Biometric Dataset Analysis Report

1.Introduction

This report presents a detailed analysis of biometric update patterns using the provided dataset. The objective is to understand age-wise behavior, regional distribution, transition spikes, and anomalies through graphical and statistical methods.

2.Dataset Overview

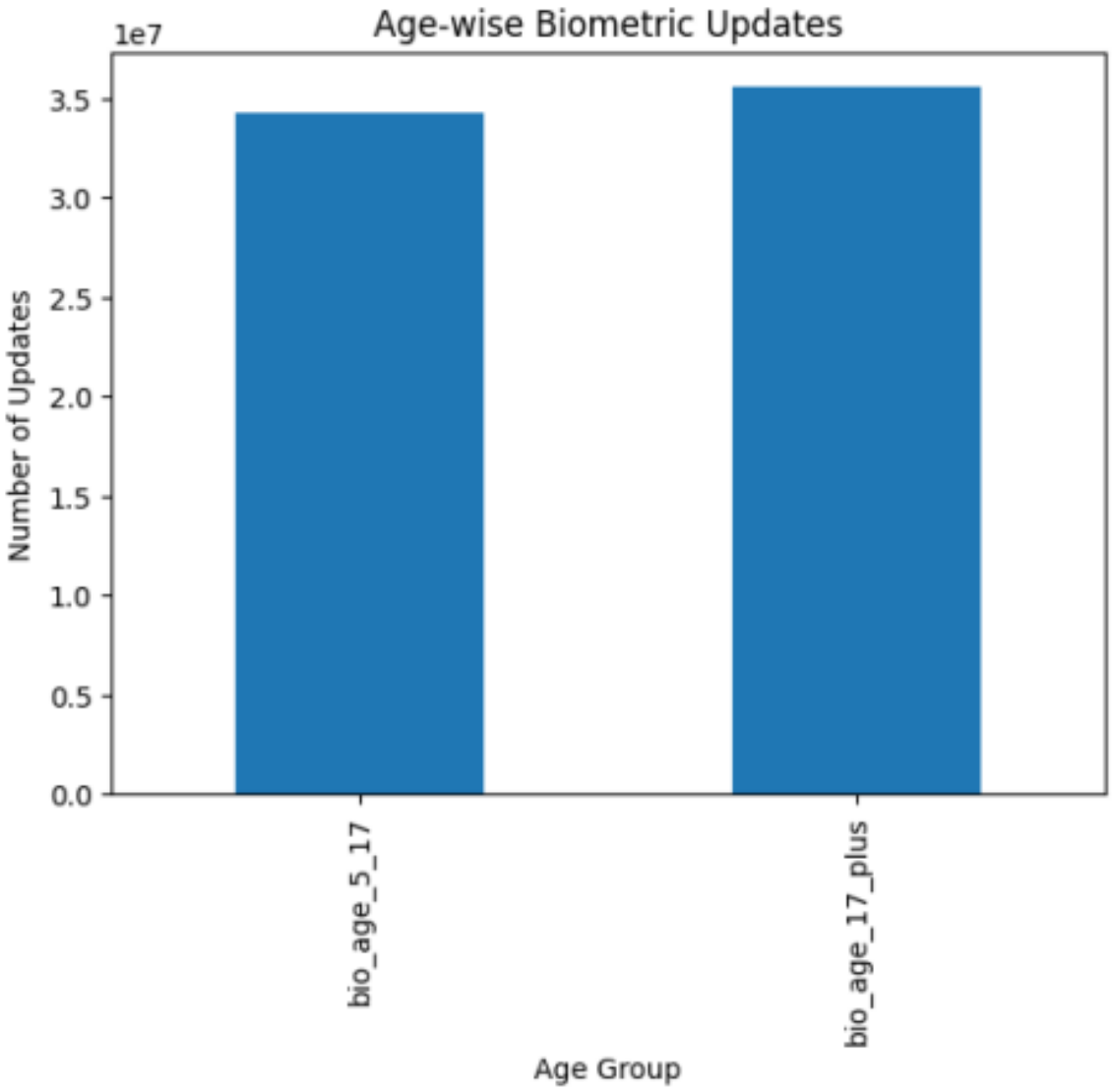
The dataset contains biometric update records categorized by date, state, district, pincode, and two age groups: 5–17 and 18+. A derived column named 'bio_total' was created by summing the age group values

3.Data Cleaning and Preparation

Data cleaning included converting date fields into datetime format, ensuring numerical consistency of age group values, handling missing values, and generating year and month features for trend analysis.

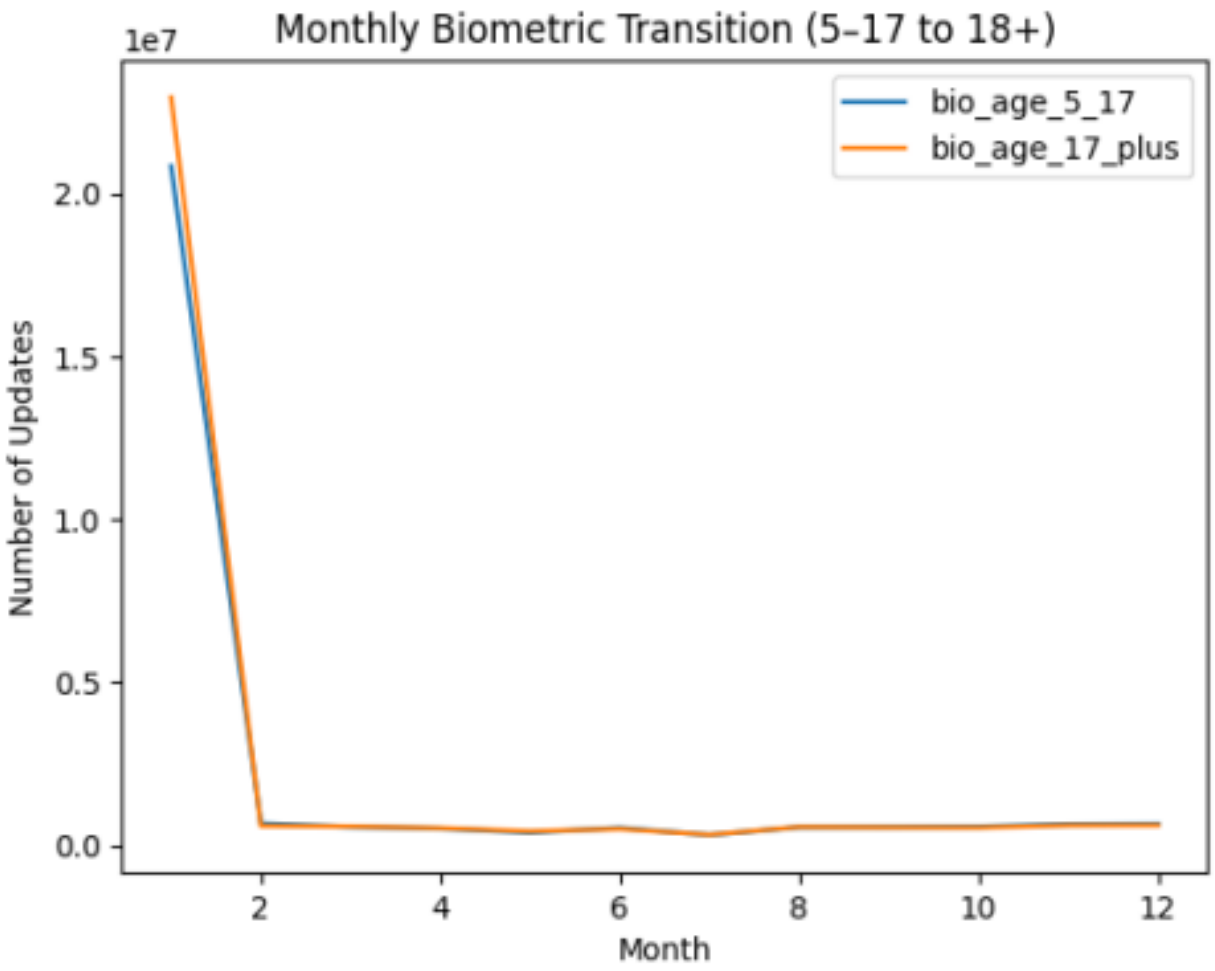
4.Age-wise Biometric Update Analysis

Age-wise analysis shows that the 18+ group slightly dominates the total number of biometric updates, indicating higher identity verification needs among adults.



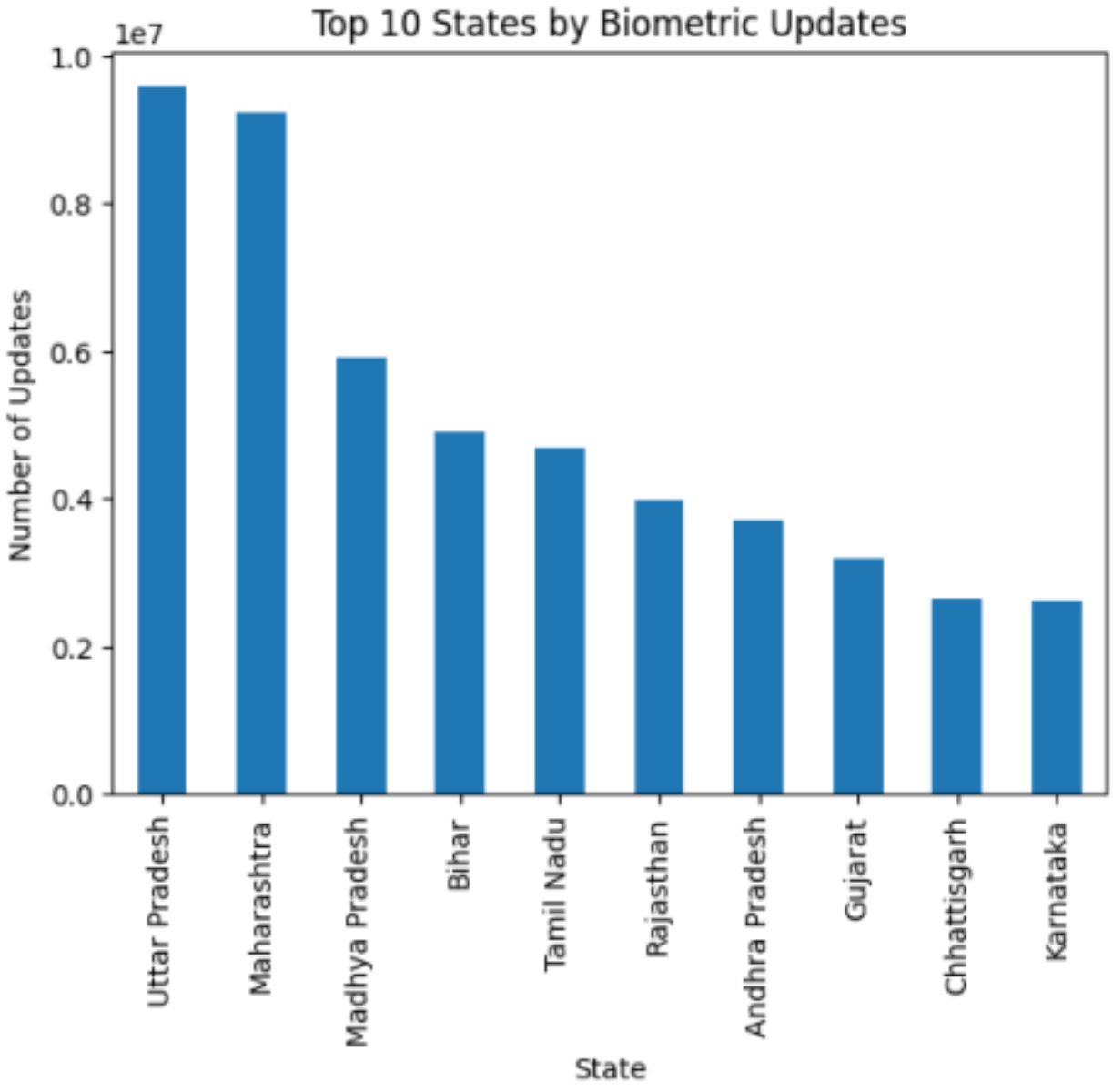
5. 5–18+ Transition Spike Analysis

The transition from the 5–17 group to the 18+ group shows noticeable spikes, reflecting identity updates related to adulthood milestones such as higher education, employment, and documentation needs.



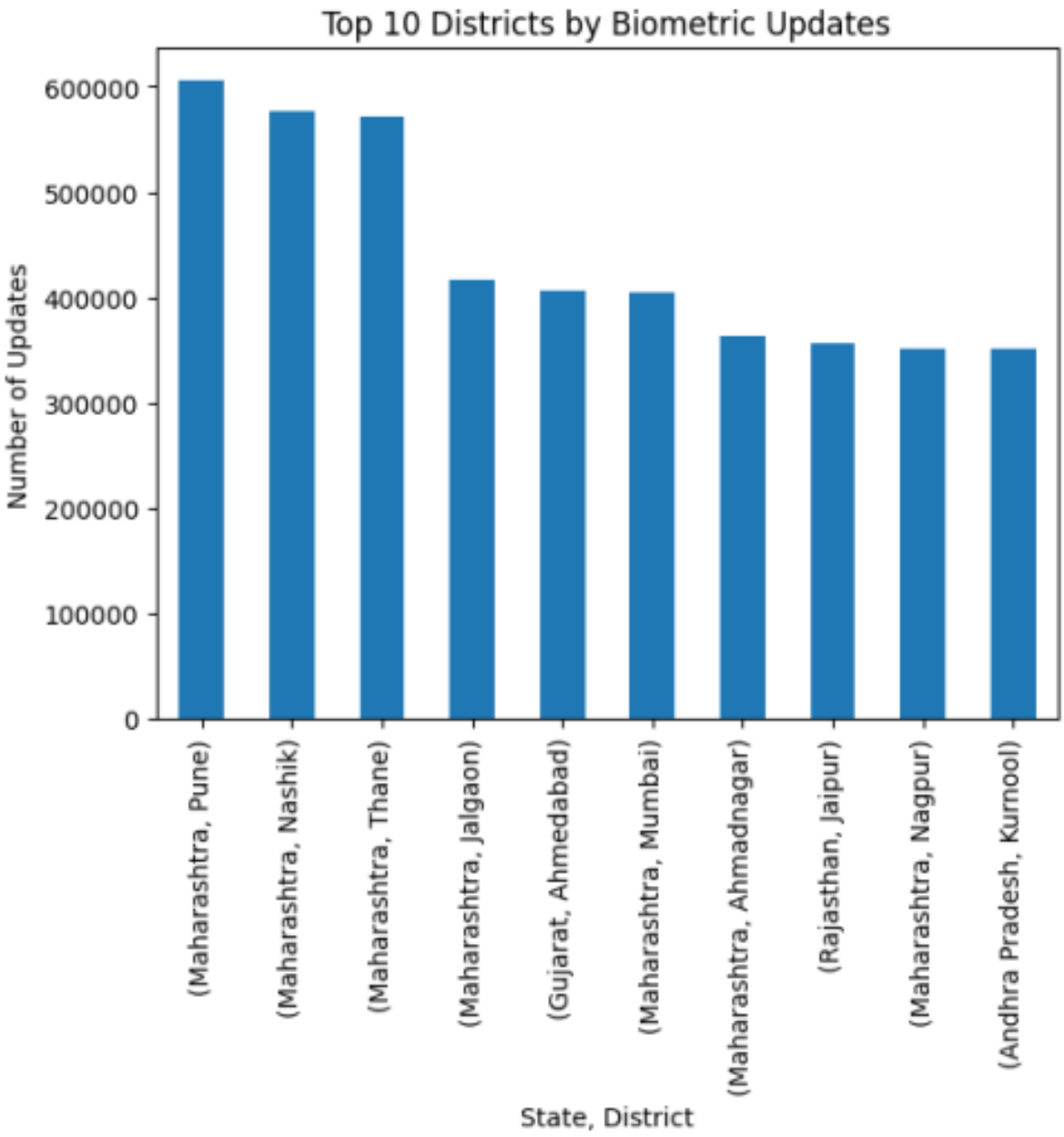
6.State-wise Distribution

State-wise analysis reveals that Uttar Pradesh, Maharashtra, and Madhya Pradesh are the top contributors, reflecting population density and administrative load.



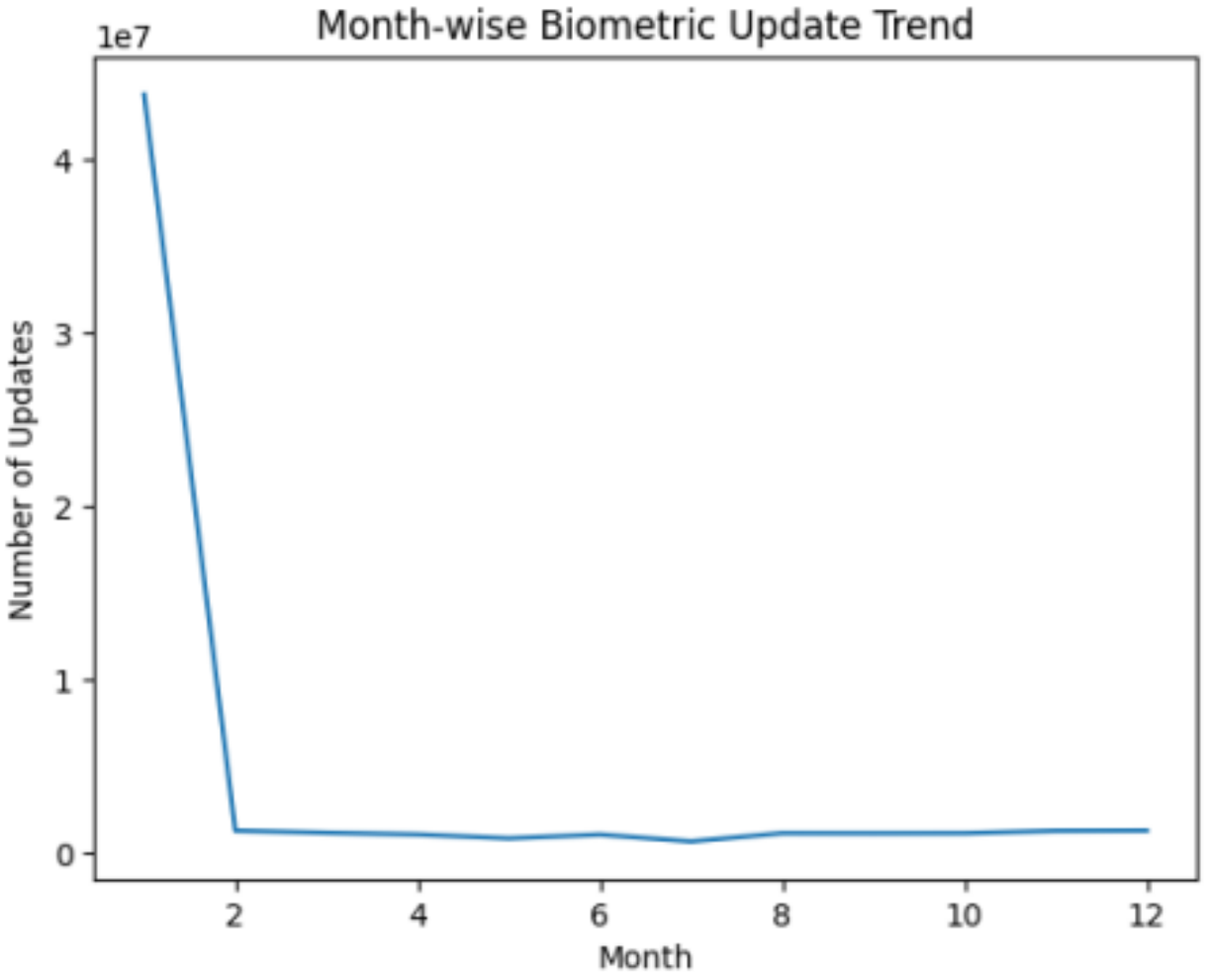
7.District-wise Distribution

District-wise analysis highlights hotspots such as Pune, Nashik, and Thane, indicating concentrated biometric service demand.



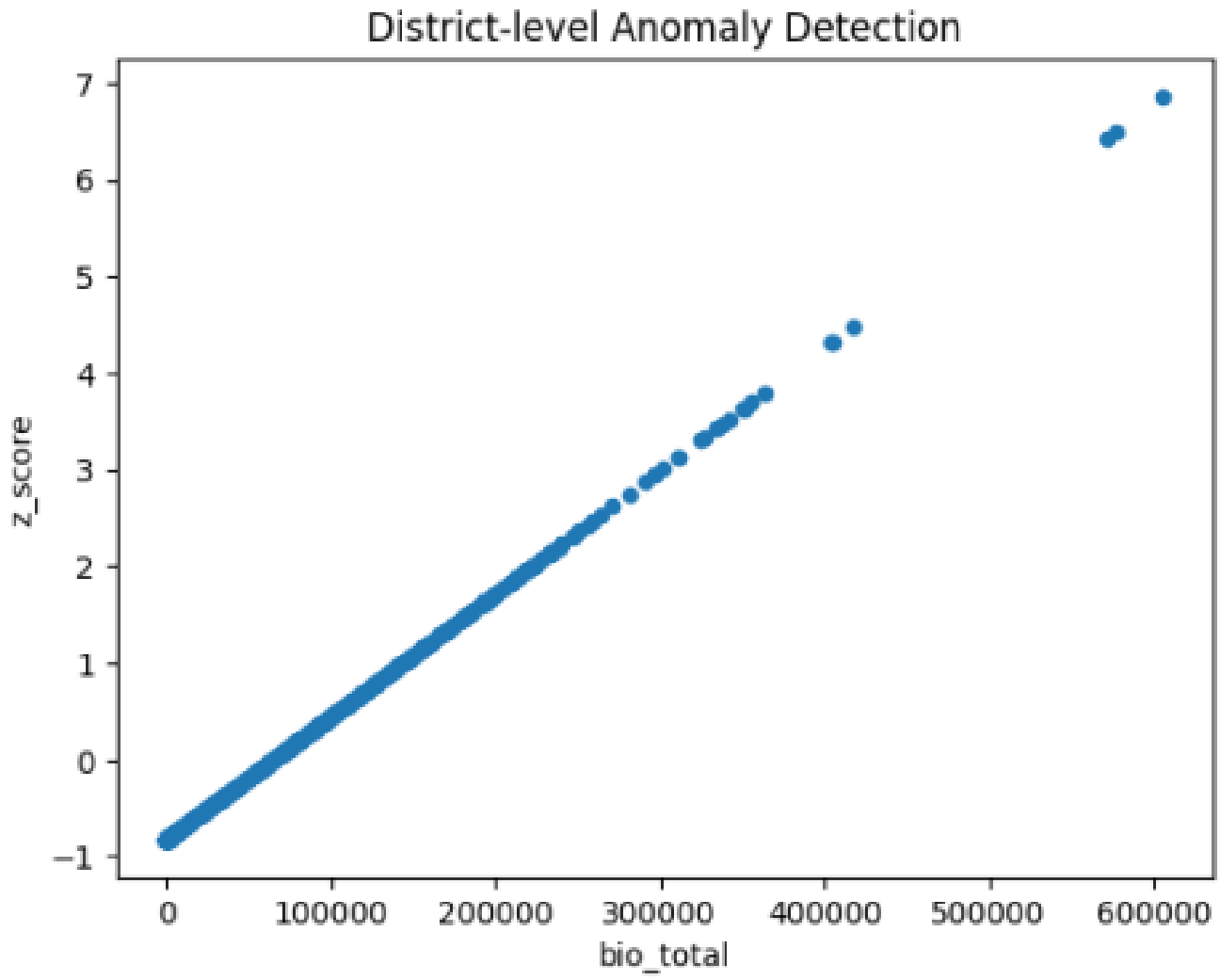
8.Monthly and Yearly Trends

Monthly trends show sharp spikes during early months, while the rest of the year maintains relatively stable activity, suggesting administrative or seasonal influences.



9.Anomaly Detection

Z-score based anomaly detection was applied at the district level. Districts with extremely high update counts were flagged as anomalies, helping identify overburdened service zones.



10,Key Insights

- Adult (18+) group slightly dominates biometric updates.
- Strong spikes occur during early months.
- Uttar Pradesh has the highest biometric load.
- Urban districts dominate the hotspot list.
- Service demand is geographically uneven.
- District-level anomalies highlight overburdened regions.
- Monthly patterns are non-uniform.
- Population density correlates with biometric volume.
- Transition spikes reflect adulthood-related documentation.
- Targeted planning can improve service efficiency.

11.Conclusion

The biometric dataset demonstrates strong age-based, regional, and seasonal patterns. The presence of concentrated hotspots and anomalies indicates the need for region-specific service planning. These insights can support policy-making, infrastructure scaling, and efficient identity management system.