Summary and quick orientation

This dataset contains district-month enrollment records (3,872 rows × 6 columns). Each row records a month, normalized state, resolved district and three age buckets: age_0_5, age_5_17, age_18_greater. The analysis below highlights where enrollment volume sits, how it changes over time, and which states/districts and age groups drive the totals.

Headline findings

- Enrollment is strongly concentrated: a small number of district-month rows and districts account for a large share of total enrollments.

- Volume rises sharply later in the year (major activity around Sep–Nov 2025).

- The 5–17 age bucket contributes a large share of total enrollments overall, but the age mix varies month-to-month and by state.

- A few states and districts dominate totals — use those for focused operational follow-up.

Below I alternate short insights with the supporting tables or charts.

Data peek (first rows)

The dataset starts like this (first 10 rows, columns in original order):

month	state_norm	district_resolved	age_0_5	age_5_17
March 2025	assam	baksa	11	14
March 2025	assam	barpeta	24	34
March 2025	assam	bongaigaon	25	33
March 2025	assam	chirang	12	29
March 2025	assam	dhemaji	20	46
March 2025	assam	dhubri	89	229
March 2025	assam	dibrugarh	10	12

| month	state_norm	district_resolved	age_0_5	age_5_17
March 2025	assam	hojai	27	42
March 2025	assam	kamrup	13	20
March 2025	assam	kokrajhar	91	169

Insight / data pairing:

- This structure (month × state × district with three age measures) makes it straightforward to aggregate by month, state or district for program reporting.

Distribution of enrollments per district-month

- Insight: Per-row totals are skewed — most district-month rows are modest, while some are extremely large.

- Supporting image (distribution histogram):

Distribution of total enrollments per district-month row
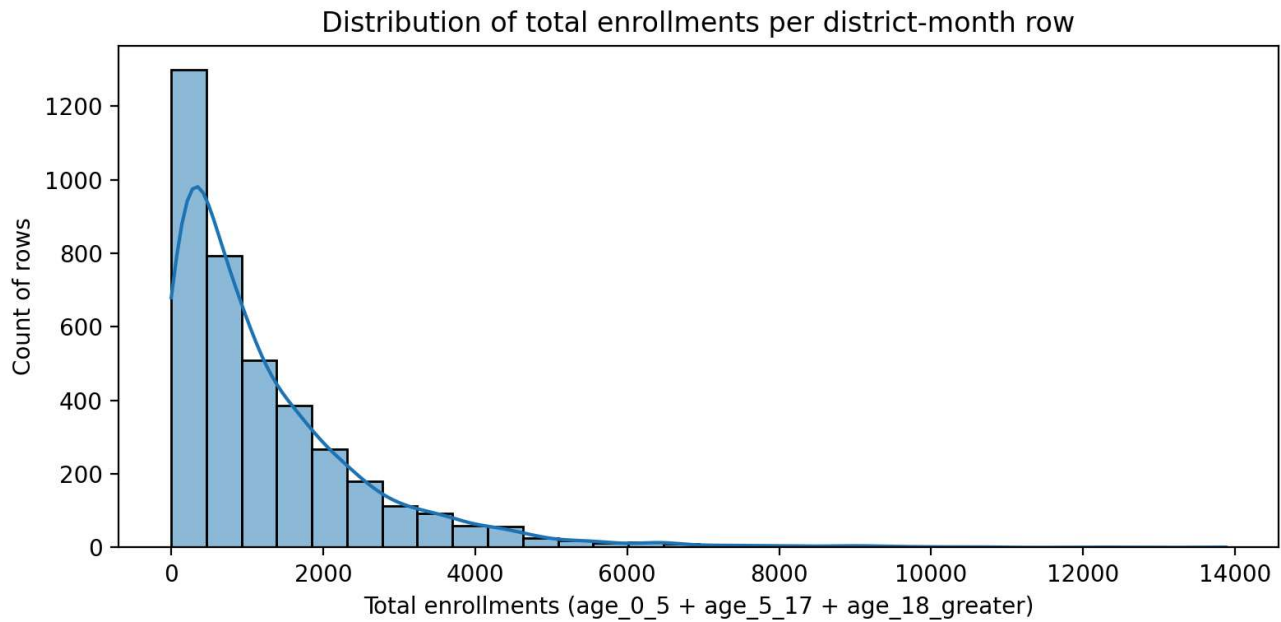
- Supporting table (summary of total_enroll per row): | index | total_enroll | |---:|---:| | count | 3872.0 | | mean | 1261.5098140495868 | | std | 1389.4513995975965 | | min | 1.0 | | 25% | 308.75 | | 50% | 825.5 | | 75% | 1725.25 | | max | 13877.0 |

Interpretation: median (~826) vs max (13,877) shows a long right tail — a few records dominate totals.

Overall time trend and peak months

- Insight: Total enrollment volume climbs later in the year, with major peaks in Sep–Nov 2025.

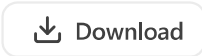- Supporting image (total enrollments over time):
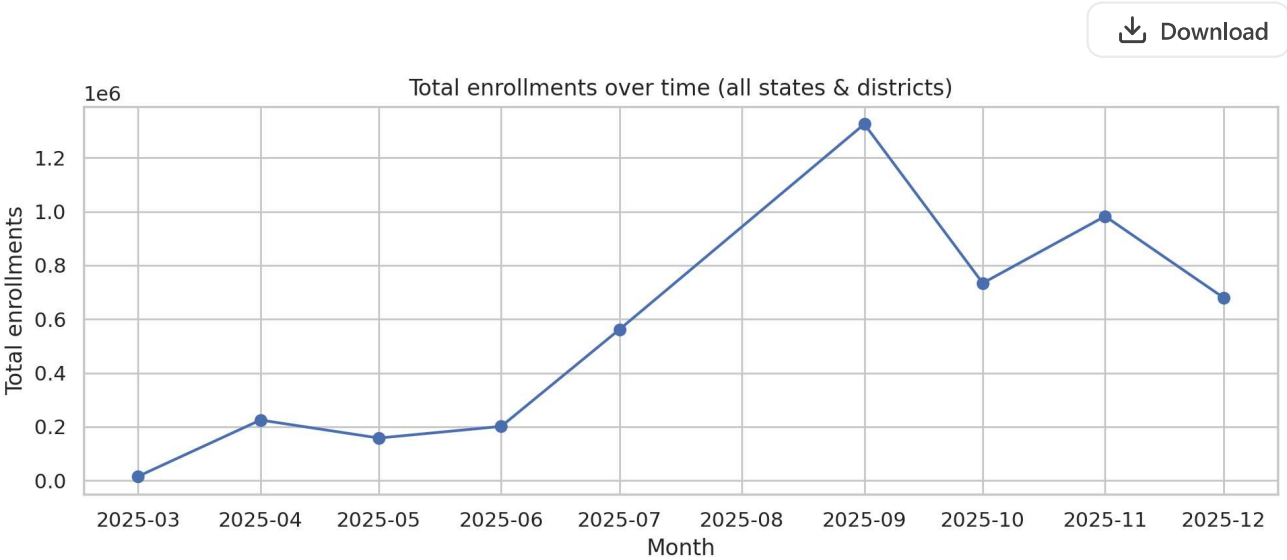
- Supporting table (top months by total): | month_dt | total | |---:|---:| | 2025-09-01 | 1,325,498 | | 2025-11-01 | 982,763 | | 2025-10-01 | 734,028 | | 2025-12-01 | 679,969 | | 2025-07-01 | 562,343 | | 2025-04-01 | 224,960 | | 2025-06-01 | 201,378 | | 2025-05-01 | 158,034 | | 2025-03-01 | 15,593 |
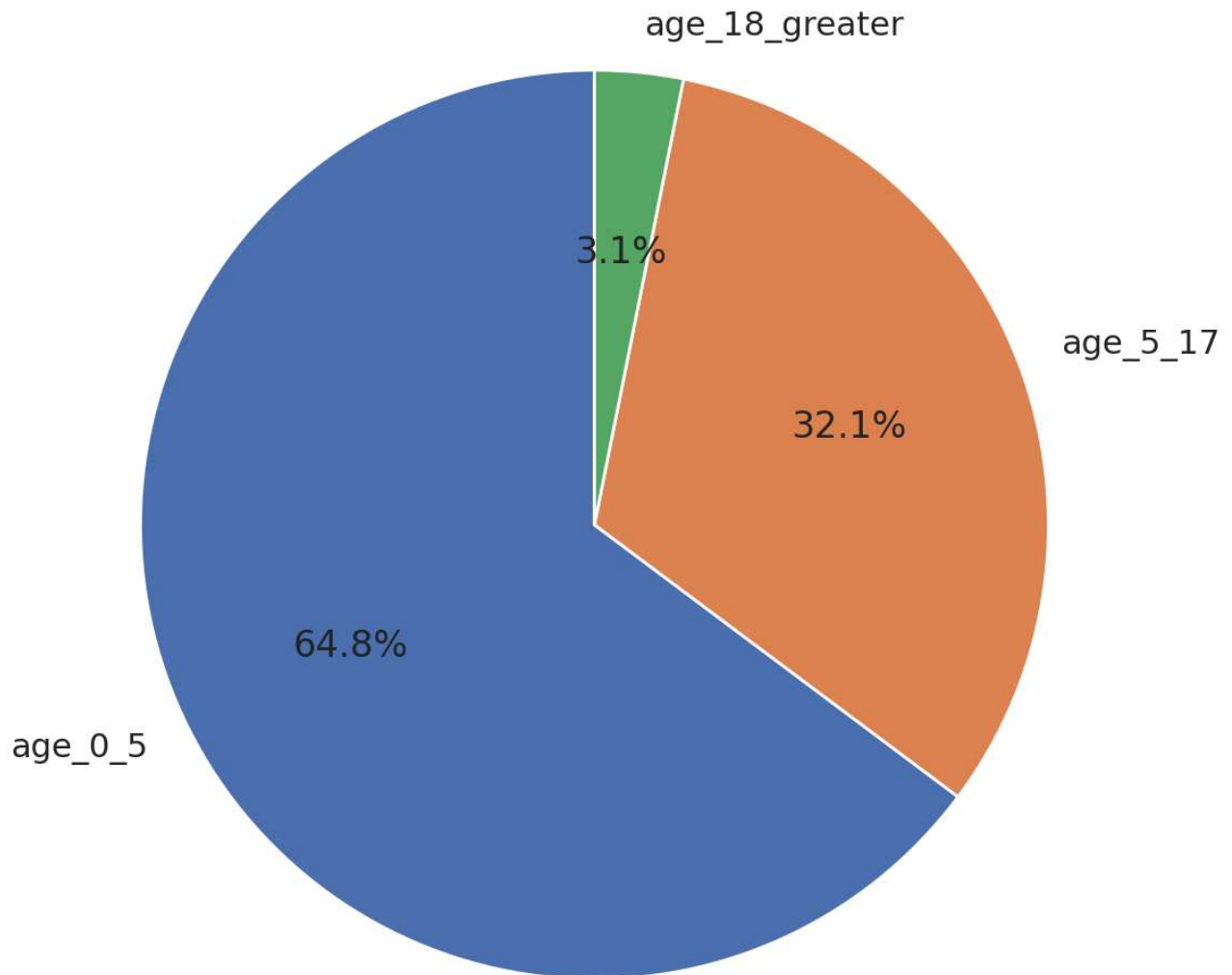
Interpretation: program activity or reporting coverage ramps up strongly into September and stays high through November and December.

Age-group composition (overall and over time)

- Insight: The 5–17 group is the largest contributor overall, but month-to-month mixes shift.

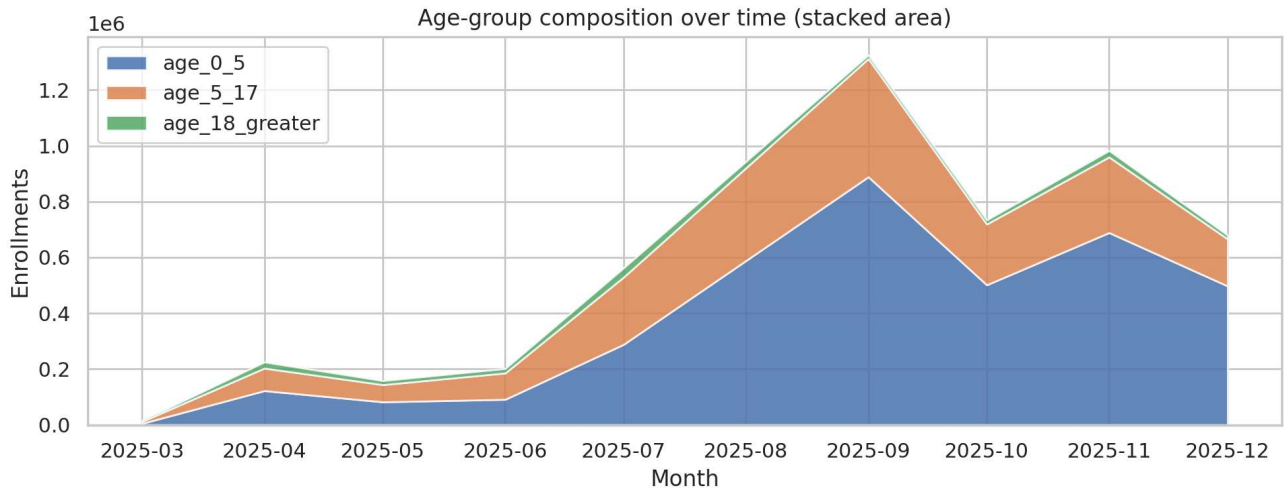- Supporting image (overall age distribution pie):

Overall enrollment share by age group

- Supporting image (age composition over time, stacked area):

[Download]

Interpretation: when totals rise late in the year, check whether growth is across all ages or dominated by one bucket — this informs targeting.
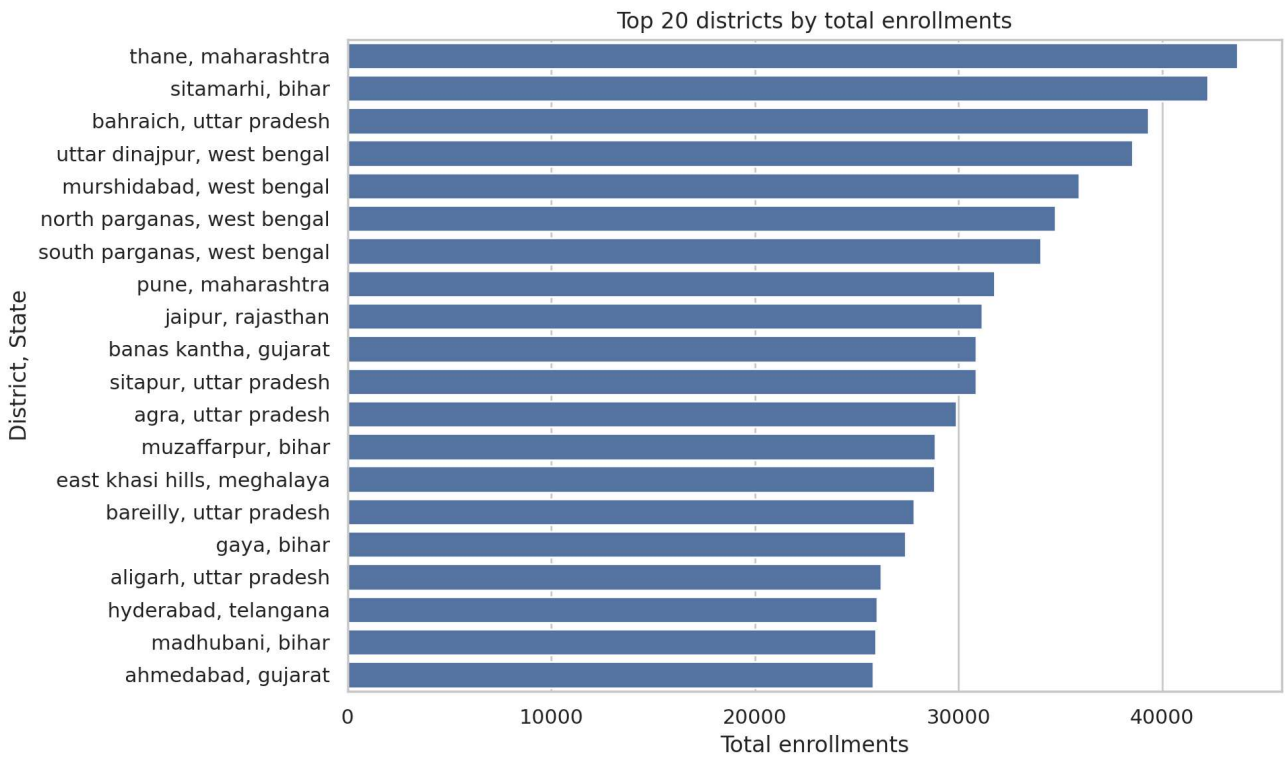
Geographic concentration by state and district

- Insight: Enrollment totals are concentrated in a limited set of states and districts.

- Supporting image (top 15 states bar chart):
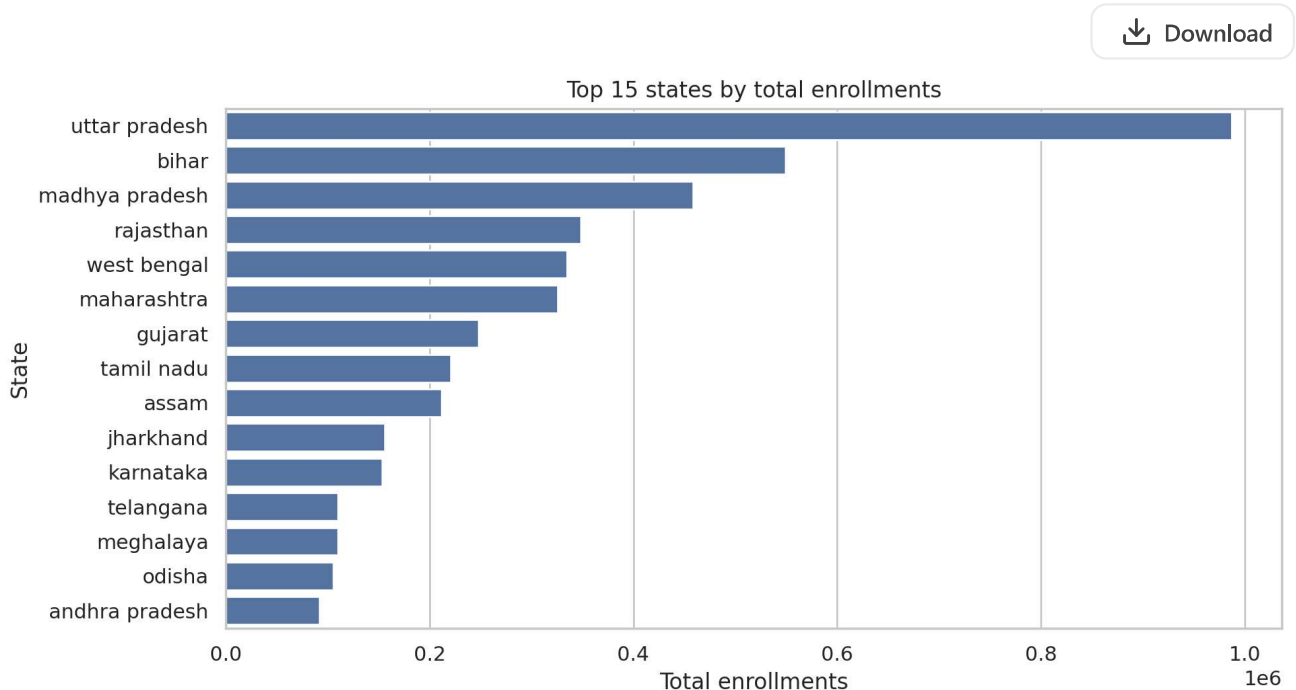
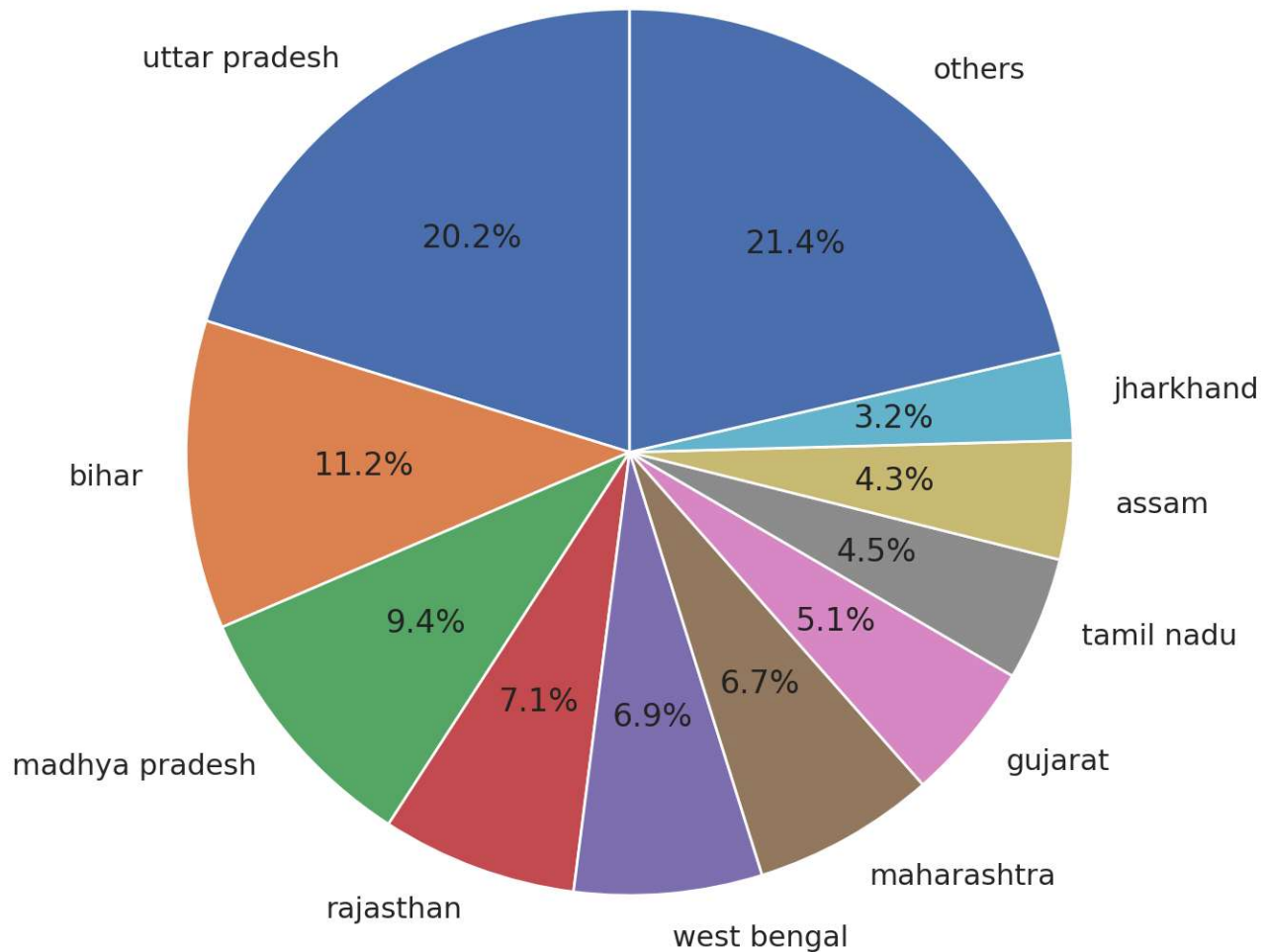- Supporting image (top 20 districts bar chart):

Top 20 districts by total enrollments

- Supporting image (state share pie: top 10 + others):

⬇ Download

## Total enrollment share by state (Top 10 + Others)

Interpretation: prioritize investigation, resource allocation, or service delivery in the top states/districts that drive most of the numbers.

Concentration behavior across districts (Pareto)

- Insight: A small number of districts account for a large cumulative share (Pareto-style); check the cumulative curve to see if 80% is reached with few districts.

- Supporting image (cumulative contribution curve):
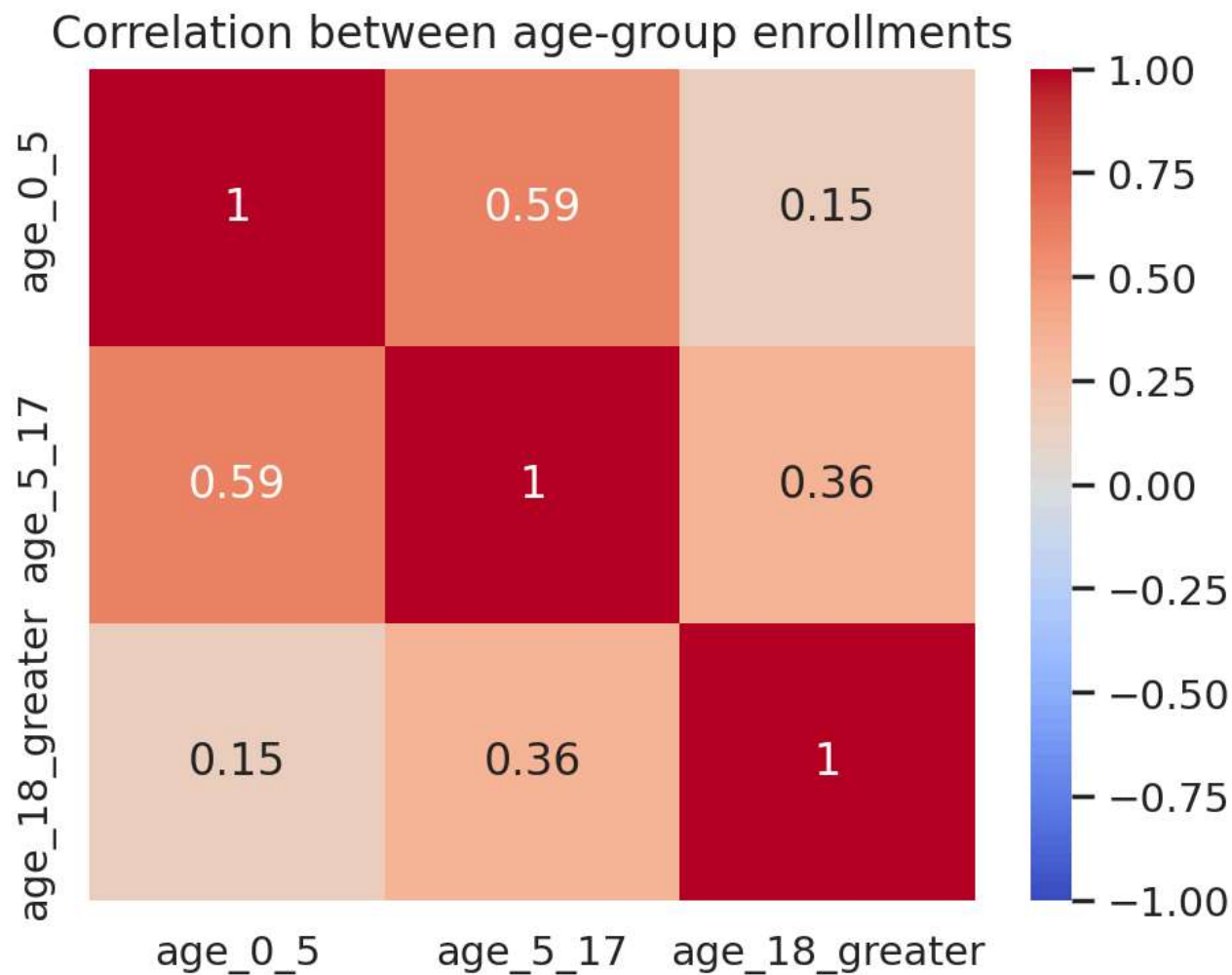
⬇ **Download**
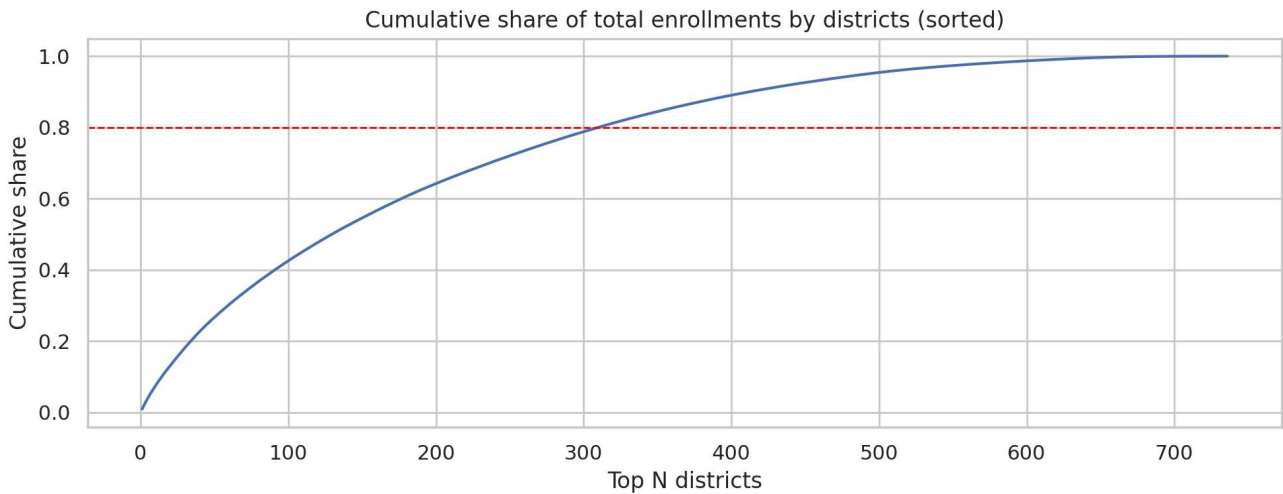
Interpretation: if 80% is reached quickly, concentrating resources on the top-ranked districts will address most beneficiaries.

Variability, correlations and relationships between age buckets

- Insight: Age buckets tend to move together (positive correlation) indicating common drivers (population, reporting).

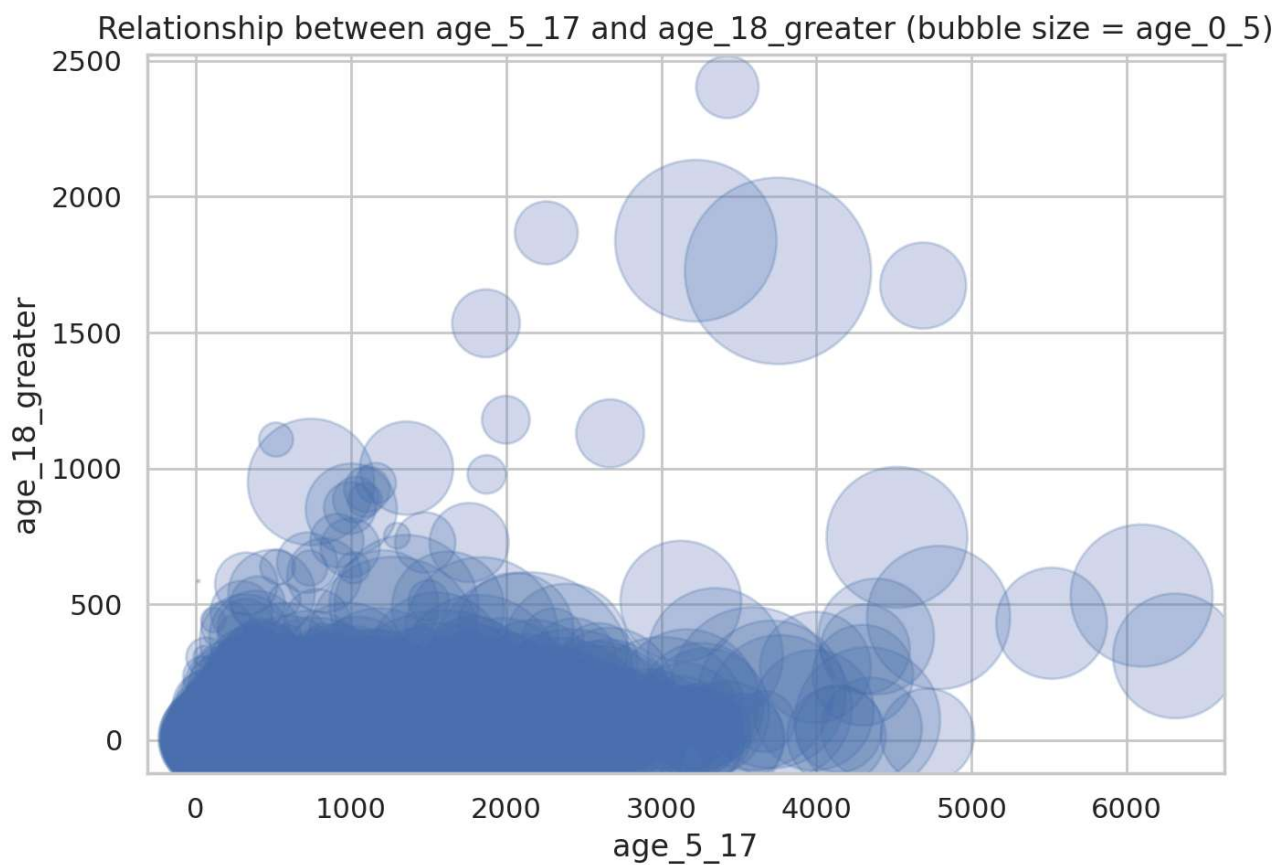- Supporting image (age-group correlation heatmap):

⬇ Download

Correlation between age-group enrollments

- Supporting image (boxplot showing spread and outliers by age group):

Distribution across district-month rows by age group
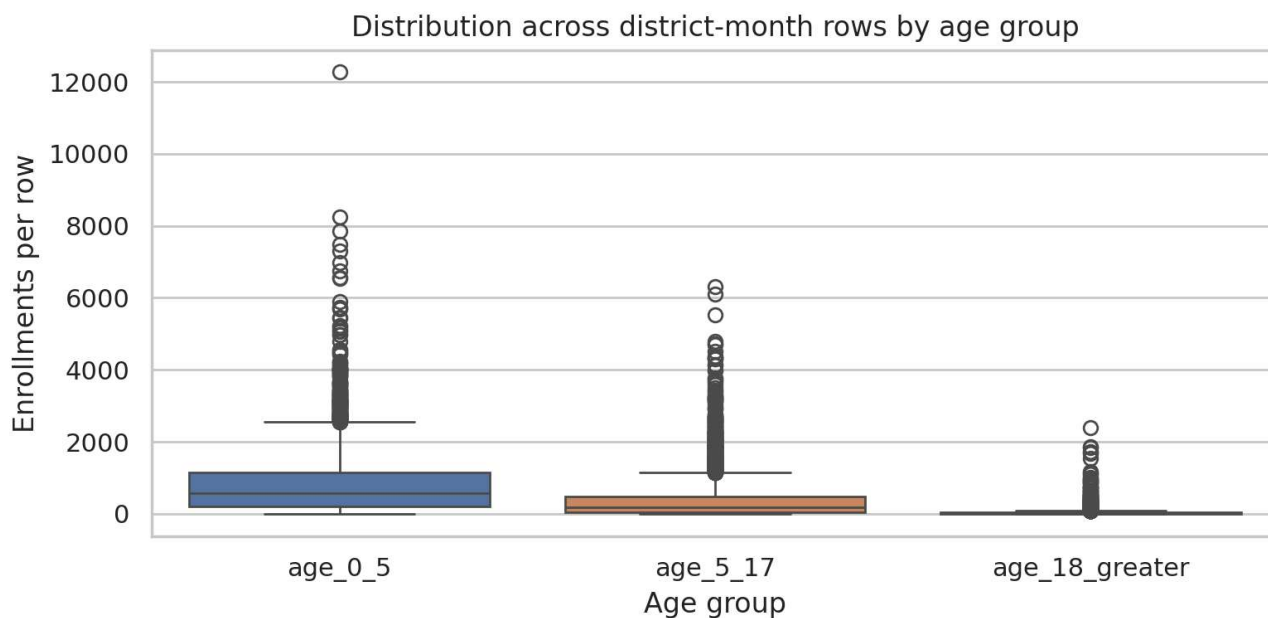
- Supporting image (scatter/bubble of age_5_17 vs age_18_greater):

Relationship between age_5_17 and age_18_greater (bubble size = age_0_5)

Interpretation: where buckets move together, target interventions by district population or reporting fixes rather than by single age group.

State-level age mixes and within-state district trends

- Insight: The age mix differs by state — stacked bars for top states show which are younger/older on average.

- Supporting image (age mix for top 10 states):
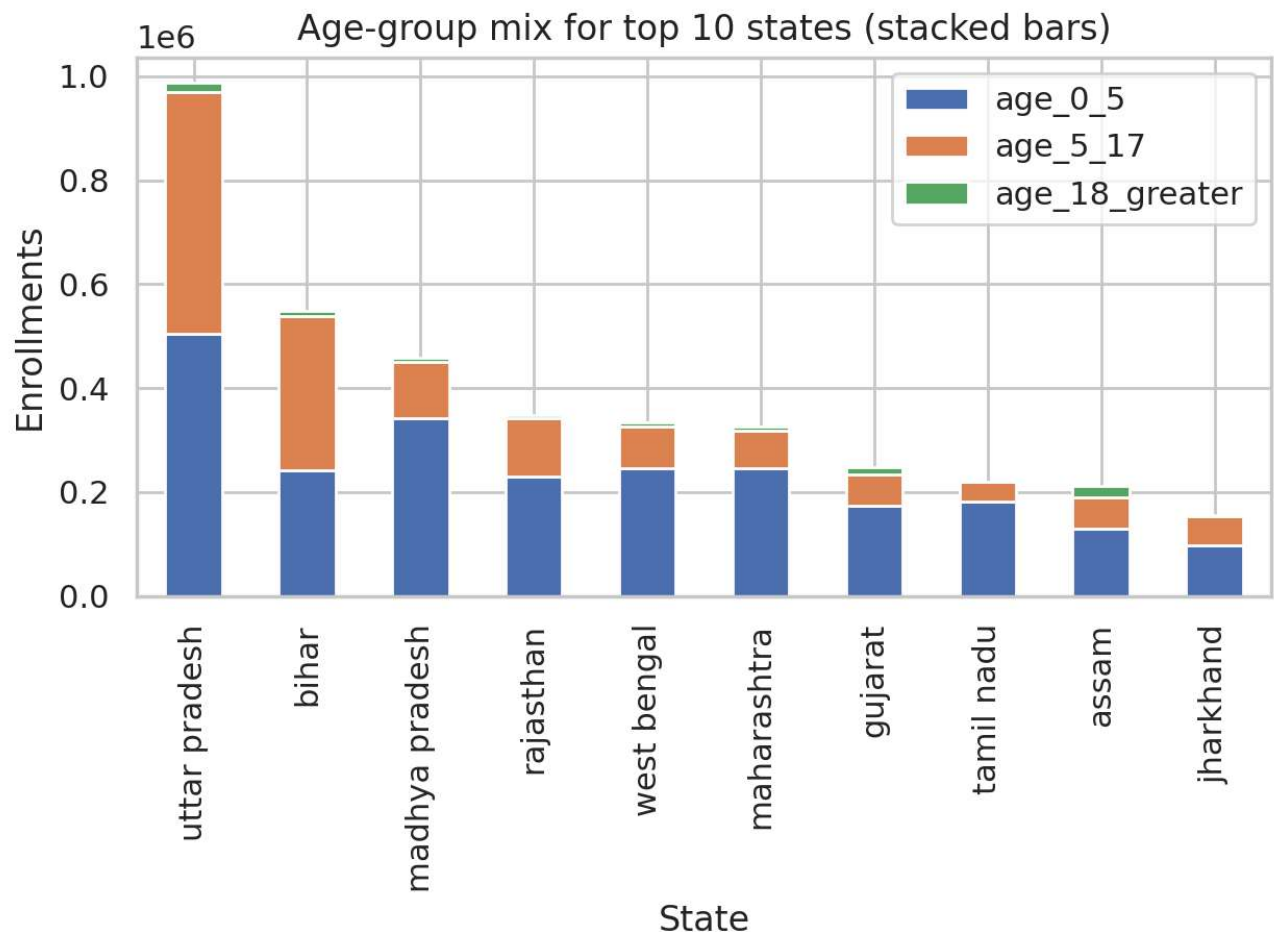
- Insight: Within high-coverage states, a few districts can be driving most of the growth — trends for the top 5 districts in the most-represented state highlight this.

- Supporting image (top 5 districts trends in most-represented state):

Top 5 districts trend in uttar pradesh
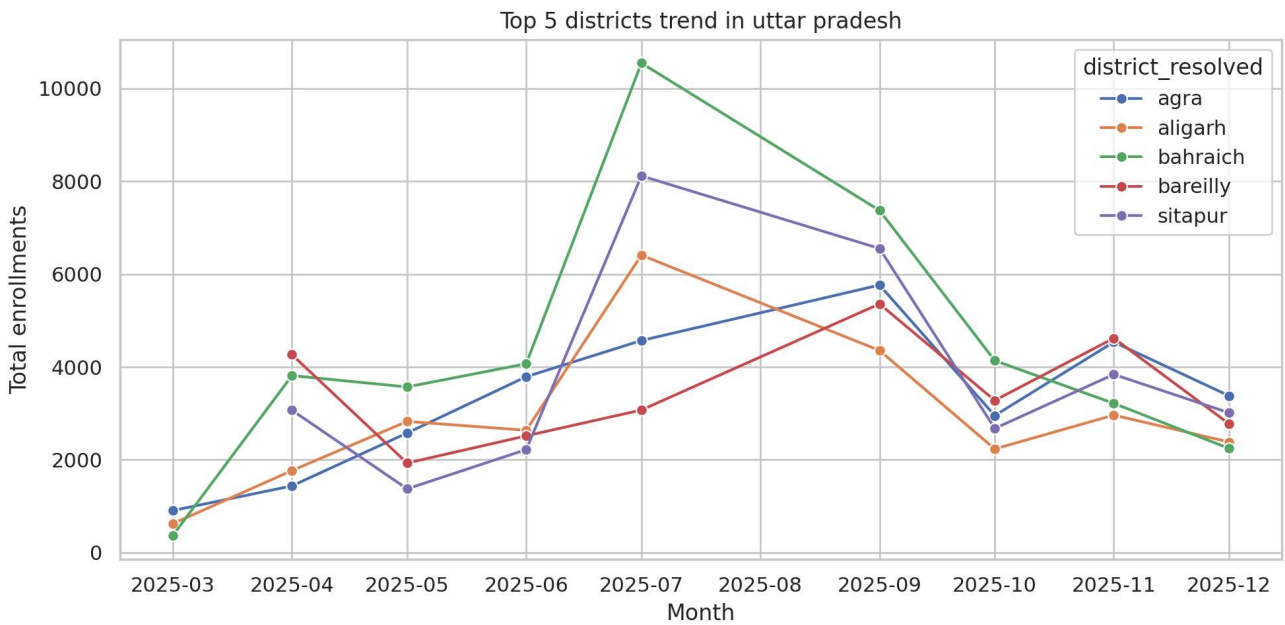
Interpretation: tailor state-level plans with district-level attention where the largest or fastest-growing districts exist.

Volatility and month-over-month growth

- Insight: MoM growth is volatile; some months show big increases consistent with the late-year surge.

- Supporting image (month-over-month growth rate):

Month-over-month growth rate of total enrollments (%)
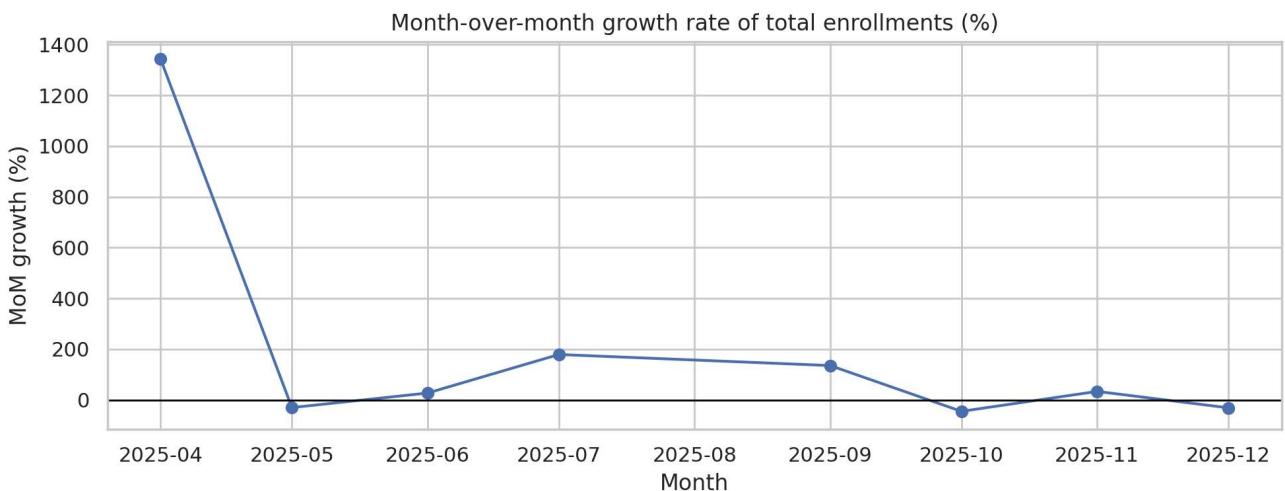
Interpretation: investigate whether growth reflects enrollment activity, data backfill, or reporting cadence changes.

Short, practical recommendations

- Focus audits, support and resourcing on the top states and districts identified by the bar charts and Pareto curve — they account for most beneficiaries.

- For reporting and planning, use month-level totals and the observed late-year surge to set realistic monthly targets and staffing.

- Where age buckets co-move strongly, design interventions at the district (or population) level rather than by specific age bucket.

- Investigate outlier district-months (very large totals) to confirm they are valid and to learn whether they reflect program scale-up, batch reporting, or data issues.

Closing summary

Overall, the dataset shows clear geographic and temporal concentration: most enrollment comes from a relatively small set of districts and from late-2025 months, and the 5–17 bucket is the largest share overall. The charts and tables above identify exactly which months, states and districts to prioritize for operational action, verification, and targeted programming. If you'd like, I can compile these visuals and the key tables into a single PDF or slide deck for easy sharing.