodes from the data. There are **19463** pincodes in the data.

3.4 Insights and Visualization:

3.4.1 Enrolment Dataset:

- Total Enrolments given in data: 5,331,760
- State wise trend analysis:

Overall Enrolments :

Rank	Top States	Bottom States
1	Uttar Pradesh: 1,002,631	Daman And Diu: 134
2	Bihar: 593,753	Lakshadweep: 413
3	Madhya Pradesh: 487,892	Andaman And Nicobar Islands: 501

Enrolments in age (0-5) :

Rank	Top States	Bottom States
1	Uttar Pradesh: 511,727	Daman And Diu: 120
2	Bihar: 363,244	Lakshadweep: 188
3	Madhya Pradesh: 274,274	Ladakh: 466

Observation : The top three states for child enrolments match the overall top three, as do the bottom two. Notably, Ladakh is the third lowest state for child enrolments.

Enrolments in age (5-17) :

Rank	Top States	Bottom States
1	Uttar Pradesh: 473,205	Lakshadweep : 11
2	Bihar: 327,043	Daman And Diu: 14
3	Madhya Pradesh: 115,172	Andaman and Nicobar Islands: 32

Observation : The top three states for child enrolments match the overall top three.

Enrolments in age (18+) :

Rank	Top States	Bottom States
1	Meghalaya : 35,078	Daman And Diu: 0
2	Assam : 22,555	Andaman and Nicobar Islands : 0
3	Uttar Pradesh: 17,699	Dadar And Nagar Haveli: 4

Observation : The dominance of Meghalaya and Assam in the 18+ category is a significant anomaly. Unlike central India, where enrolments are driven by birth (0-5 age group), these North-Eastern states exhibit a 'Late Enrolment' trend.

Recommendation: UIDAI should investigate if this is driven by distinct administrative requirements (e.g., NRC related documentation) or recent saturation drives in border districts.

- **District wise trend analysis:**

Overall Enrolments :

Rank	Top District	Bottom District
1	Thane : 43,142	Bagpat : 1
2	Sitamarhi : 41,652	Balotra : 1
3	Bahraich : 38,897	Beawar : 1

Observation : High enrolment rates are observed in districts situated near the Bangladesh border, specifically Murshidabad, North 24 Parganas, and South 24 Parganas.

Enrolments in age (0-5) :

Rank	Top District	Bottom District
1	Murshidabad : 30,593	Lakshadweep : 0
2	Thane : 28,692	Bagpat : 0
3	Bengaluru : 25,845	Jharjar : 0

Enrolments in age (5-17) :

Rank	Top District	Bottom District
1	Bahraich : 22,132	Balotra : 0
2	Gaya : 19,786	Bandipur : 0
3	Sitamarhi : 18,600	Bardez : 0

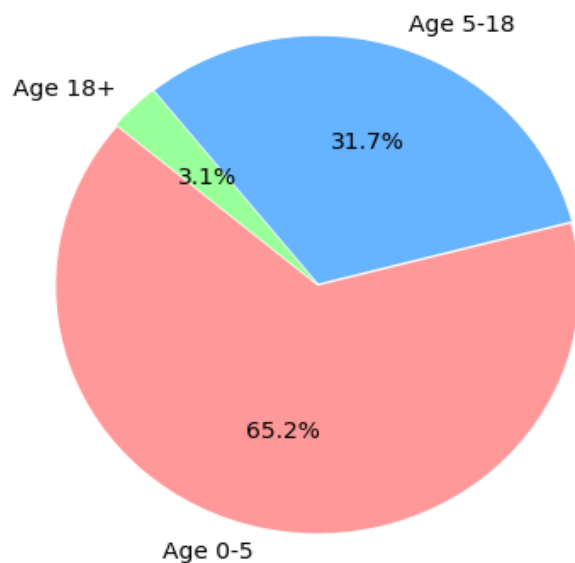
Enrolments in age (18+) :

Rank	Top District	Bottom District
1	East Khasi Hills : 9,839	Anthapur : 0
2	West Khasi Hills : 5,259	Ananthapur : 0
3	West Garo Hills : 4,532	Andamans : 0

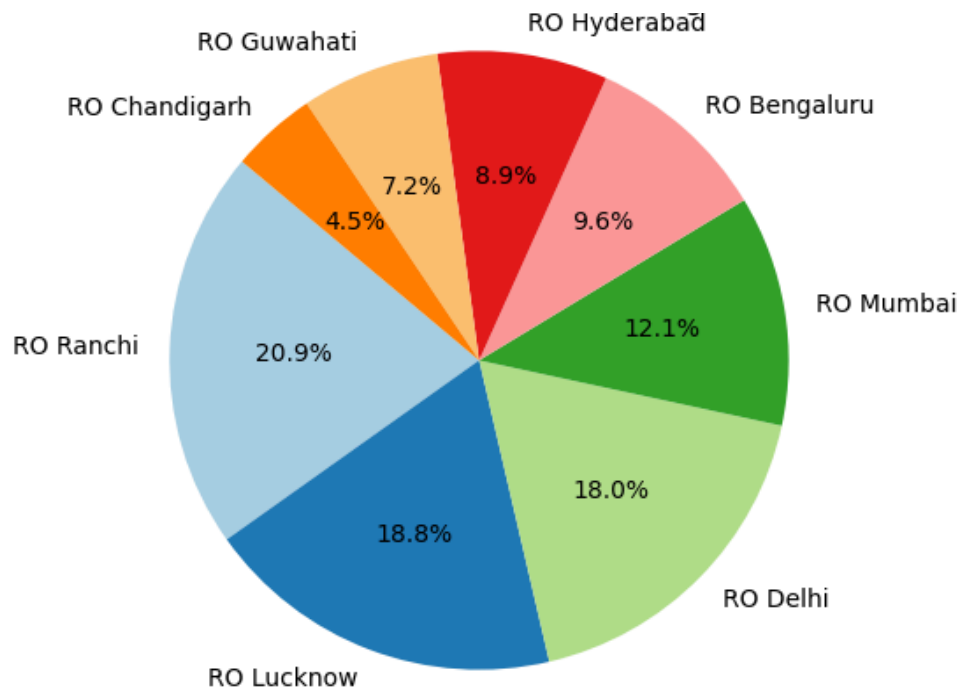
Observation : unexpected that bengaluru and bengaluru urban are in top 10 districts for 18+ enrolments and covers 4.58% of total 18+ enrolments. Notably, 5 of top 10 districts for enrolments above 18+ are from North East states.

- Age wise analysis:

Total Enrolments Distribution by Age Group

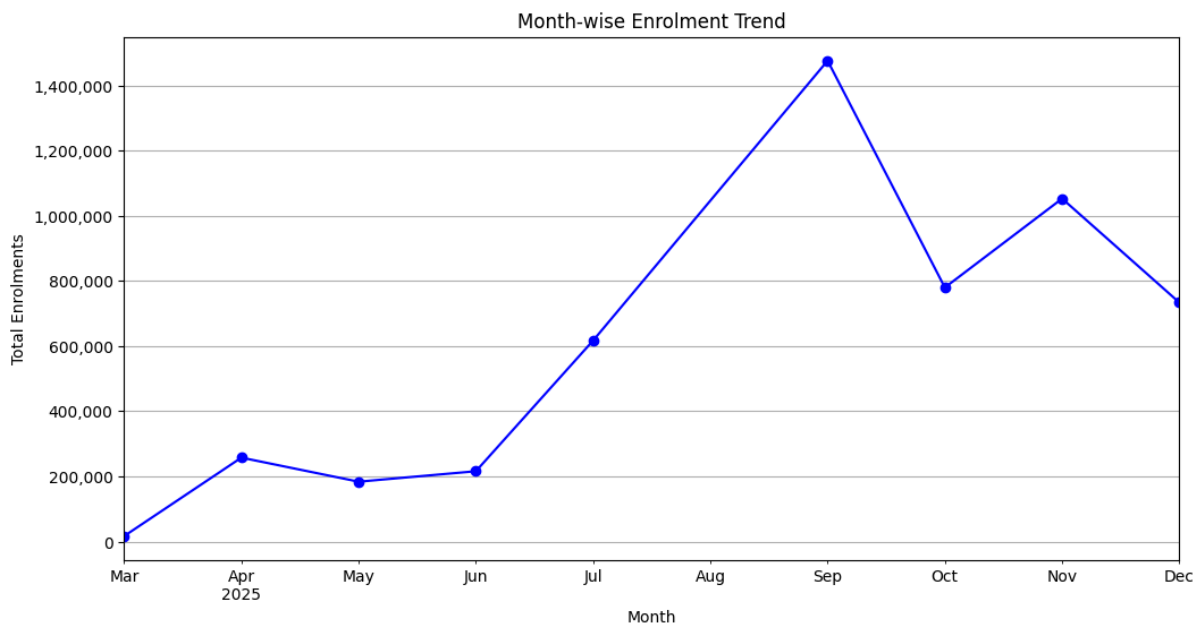


- Regional Office wise analysis:



Distribution of states/ UTs across UIDAI Regional Offices

- Month Wise Enrolment Trend:



3.4.1 Biometric Dataset:

- Total Biometric Updates given in data: 68,261,059
- State wise trend analysis:

Overall Updates:

Rank	Top States	Bottom States
1	Uttar Pradesh: 9,367,083	Dadar and nagar Haveli And Daman And Diu: 2,507
2	Maharashtra: 9,020,710	Lakshadweep: 4,745
3	Madhya Pradesh: 5,819,736	Ladakh: 5,659

Biometric Updates in age (5-17) :

Rank	Top States	Bottom States
1	Uttar Pradesh: 6,076,420	Dadar and nagar Haveli And Daman And Diu: 747
2	Maharashtra: 3,437,083	Lakshadweep: 2195
3	Madhya Pradesh: 3,148,670	Ladakh: 2752

Biometric Updates in age (18+):

Rank	Top States	Bottom States
1	Maharashtra : 5,583,627	Dadar and nagar Haveli And Daman And Diu: 1760
2	Uttar Pradesh : 3,290,663	Lakshadweep: 2550
3	Madhya Pradesh : 2,671,066	Ladakh : 2907

Observation : As the maximum candidates are from Maharashtra and Uttar Pradesh , so we can say that they have maximum Biometric Updates.

- **District wise trend analysis:**

Overall Biometric Updates:

Rank	Top District	Bottom District
1	Pune: 593,592	Banipur : 1
2	Nasik: 561,893	Chandauli : 1
3	Thane: 559,311	Gautam Buddha Nagar : 1

Biometric Updates in age (5-17) :

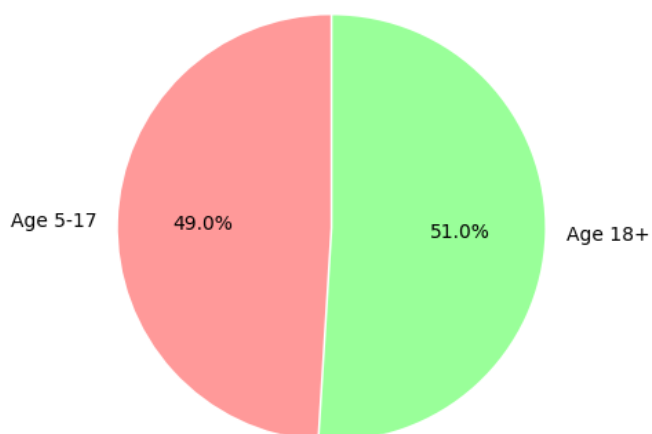
Rank	Top District	Bottom District
1	Pune: 277,443	Akhera: 0
2	Kurnool : 237,941	Gautam Buddha Nagar : 0
3	Nashik : 208,401	Hingoli : 0

Biometric Updates in age (18+) :

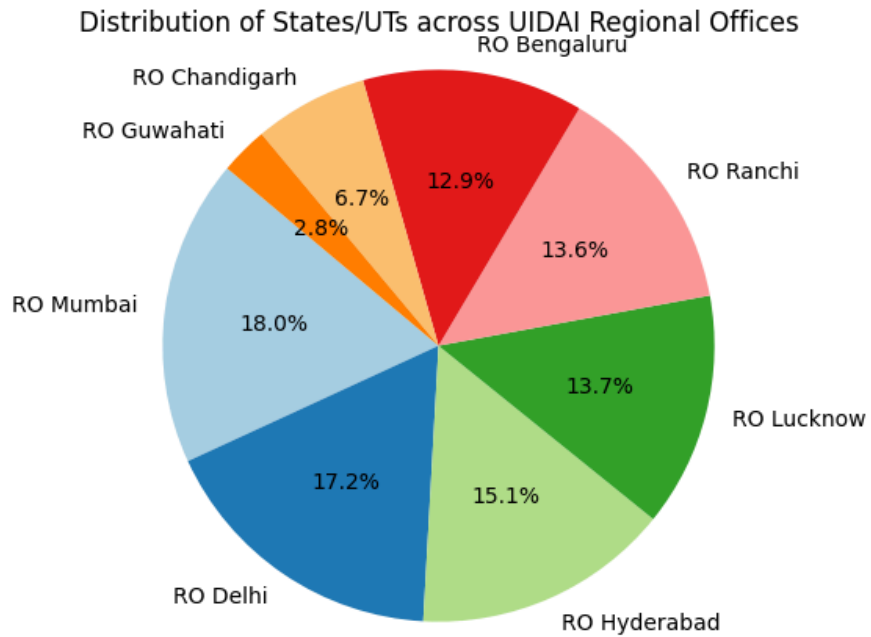
Rank	Top District	Bottom District
1	Thane: 361,166	Auriya : 0
2	Nashik : 353,492	Pherzawl : 0
3	Pune: 316,149	Poonch : 0

- **Age wise analysis:**

Total Updates Distribution by Age Group

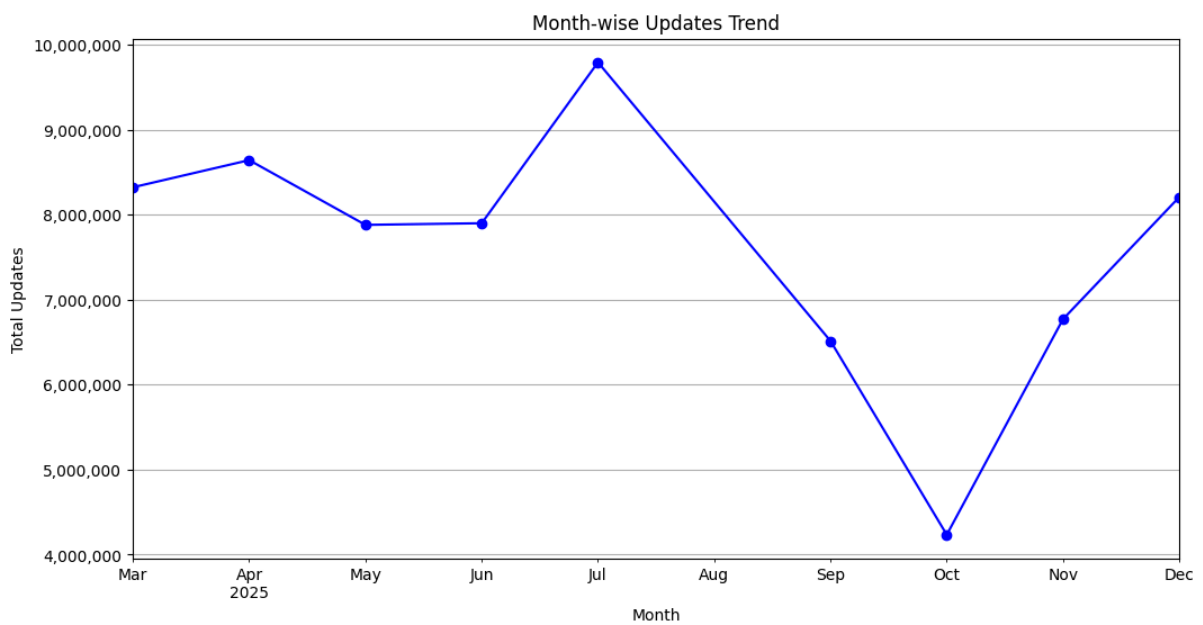


- Regional Office wise analysis:



Distribution of states/ UTs across UIDAI Regional Offices

- Month Wise Biometric Trend:



3.4.3 Demographic Dataset:

- Total Demographic Updates given in data: 36,597,559
- State wise trend analysis:

Overall Updates:

Rank	Top States	Bottom States
1	Uttar Pradesh: 6,460,511	Lakshadweep : 915
2	Maharashtra: 3,824,892	Daman and Diu : 1735
3	Bihar: 3,638,845	Dadar and Nagar Haveli and Daman and Diu : 2768

Demographic Updates in age (5-17) :

Rank	Top States	Bottom States
1	Uttar Pradesh : 603,453	Lakshadweep : 131
2	Madhya Pradesh : 289,711	Daman And Diu : 211
3	Bihar : 285,883	Dadar And Nagar Haveli and Daman And Diu : 227

Demographic Updates in age (18+):

Rank	Top States	Bottom States
1	Uttar Pradesh : 5,857,058	Lakshadweep : 784
2	Maharashtra : 3620599	Daman And Diu : 1524
3	Bihar : 3,352,962	Dadar And Nagar Haveli and Daman And Diu : 2,541

- District wise trend analysis:

Overall Demographic Updates:

Rank	Top District	Bottom District
1	Thane: 326749	5Th Cross : 1

2	Pune: 326083	Bhadrak(R): 1
3	South 24 Pargans: 300121	Udupi: 1

Demographic Updates in age (5-17) :

Rank	Top Districts	Bottom Districts
1	Bengaluru: 30756	Bandipur : 0
2	North West Delhi: 27972	Medchal-malkajgiri : 0
3	Hyderabad: 26280	Mohali : 0

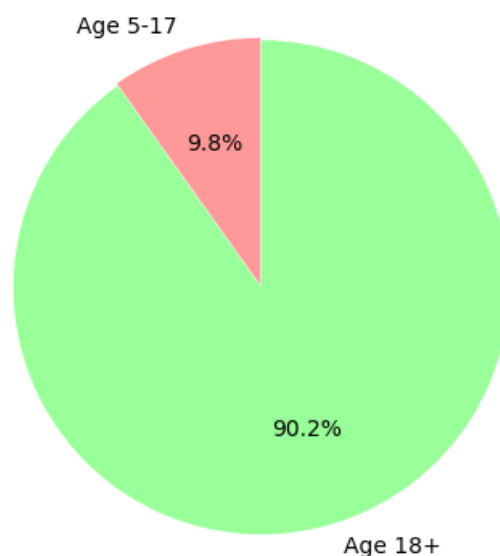
Demographic Updates in age (18+) :

Rank	Top Districts	Districts
1	Pune : 301944	Chandauli : 0
2	Thane : 300816	5Th Cross : 1
3	South 24 Parganas : 279466	Chitrakoot : 1

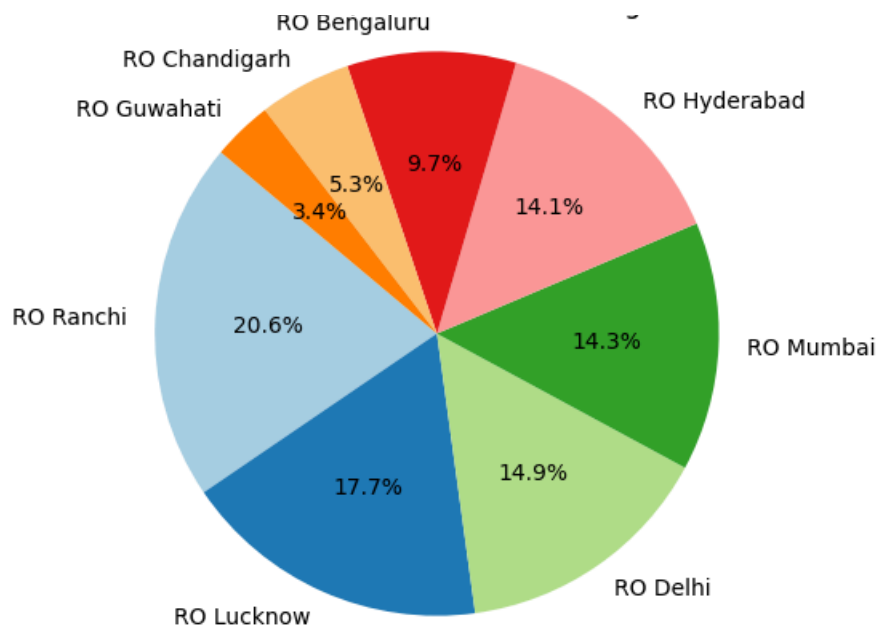
Observation : The top 5 districts for Demographic updates (Pune, Nashik, Bengaluru, Ahmedabad) are major industrial/IT hubs, suggesting updates are driven by workforce migration.

- **Age wise analysis:**

Total Updates Distribution by Age Group

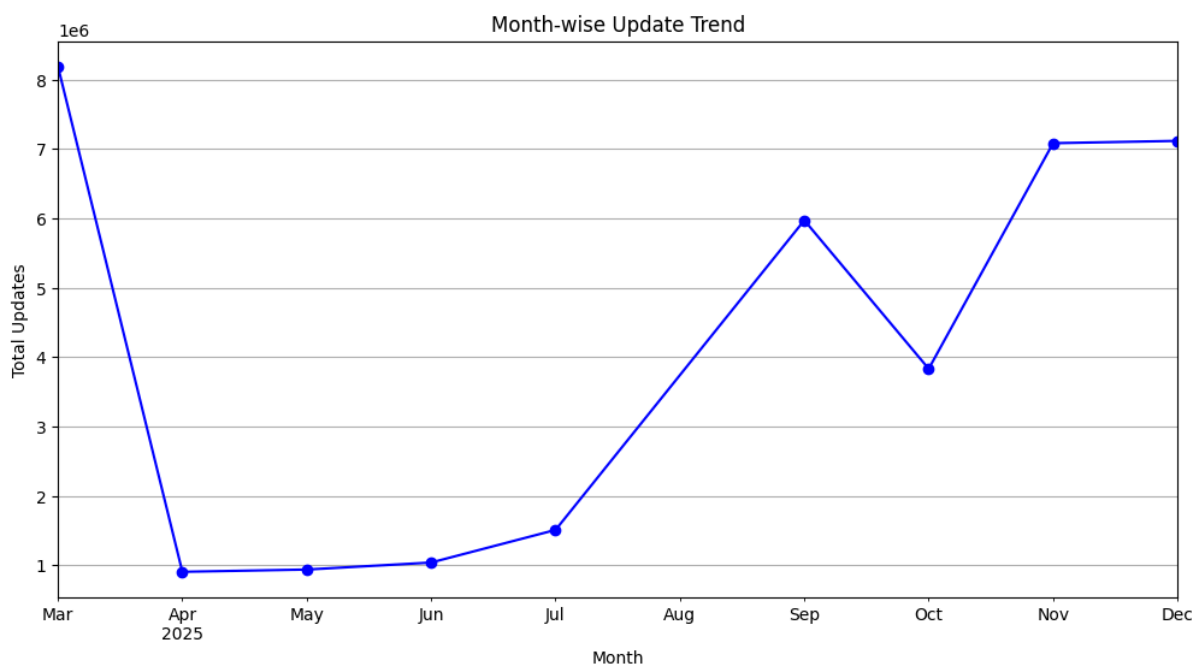


- Regional Office wise analysis:

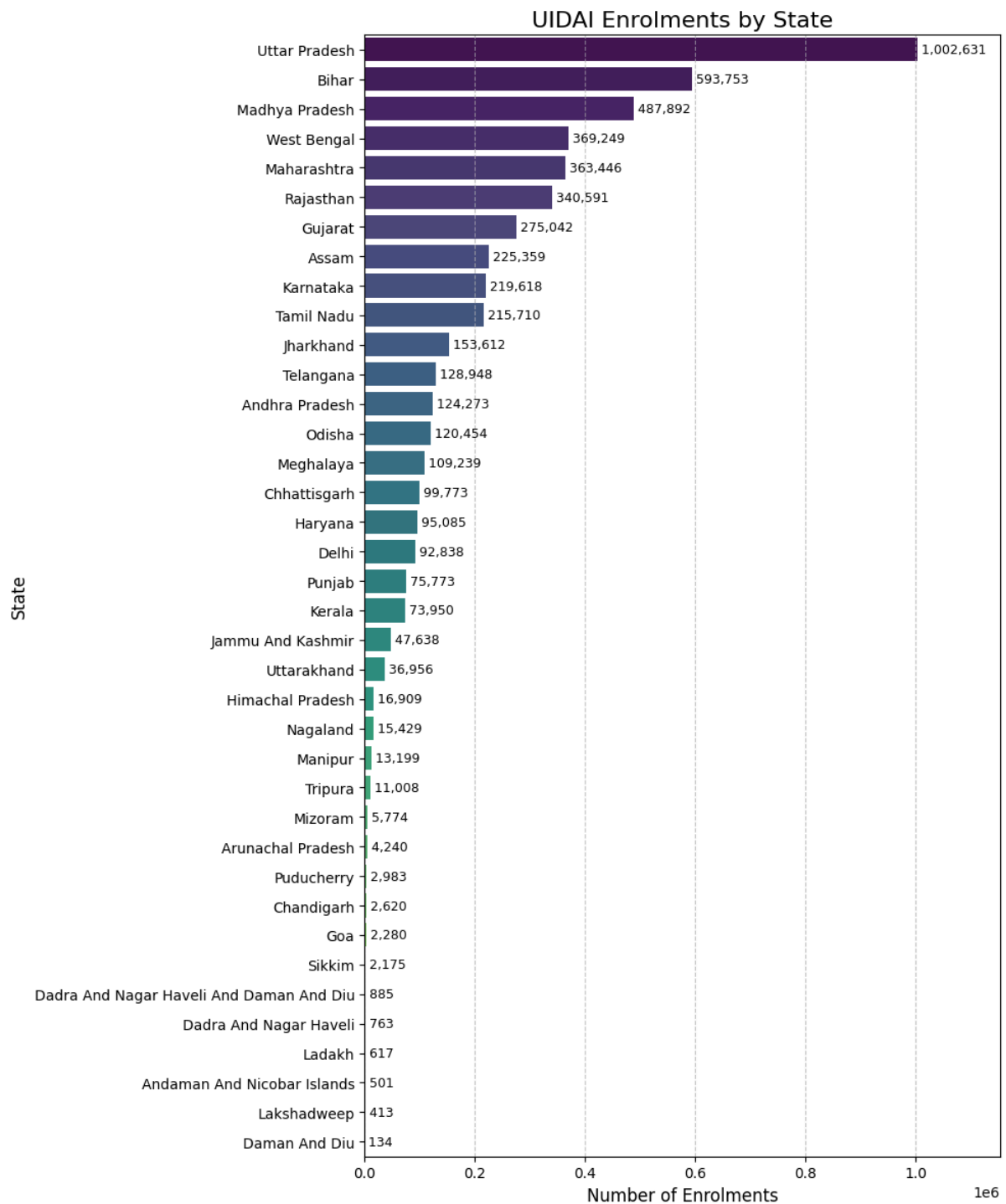


Distribution of states/ UTs across UIDAI Regional Offices

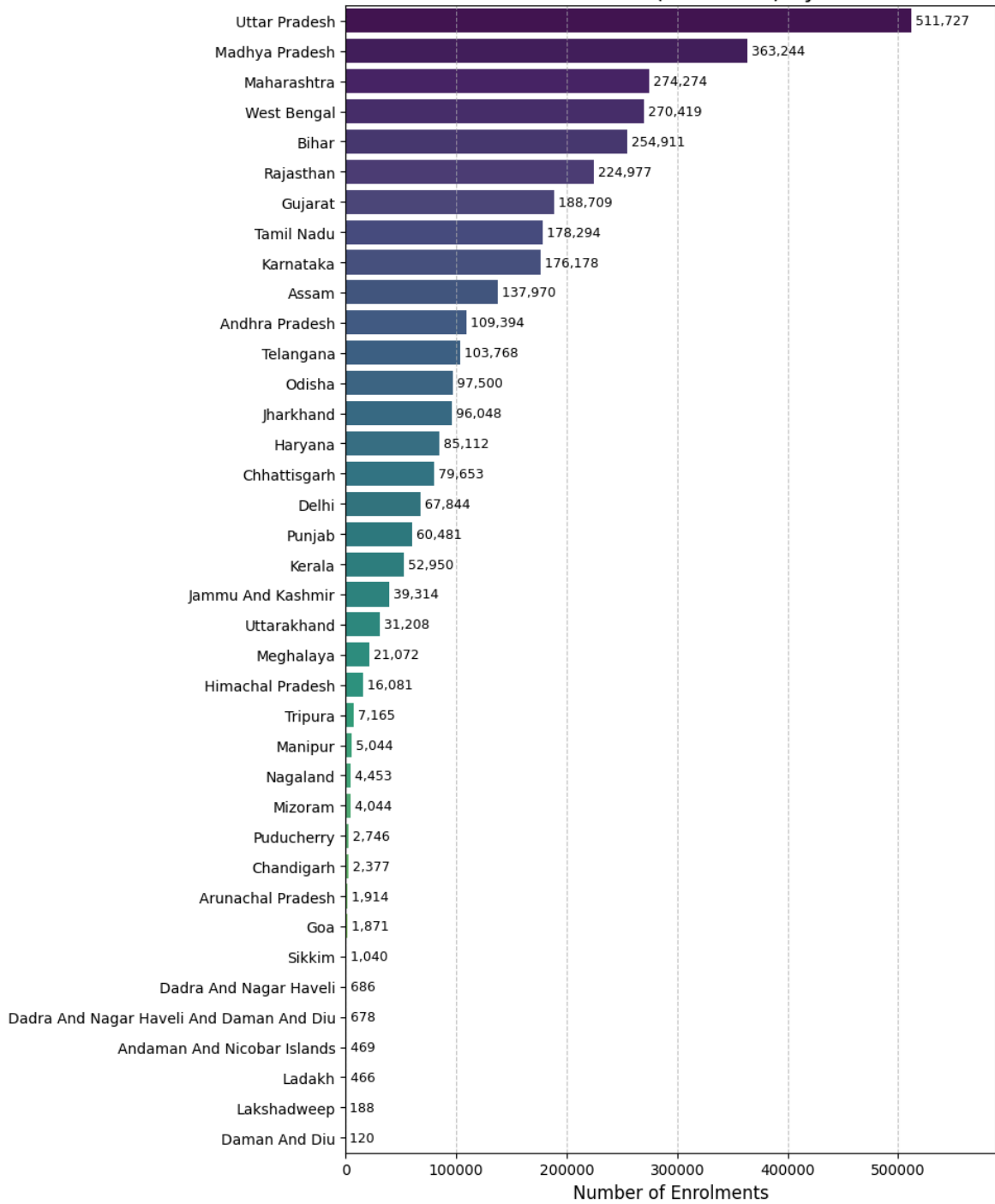
- Month Wise Demographic Trend:

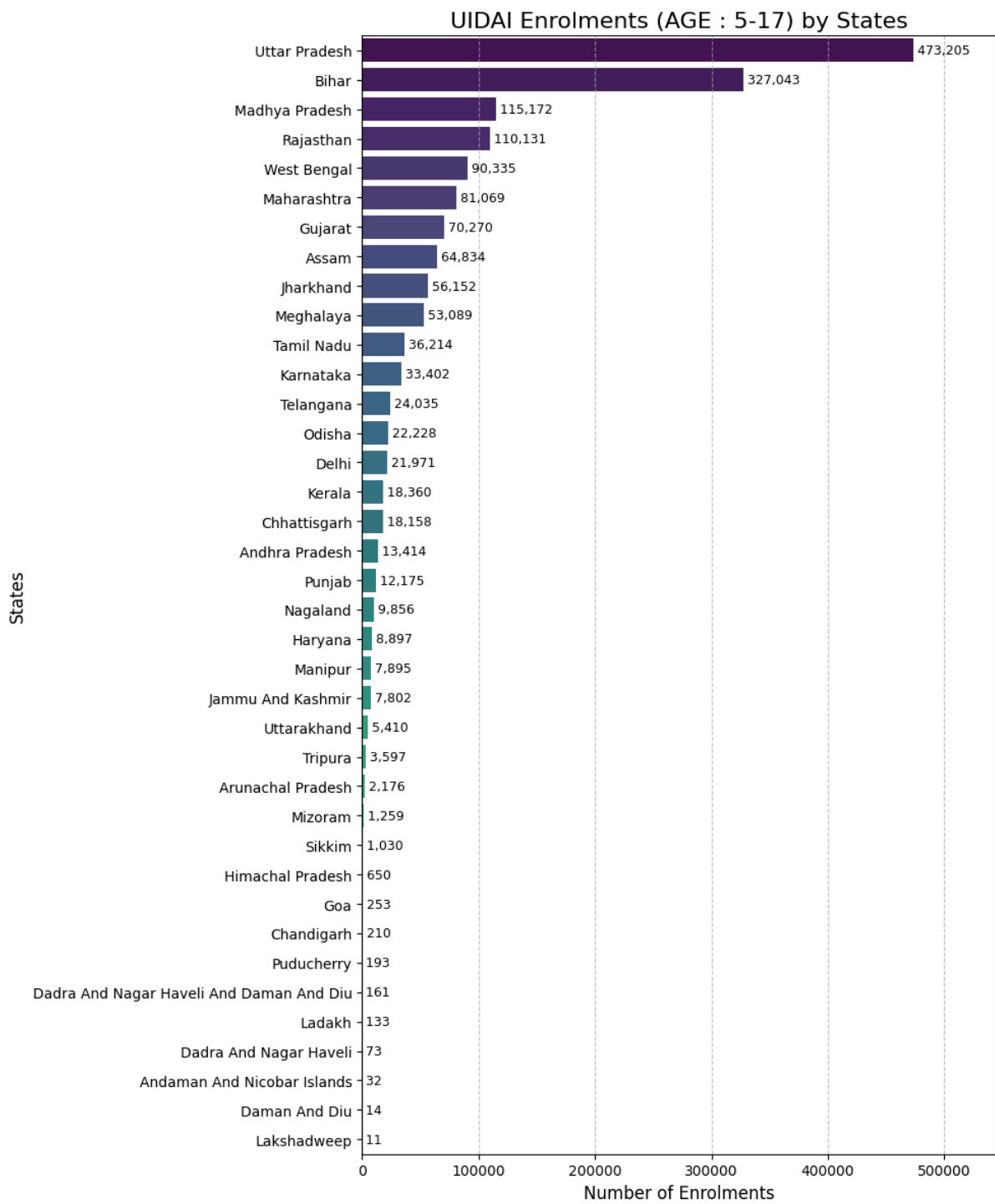


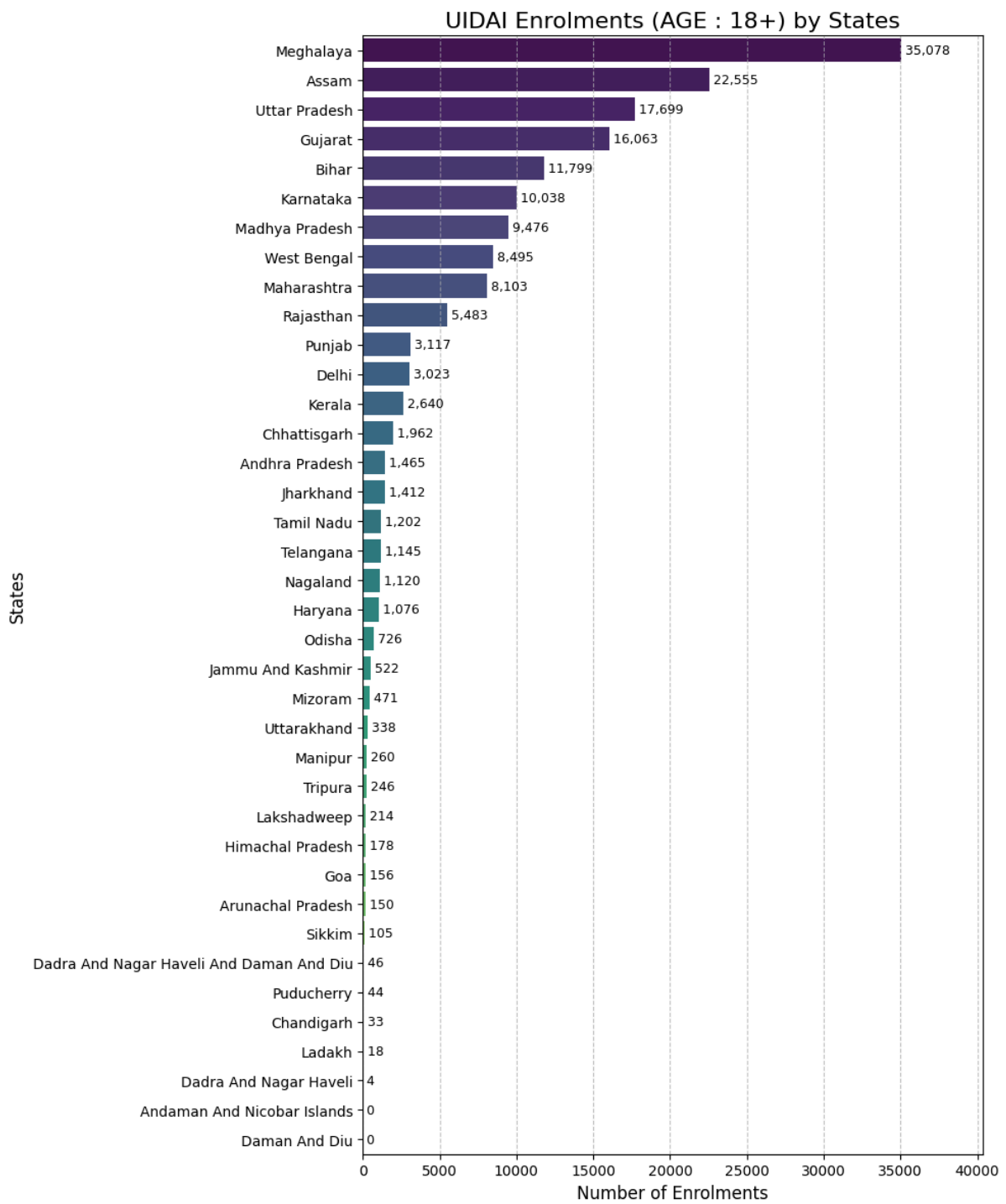
Enrolment graphs for reference:

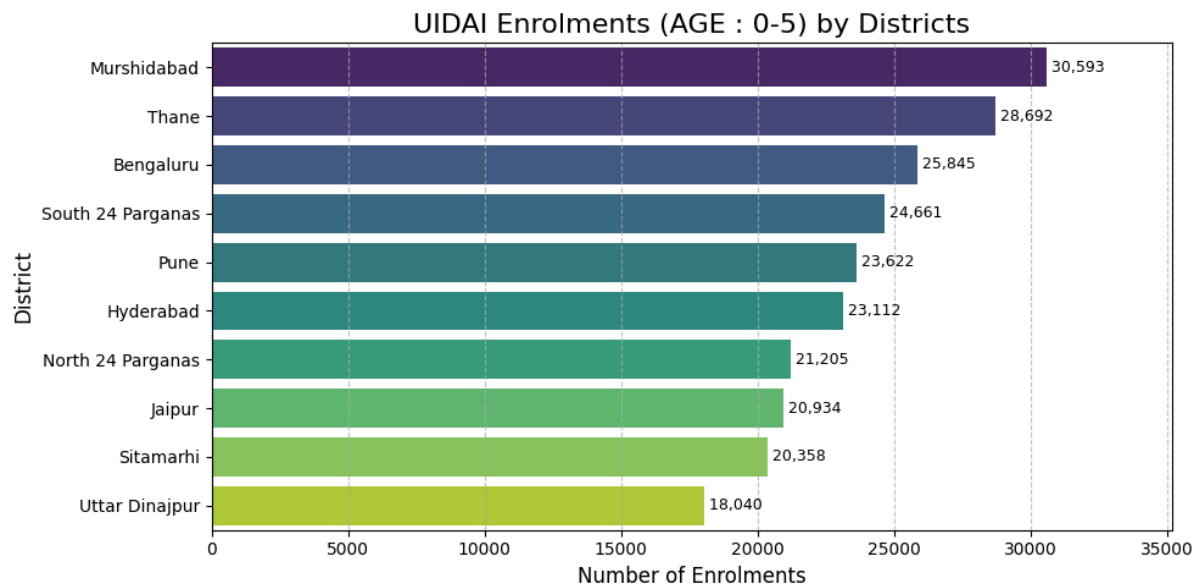
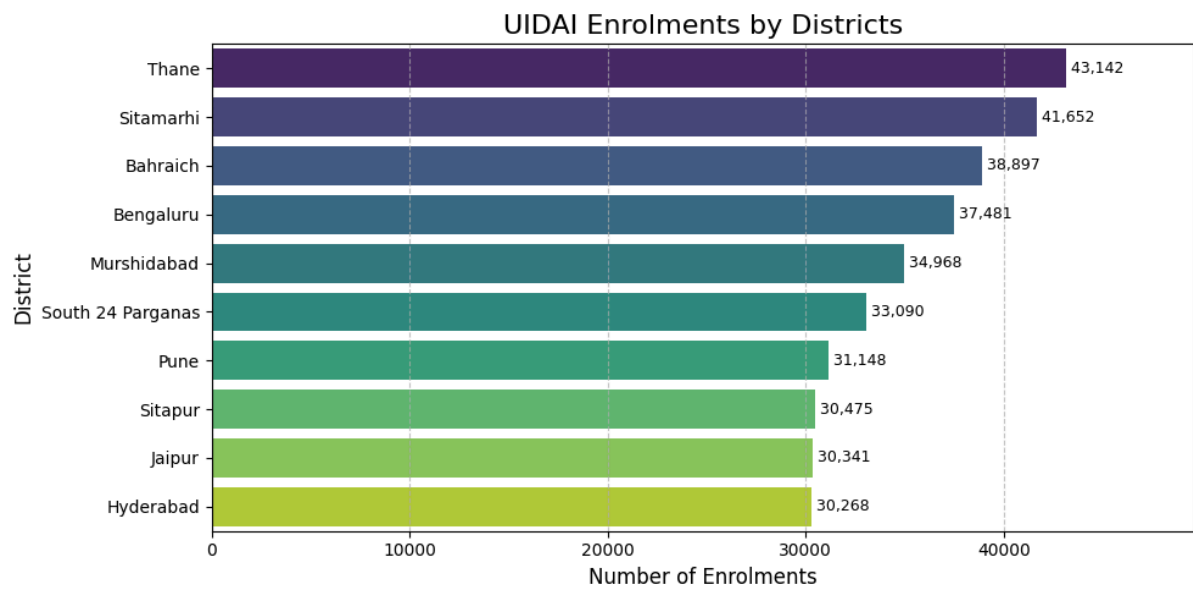


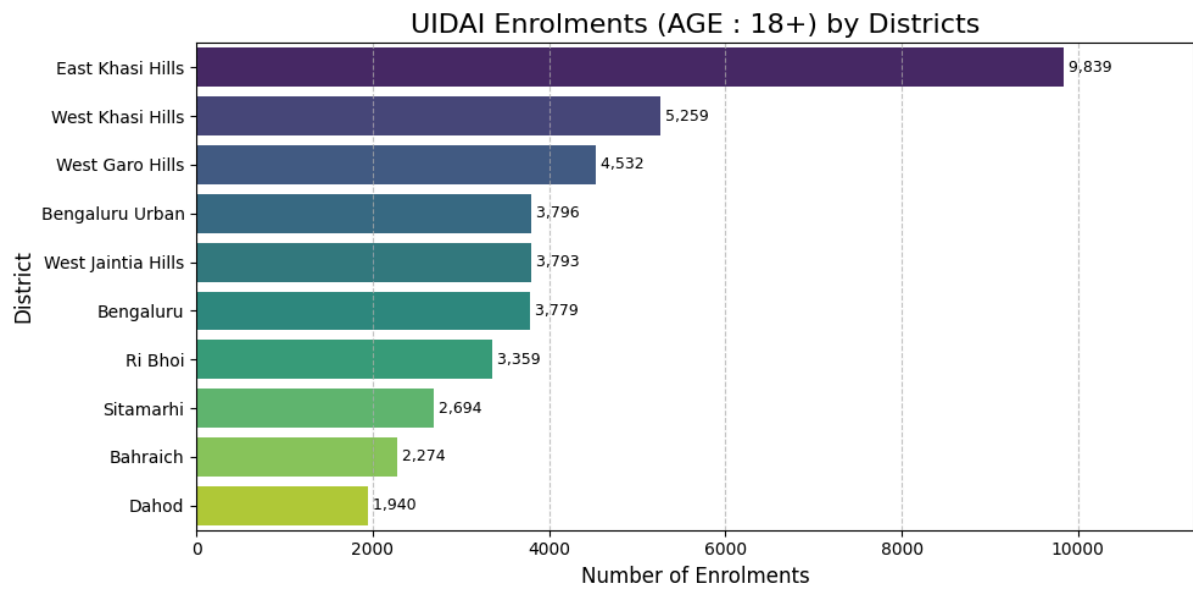
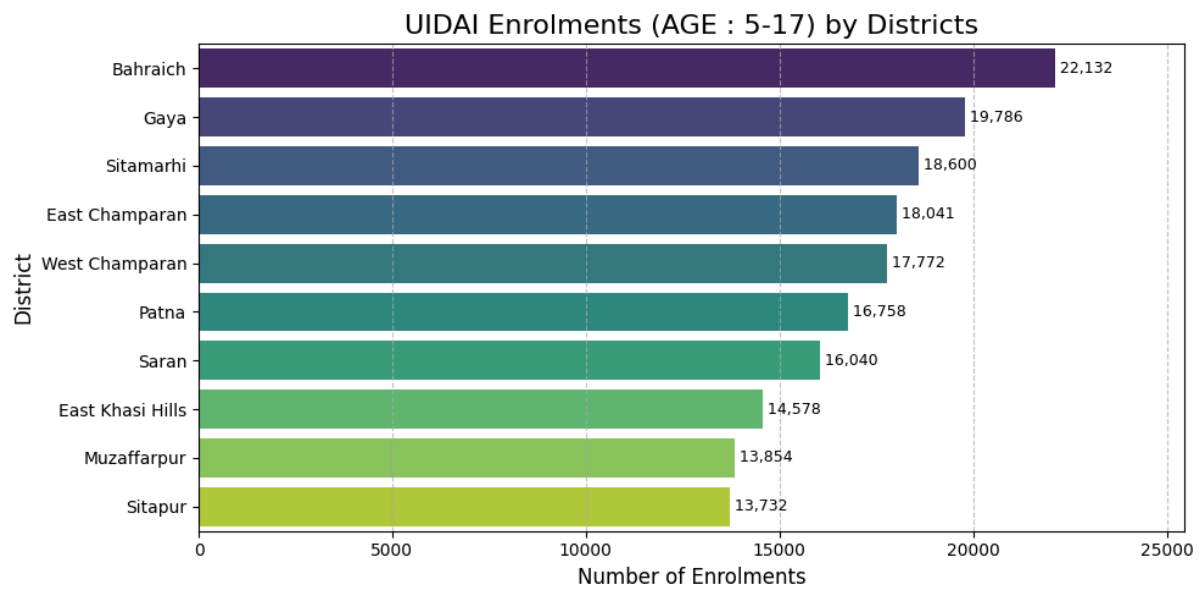
UIDAI Enrolments (AGE : 0-5) by States



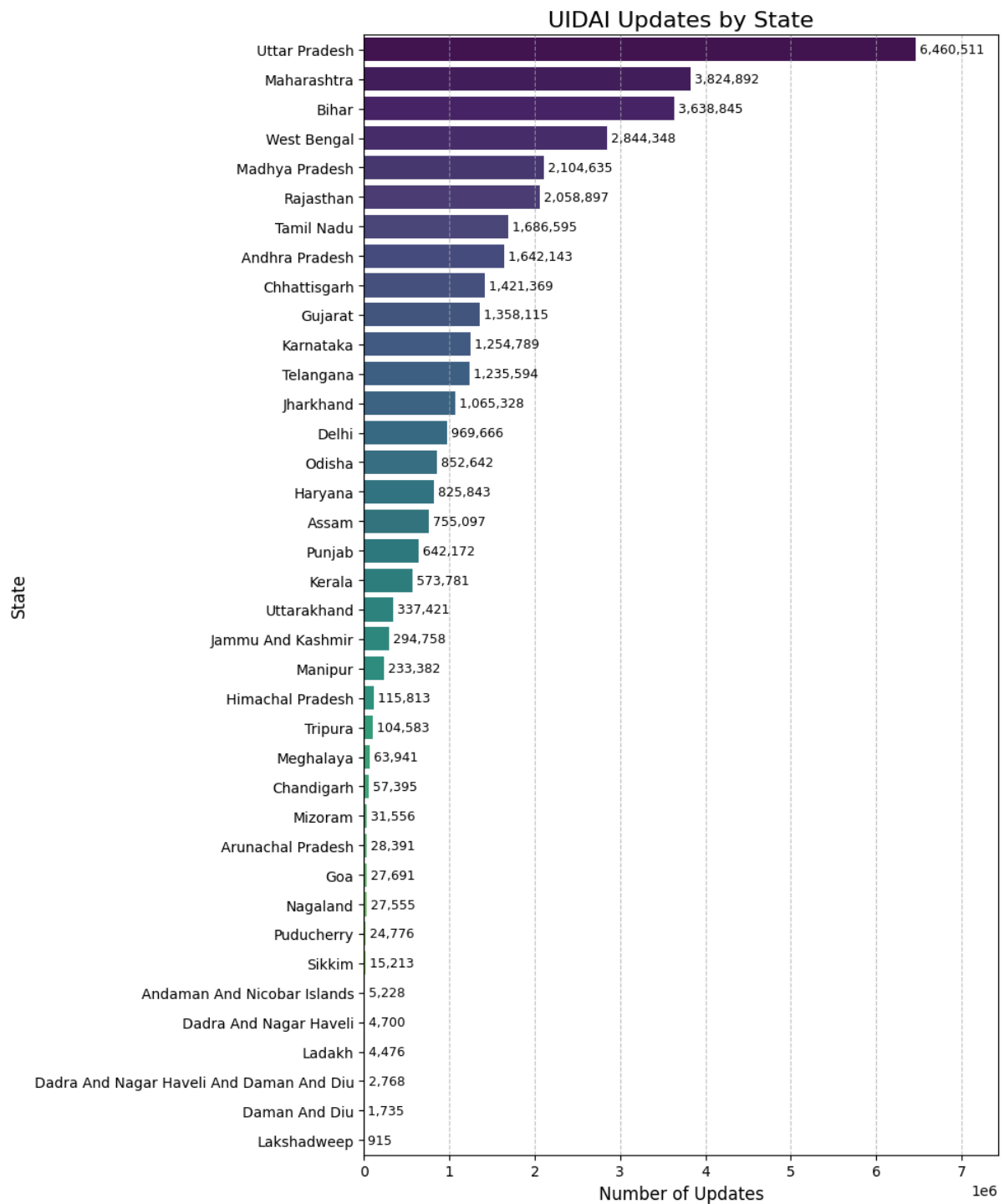






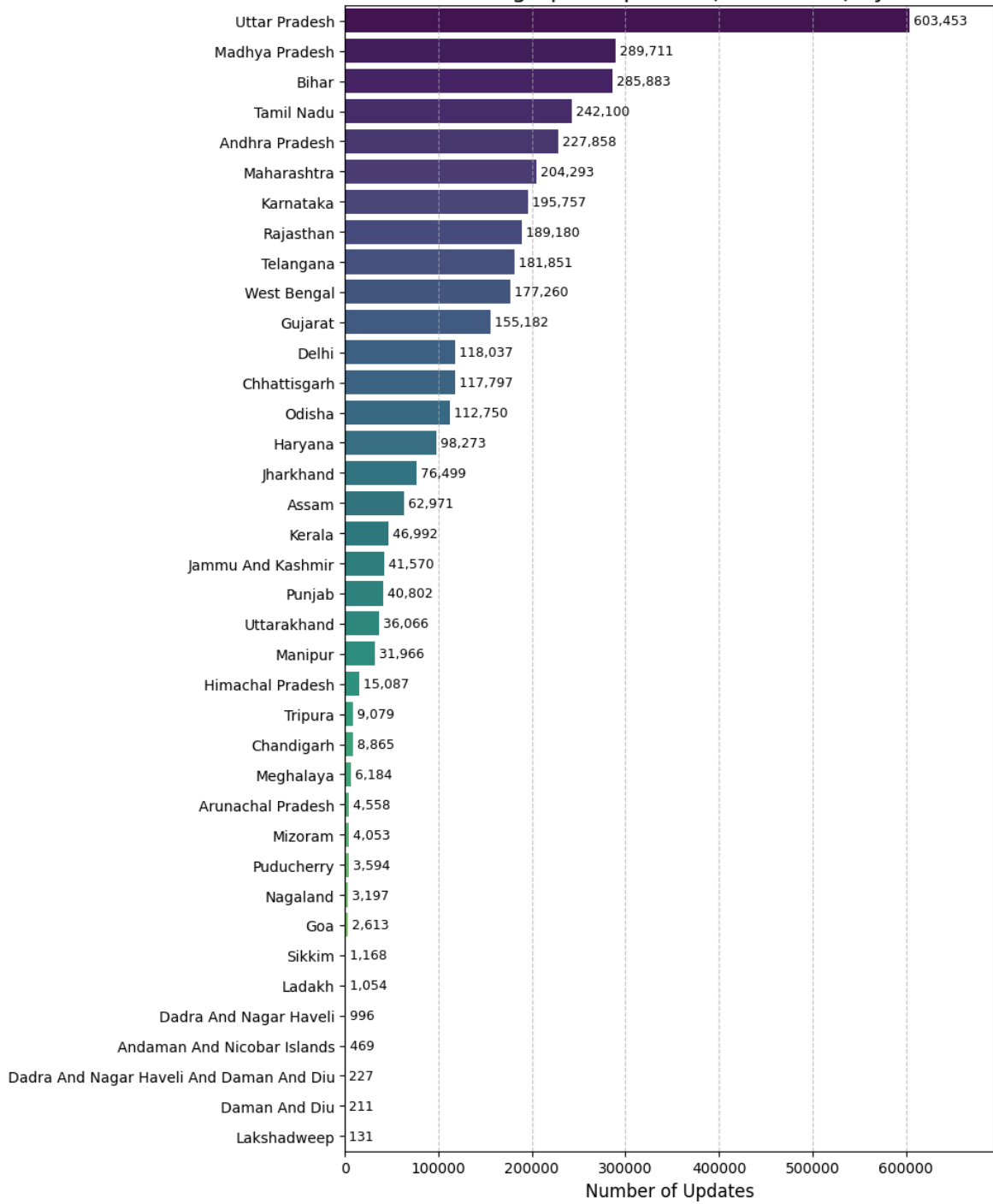


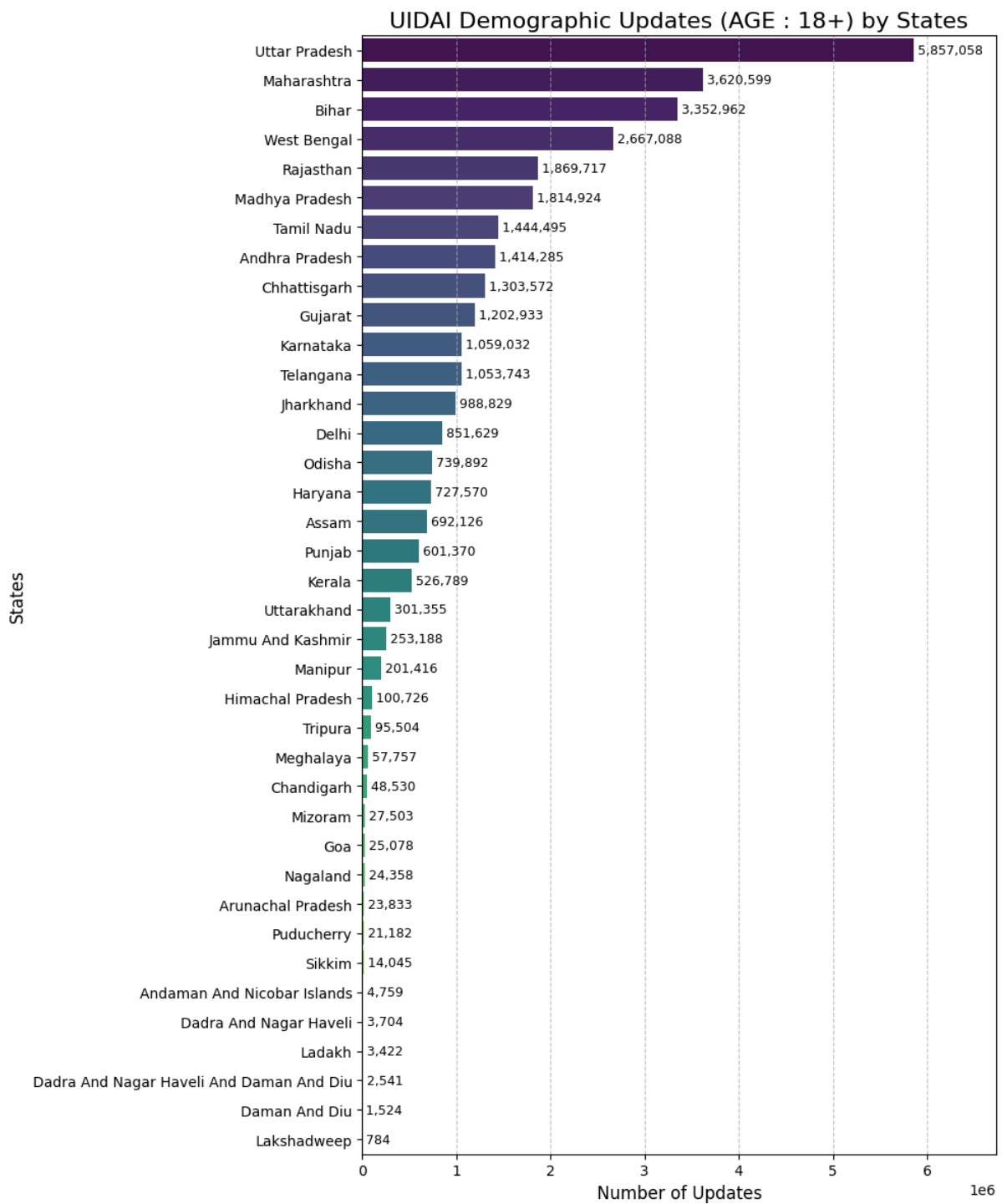
Demographic graphs for reference:

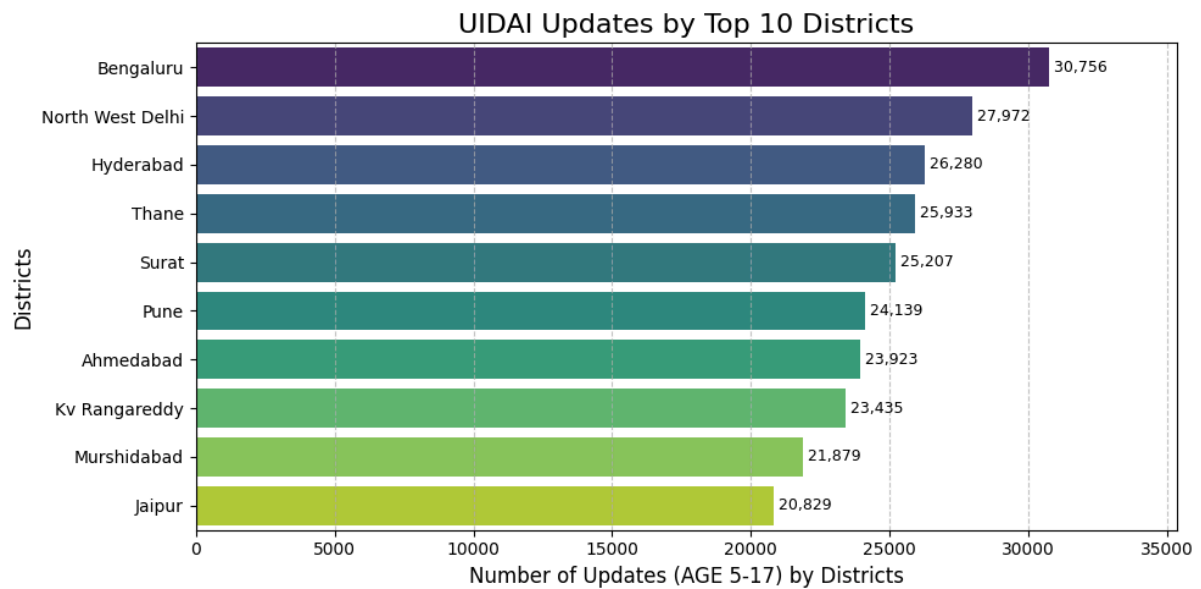
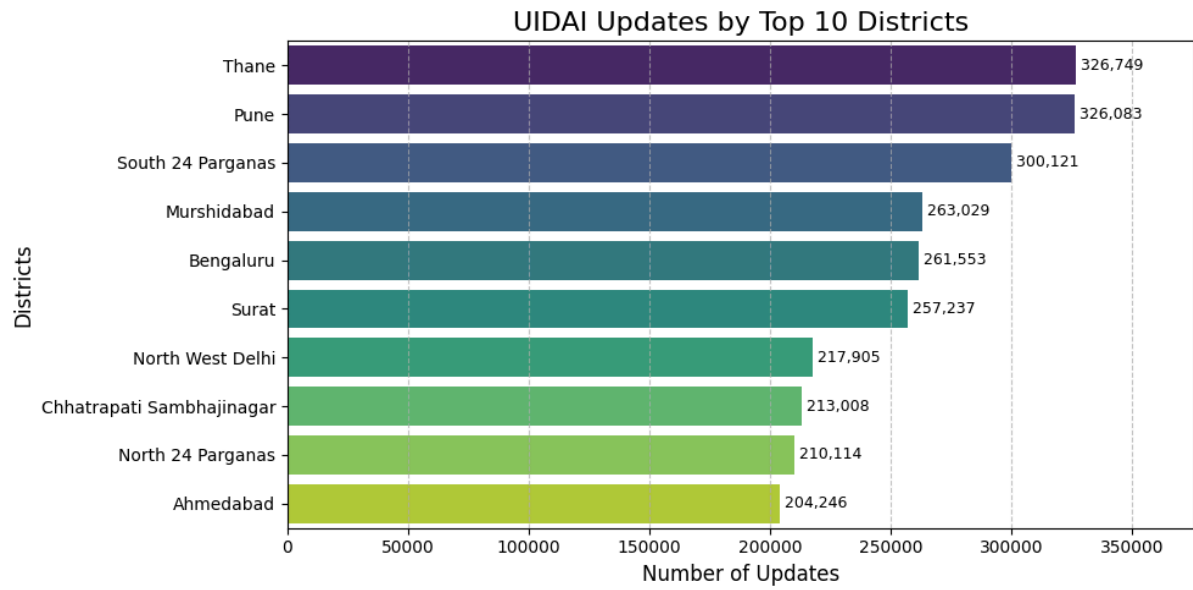


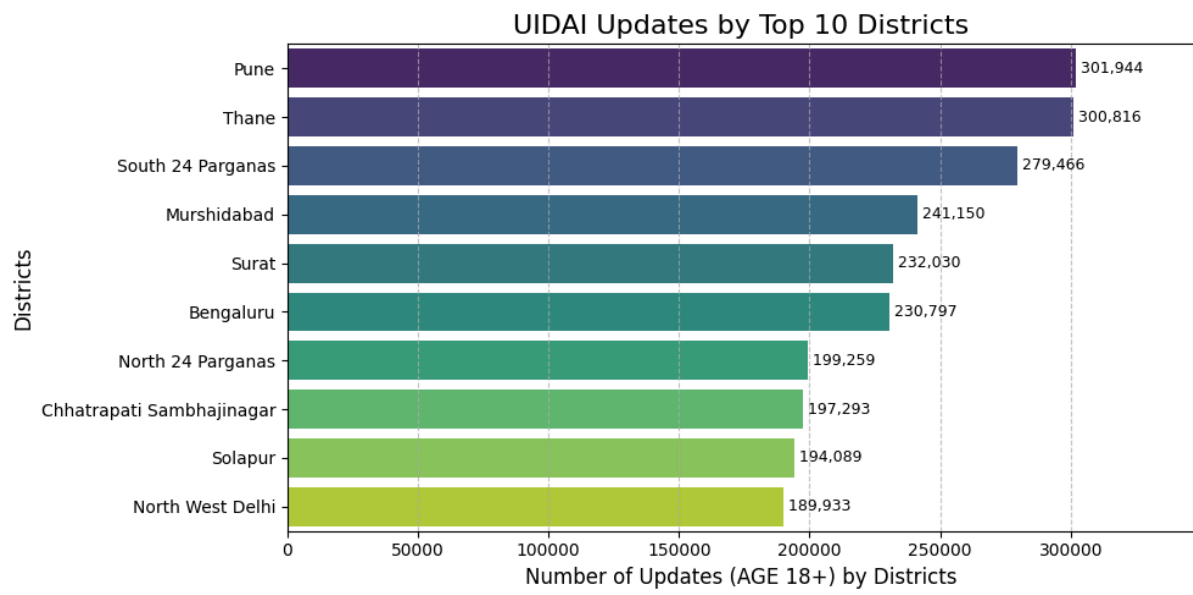
States

UIDAI Demographic Updates (AGE : 5-17) by States

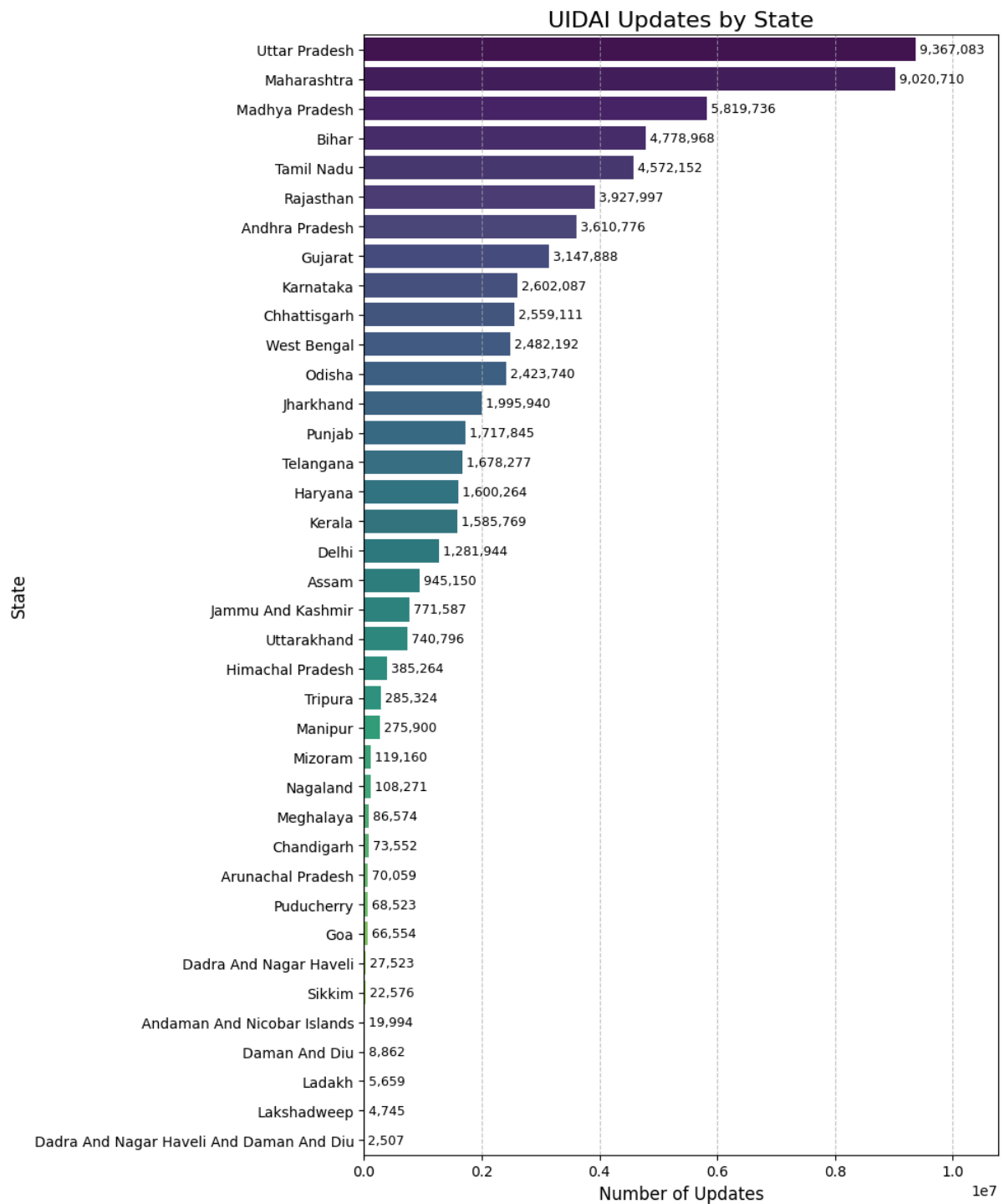


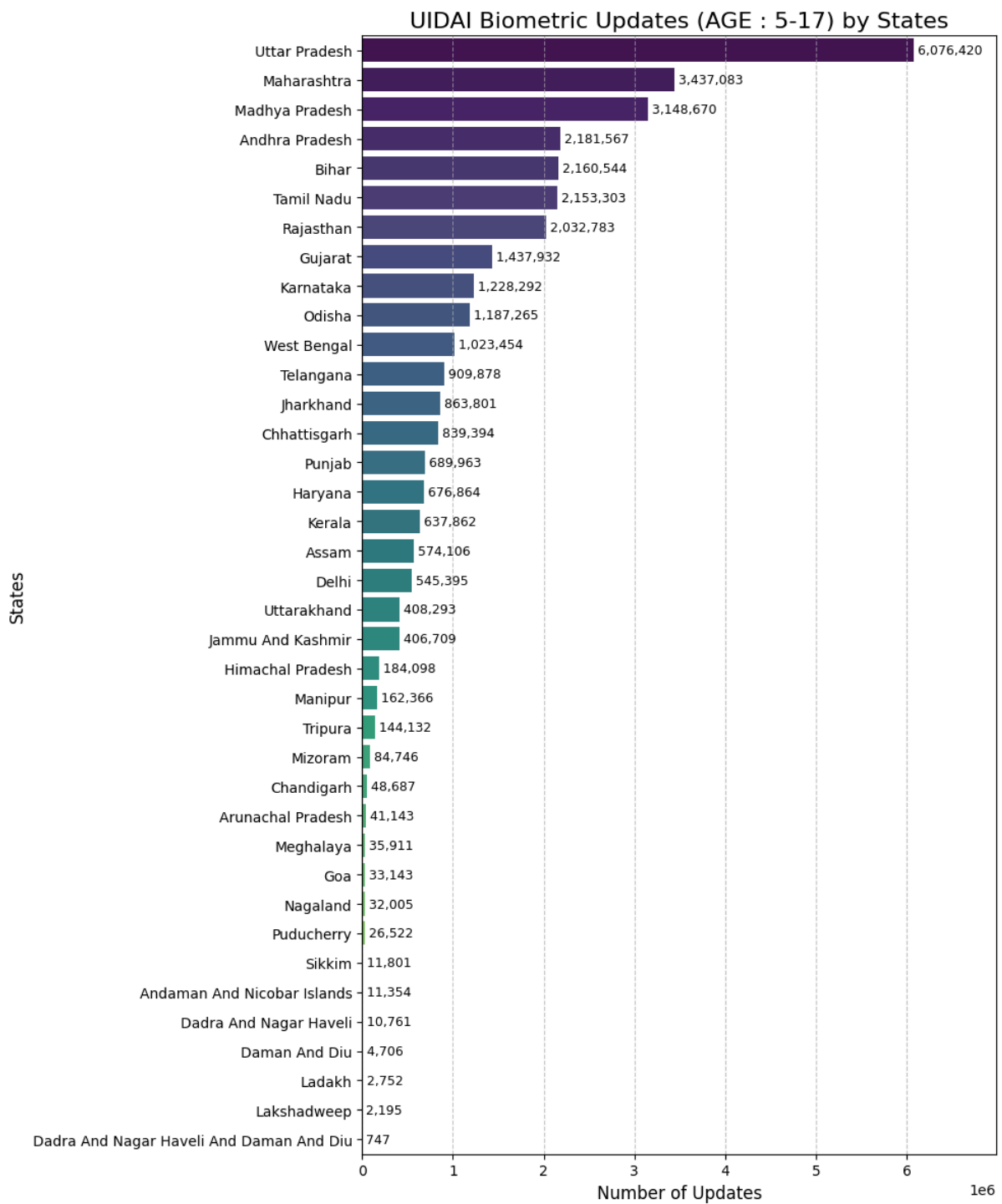


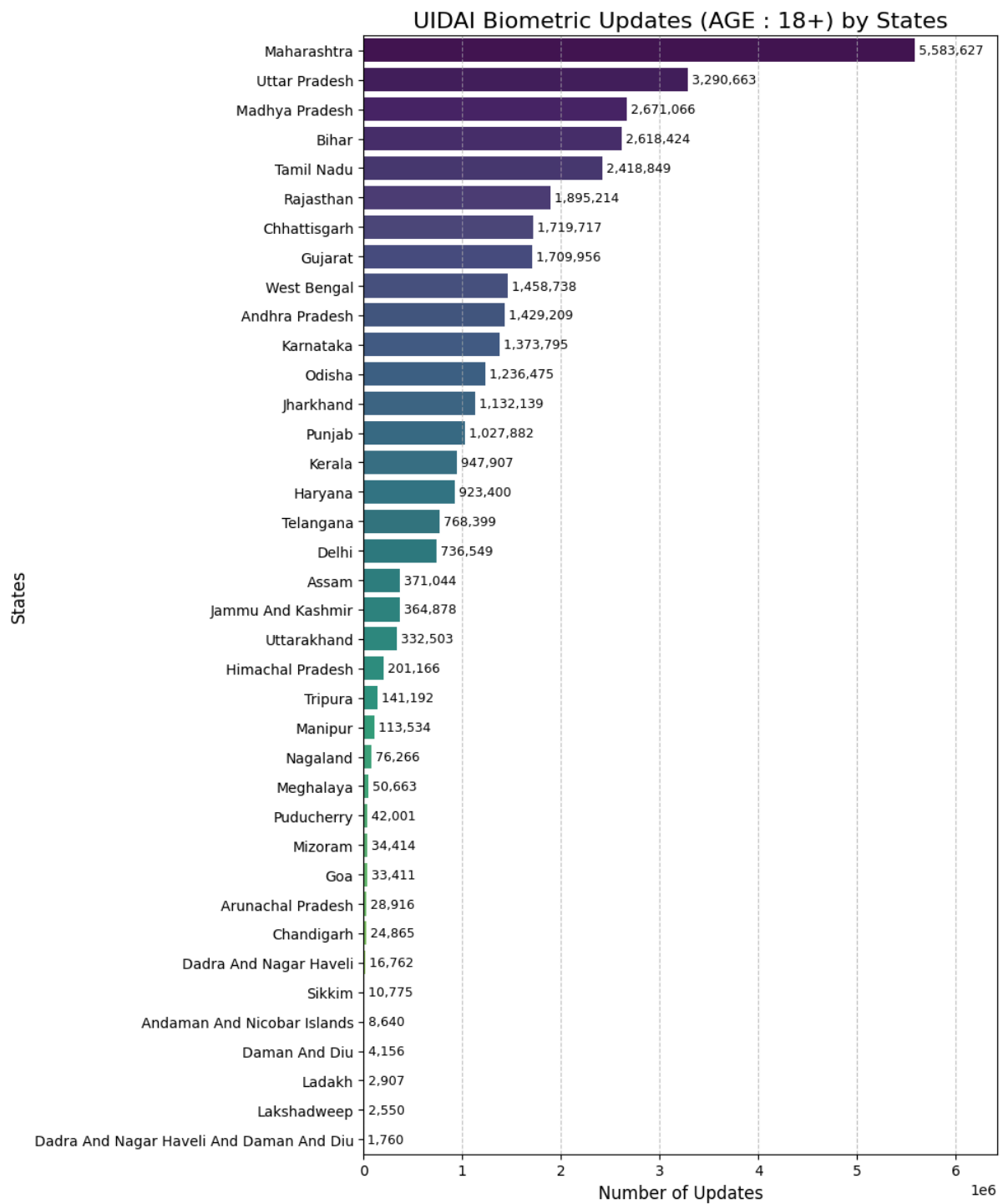


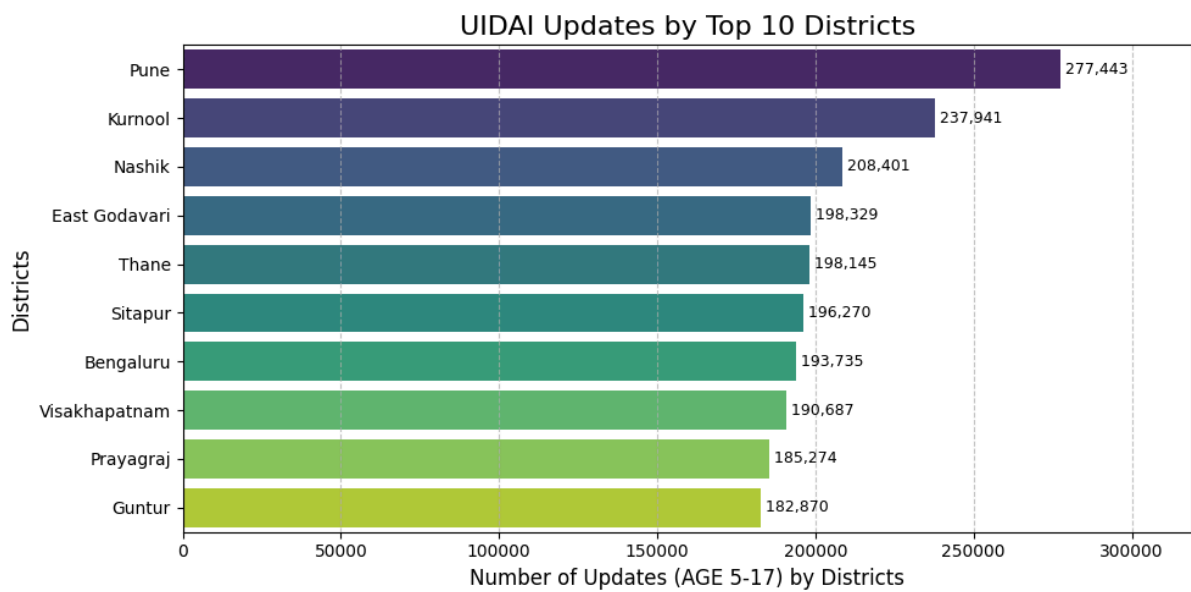
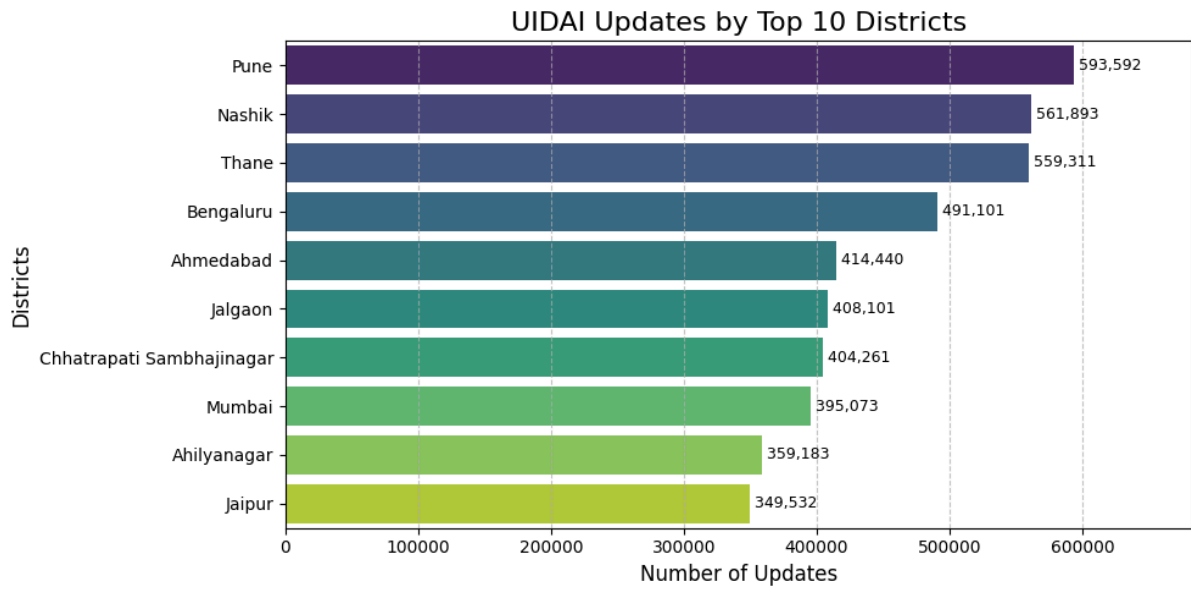


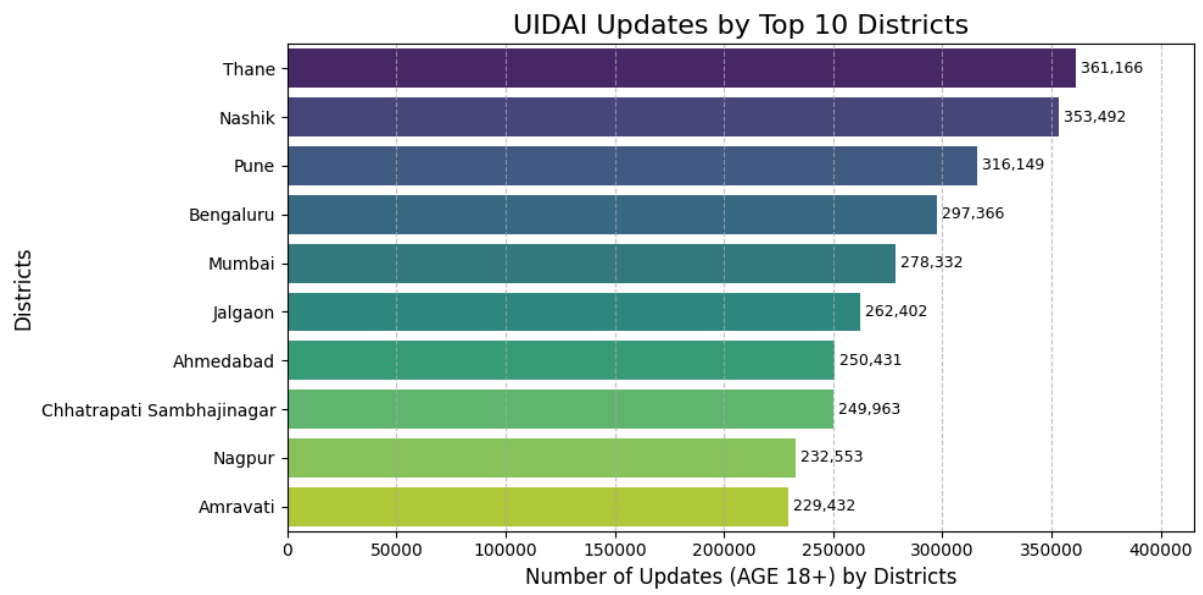
Biographic graphs for reference:











THANKYOU UIDAI FOR THIS OPPORTUNITY.