Telangana PDS Analytics

Multi-Dimensional Shop Performance Clustering & Behavioral Profiling

1. Executive Summary

This project applies unsupervised learning techniques to analyze Fair Price Shops (FPS) under the Telangana Public Distribution System (PDS). Using KMeans clustering and HDBSCAN anomaly detection, shops were segmented into behavioral personas. A Behavioral Intensity Index was developed to enable monitoring prioritization in a policy-neutral framework.

2. Data Acquisition & Methodology

Multi-year datasets (2023–2025) were consolidated using shop identifiers. Feature engineering included Utilization Ratio, Portability Ratio, Rice-Wheat Intensity, Log Transaction Volatility, and a composite Behavioral Intensity Index. Clustering was performed using KMeans, while HDBSCAN identified behaviorally distinct shops. PCA was used for dimensionality reduction and visualization.

3. Cluster Insights (KMeans Personas)

Stable Rural Mainstream: Moderate utilization, low portability, stable demand patterns.
Urban Mobility-Driven: High portability and elevated transaction intensity.
Low-Variability Controlled: Consistent transaction behavior with minimal volatility.
High-Portability Transit Hubs: High volatility and mobility-driven demand dynamics.

4. HDBSCAN – Behavioral Distinct Shops

HDBSCAN identified a small percentage of shops as statistically distinct. These shops demonstrate extreme volatility or portability patterns relative to cluster norms. They represent monitoring-priority cases for contextual review rather than direct compliance assumptions.

5. Modeling Validation Metrics

Cluster quality was evaluated using Silhouette Score and PCA-based separability analysis. KMeans demonstrated moderate-to-strong separation across behavioral personas. HDBSCAN provided robust anomaly isolation with minimal noise percentage, indicating strong density-based cluster structure.

6. District-Level Behavioral Intelligence

District Behavioral Intensity Ranking revealed significant inter-district variation. Urban districts demonstrated higher portability-driven intensity, while rural districts maintained stable usage patterns. This supports district-sensitive monitoring strategies rather than uniform policy enforcement.

7. Dashboard & Visualization Framework

An interactive Streamlit dashboard was developed with:
• District & Year filters
• Monitoring Sensitivity Slider
• Geospatial Cluster Mapping (Folium)
• PCA Cluster Projection
• Radar-Based Persona Comparison
• Shop-Level Deep Dive Analysis
• Anomaly Intelligence View
• Downloadable Filtered Data

8. Data-Driven Recommendations

• Implement Behavioral Intensity Index in periodic state monitoring.
• Allocate dynamic stock buffers in high-portability regions.
• Conduct contextual audits for behaviorally distinct shops.
• Integrate dashboard insights into policy review cycles.
• Refresh clustering model periodically to capture behavioral drift.

9. Strategic Impact & Conclusion

This project demonstrates scalable application of unsupervised learning in public policy analytics. By transforming administrative transaction data into behavioral intelligence, the system supports data-driven governance, monitoring prioritization, and strategic resource allocation across districts.