**Market Insights Report Buyhatke – Section 2**

**Introduction:**

This report transforms a **2-month transactional audio dataset** into actionable, market-level analytics using Python's data analysis libraries like Pandas and Seaborn. It moves beyond basic summaries to highlight:

* A Pareto analysis of **category and brand concentration** to identify top market drivers.
* **Price sensitivity**, explored by segmenting products into four distinct tiers based on price quantiles.
* Key **customer archetypes** defined by a robust RFM (Recency, Frequency, Monetary) analysis.
* **Geographic sales trends** across top-performing states and cities.
* **Customer retention patterns** visualized through a monthly cohort analysis.

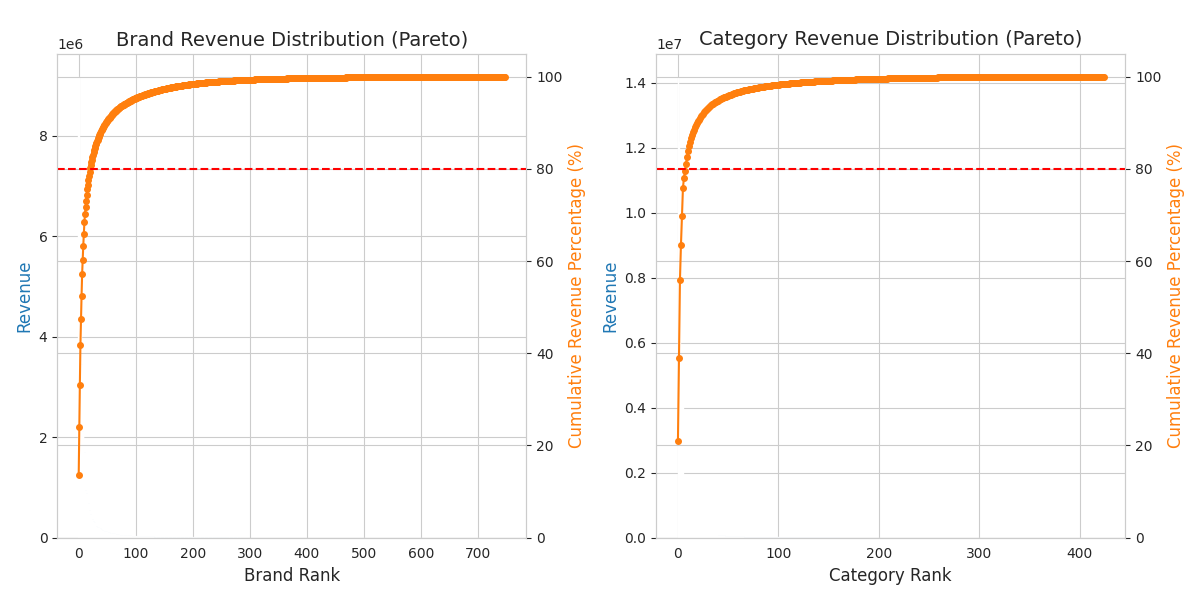
The resulting insights are translated into actionable priorities designed to grow revenue, reduce churn, and exploit cross-sell synergies across key categories and regions.

**1. Key Findings**

**Category & Brand Concentration:**

The analysis reveals a sharp Pareto (80/20) effect, which is far more pronounced than the typical rule. A mere ,

1.88% of product categories and 2.67% of brands are responsible for generating a full 80% of total revenue. This highlights an extreme market concentration. An analysis of the top-performing categories further shows that a few key brands consistently dominate sales, indicating that this concentration exists at both the market and category levels



**Figure 1:** Brand & *Category revenue share (Pareto plot)*

**Price Sensitivity & Elasticity:**

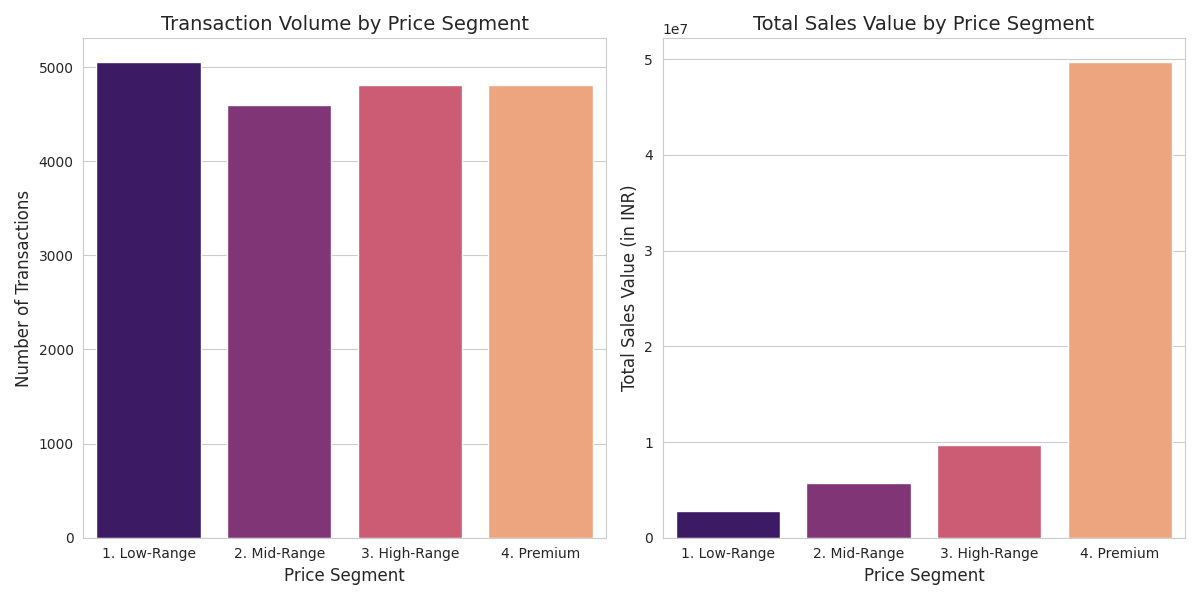
The analysis shows a clear difference in market behavior across price points. By segmenting products into four tiers based on price, we found that transaction volumes are surprisingly consistent across them, with the "Low-Range" tier accounting for 5,053 transactions and the "Premium" tier for 4,808.

However, the value generated is highly concentrated at the top. The "Premium" segment alone drove ₹49.7 million in sales, vastly outperforming the ₹2.8 million from the "Low-Range" segment. This indicates that while lower-priced items have high demand elasticity (driving volume), the premium category is the primary revenue driver and appears to be far more price-inelastic

Note: In code (step -6) the products are splitted based on their price into 4 parts like:-

Low-Range : segment containing the products with the lowest prices

**Premium**: segment containing the products with the highest prices

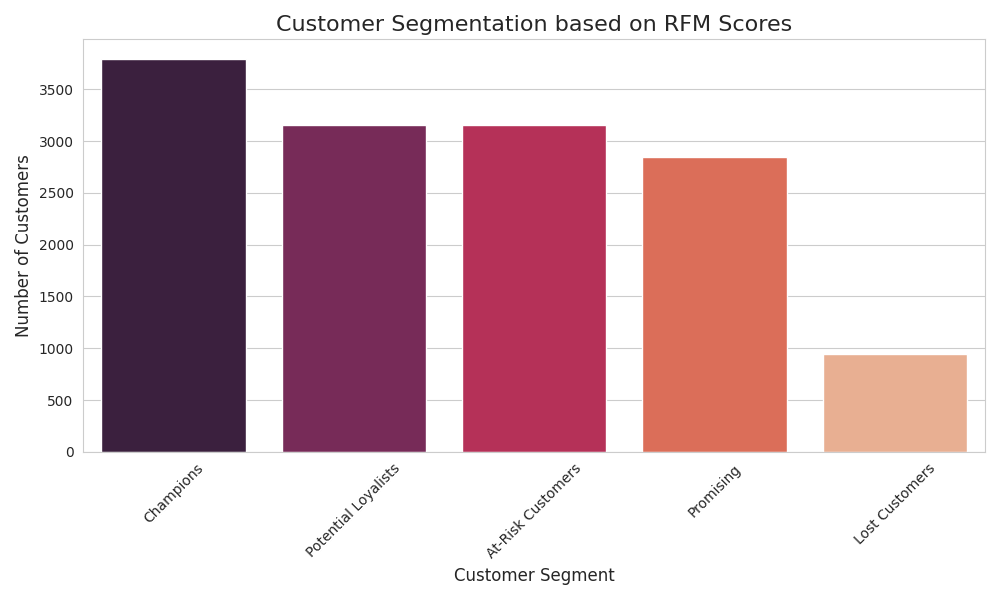


**Figure 2:** *Transaction Volume and Sales Value by Price Segment***Customer**

**Customer Behavior (RFM Analysis):**

An RFM (Recency, Frequency, Monetary) analysis was used to group customers into five distinct segments based on their transaction history. The analysis shows a large base of high-value customers but also a significant group at risk of churning:

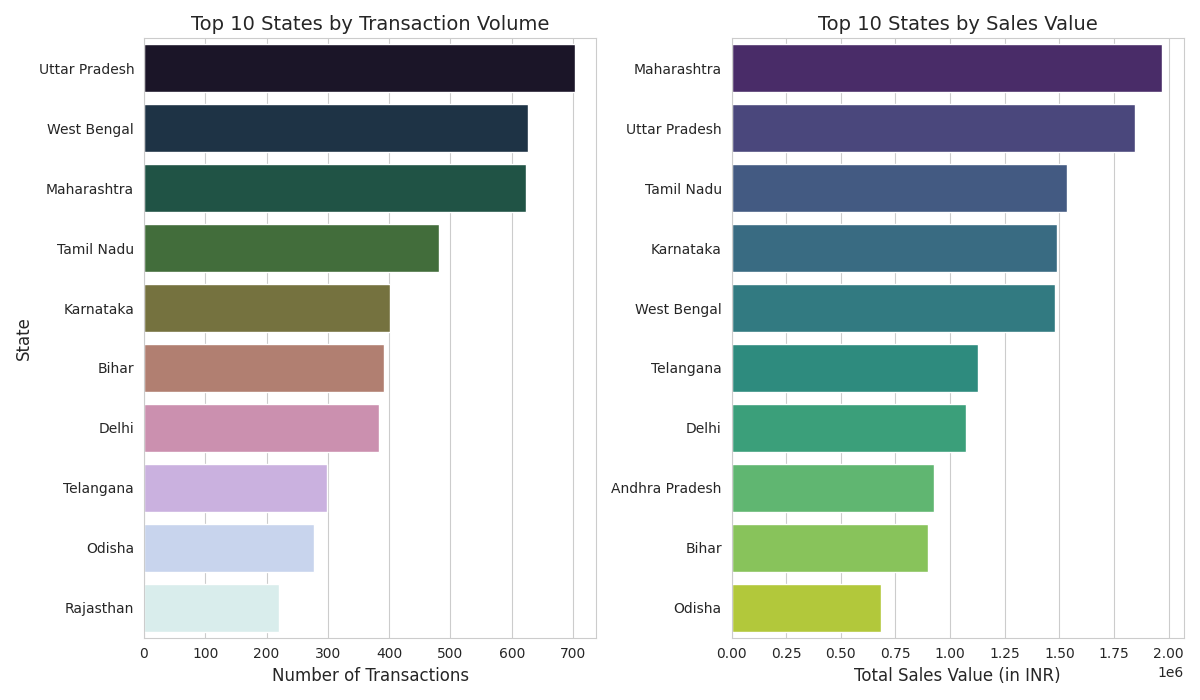
* Champions (~3,700 customers): The most valuable group, consisting of recent and frequent buyers (High Recency + High Frequency scores).
* Potential Loyalists (~3,150 customers): Recent customers who are not yet frequent purchasers (High Recency + Low Frequency). They represent a key opportunity for growth.
* At-Risk Customers (~3,150 customers): Customers who were frequent buyers but have not purchased recently (Low Recency + High Frequency). They require targeted reactivation campaigns.
* Promising (~2,850 customers): A broad group of customers with varied purchasing patterns who do not fall into the other main categories.
* Lost Customers (~950 customers): The lowest value segment, with low scores in recency and frequency, indicating a high probability of churn.



**Figure 3:** *RFM(Recency, frequency& Monetary Value) Analysis Result*

**Geographic Trends:**

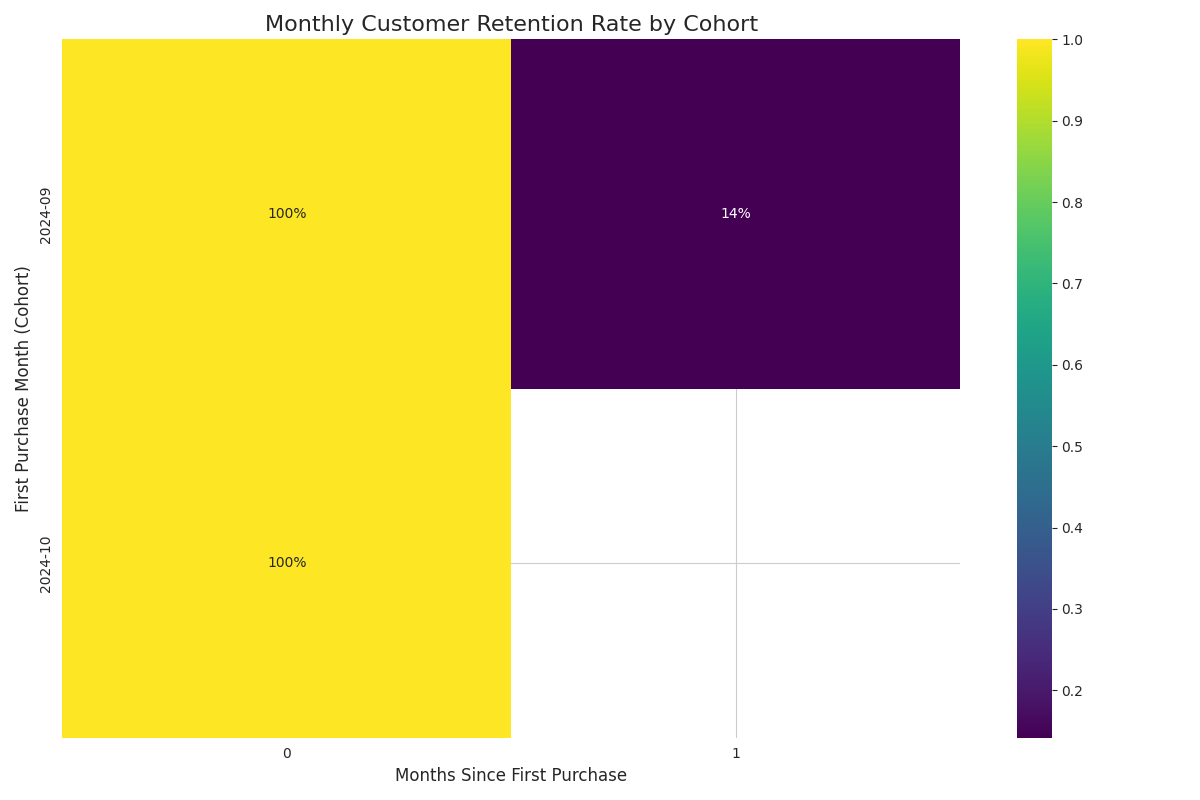
An analysis of the 5,954 transactions with available location data reveals that sales are highly concentrated in a few key states and cities. There is a notable difference between transaction volume and sales value at the state level. While Uttar Pradesh leads in the number of transactions, Maharashtra is the top state for total sales value, indicating higher average order values. Other states like West Bengal, Tamil Nadu, and Karnataka also rank in the top five for both metrics. At the city level, New Delhi is the clear market leader, ranking highest in both transaction volume and total sales value.



**Figure 4:** *Geographic analysis by Transaction or Value*

**Cohorts & Retention:**

A cohort analysis was performed by grouping customers based on the month of their first purchase. The resulting retention heatmap shows a clear result for the initial cohort from September 2024. Of the customers acquired in that month, 14% returned to make a repeat purchase in the following month (October 2024). This provides a baseline metric for month-over-month customer retention.



**Figure 5:** *Customer Retention Trends Across Monthly Cohorts*

**2. Market-level Insights (what the data implies)**

* The market is extremely concentrated. The Pareto analysis shows that just 1.88% of categories and 2.67% of brands drive 80% of all revenue. This means focusing on these core drivers can yield rapid results.
* A single pricing strategy is ineffective across the board. While "Low-Range" items generate high transaction volumes, the "Premium" segment drives disproportionately high revenue (₹49.7M), suggesting it is far less price-sensitive and margins should be protected.
* The customer base is clearly segmented by behavior. There is a large, valuable core of "Champions" to nurture, but also a significant group of "At-Risk Customers" (who were frequent but haven't purchased recently) that requires immediate reactivation efforts.
* Geographic performance varies significantly. For example, Uttar Pradesh leads in transaction volume, but Maharashtra generates higher total sales value. This points to opportunities for localized assortment and marketing campaigns.
* The initial month is critical for long-term customer retention. The cohort analysis showed that 86% of new customers from the first cohort did not return in the following month, highlighting the need for strong onboarding and early engagement strategies.
* The clear product tiers and customer segments suggest a strong potential for synergy. Building cross-category bundles (e.g., pairing a "Premium" audio product with a "Low-Range" accessory) could lift average order value (AOV) and improve retention among "Potential Loyalists."

**3. Actionable Recommendations**

1. Laser-Focus on Top Categories & Brands: Instead of a broad "top 20%" approach, allocate the majority of merchandising and marketing budget to the critical ~2% of categories and ~3% of brands that drive 80% of revenue. Secure partnerships or exclusives with these key brands to solidify market share.
2. Implement Tier-Based Pricing Strategies: Use the price segmentation analysis to create targeted rules. Deploy aggressive discounts and promotions for the high-elasticity "Low-Range" and "Mid-Range" segments to drive volume, while protecting margins and using value-based pricing for the low-elasticity "Premium" segment.
3. Deploy Segment-Specific Lifecycle Marketing: Build targeted customer journeys based on the RFM analysis:
   * Champions: Nurture with a VIP loyalty program.
   * Potential Loyalists: Engage with targeted cross-sell and up-sell campaigns to increase their purchase frequency.
   * At-Risk Customers: Launch automated win-back campaigns with behavior-triggered coupons to prevent churn.
4. Optimize Geo-Targeted Campaigns: Use the geographic insights to tailor marketing. For example, focus on volume-driving offers in Uttar Pradesh while highlighting premium product availability in the higher-value Maharashtra market.
5. Aggressively Target First-Month Retention: Given the 86% customer drop-off after the first month, immediately implement a dedicated 0–30 day onboarding experience with welcome emails, a first-purchase follow-up offer, and quick cross-sell recommendations to significantly raise the repeat purchase rate.
6. Measure & Iterate: Track key performance indicators (LTV, repeat purchase rate, AOV) for each of the identified RFM and geographic segments. Use these segments as a foundation for running A/B tests on new pricing, bundling, and onboarding strategies**.**

**4. Appendix / Next Steps**

* Plots and Visualizations: This analysis produced several key visualizations, including Pareto charts, price segment comparisons, RFM segment distributions, top state/city bar charts, and a cohort retention heatmap. A future iteration could enhance the geographic analysis with a geo-heatmap to identify more granular, pincode-level opportunities.
* Next Analytic Steps: The logical next steps are to combine this transactional data with session-level web behavior to improve personalization, compute SKU-level inventory concentration to optimize stock, and run formal uplift tests for the recommended promotions to measure their direct impact on revenue.