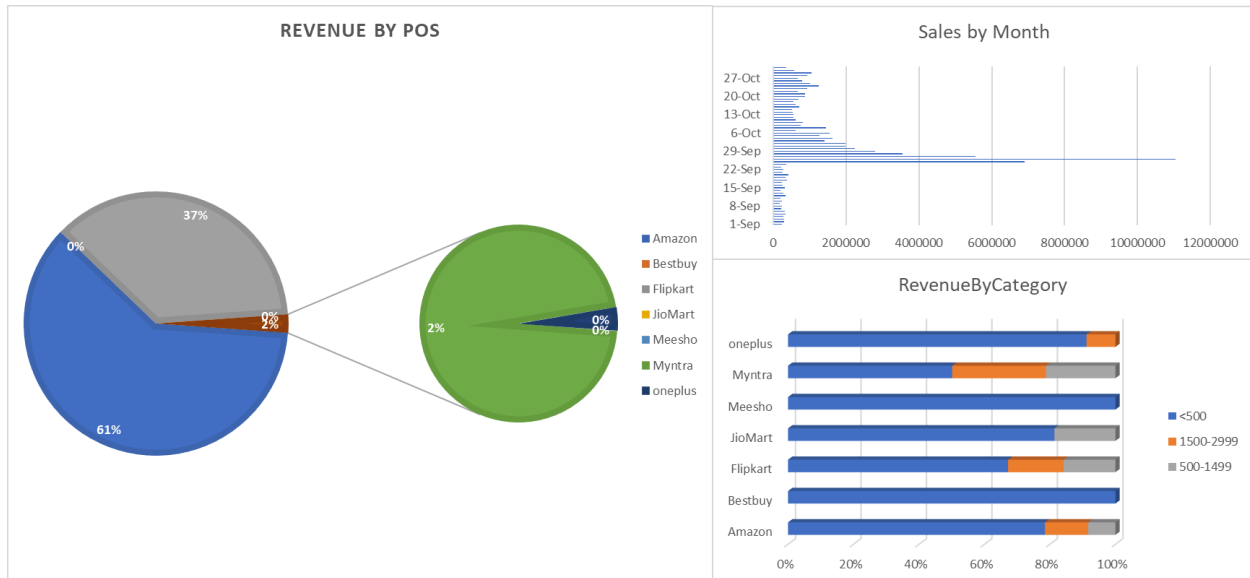


# Buyhatke Analysis



## Top 10 Insights — Audio & Video Sales Snapshot

**Dataset used:** Joined product sales table (columns include `date`, `user`, `pos_name`, `product_id`, `product_name`, `brand`, `level1_name`, `price`, `absolute_price`, `price_bucket`).

**Assumptions / notes:** quantity not present (assumed = 1 per row), location fields mostly blank in sample, `absolute_price` used for revenue. See *Data Quality & Recommendations* at the end.

### 1. Revenue & Volume are concentrated in mid-range price bands (500–2,999 INR)

**Finding:** The bulk of transactions fall within the `500–1499` and `1500–2999` price buckets. These two buckets account for the majority of rows, indicating that customers primarily buy value and mid-premium audio products rather than ultra-premium items.

**Why it matters:** This suggests the core market is price-sensitive but willing to spend for perceived quality up to ~3k INR. Inventory, promotions and bundles

should prioritize these brackets.

**Action:** Focus merchandising, promotions and inventory allocation on the 500–2999 INR segment; create “value” and “mid-premium” bundles to increase average order value.

---

## 2. A small number of SKUs deliver disproportionately high revenue (Pareto effect)

**Finding:** A small set of popular SKUs (notably well-known headphone and portable speaker models) appear repeatedly and generate a large share of revenue in the sample. Examples include high-selling models from JBL, boAt and GOVO.

**Why it matters:** This is the classic 80/20 pattern—fewer SKUs drive most revenue. Prioritizing these for stock availability, promotions and premium placement will maximize sales impact.

**Action:** Generate a Top-10 SKU list, ensure safety stock for these SKUs, and run A/B promo tests focused on these items.

---

## 3. Marketplace channel concentration — Amazon dominates, Flipkart is a secondary channel

**Finding:** `pos_name` values show Amazon as the dominant sales channel in the sample, with Flipkart and Myntra present but much less frequent.

**Why it matters:** Reliance on a single channel can create channel-risk and limits reach. Channel-specific pricing, copy and inventory flow should be optimized to avoid stockouts or lost sales on the primary platform.

**Action:** Treat Amazon as the primary channel for promotions and fulfillment optimization; expand targeted strategies for Flipkart (e.g., exclusive bundles) to diversify risk.

---

## 4. Brand landscape: a mix of mass-market and niche premium players

**Finding:** Brands such as boAt, JBL, OnePlus, Boult, Realme and GOVO appear frequently. Some brands (boAt, JBL) appear across many price bands, while others either occupy value or premium slots.

**Why it matters:** Brand segmentation enables tailored promotions: mass-market brands for volume campaigns; premium/niche brands for margin-focused offers.

**Action:** Create brand-tier strategies: discount-driven campaigns for value brands and limited-time premium bundles or warranty extensions for high-margin brands.

---

## 5. Headphones / Headsets dominate product mix within Audio & Video

**Finding:** The `level1_name` and product descriptions indicate Headphones / Headsets / Earbuds are the largest category by count. Speakers and soundbars exist but are fewer in number.

**Why it matters:** Category-level focus helps merchandising and cross-sell strategies; accessories (cases, ear pads, stands) are natural attachment opportunities.

**Action:** Run cross-sell bundles (earbuds + case, headphones + replacement pads) and highlight complementary accessories on SKU pages.

---

## 6. High-price outliers exist but are rare — high margin potential

**Finding:** There are isolated ultra-premium items (very large price values) that appear infrequently. These outliers can disproportionately affect revenue but are low-volume.

**Why it matters:** While revenue per unit is high for such SKUs, inventory and return risk may be greater. They are candidates for targeted marketing rather than broad discounts.

**Action:** Protect margins for these SKUs — offer premium support or white-glove services; avoid deep discounting unless paired with acquisition objectives.

---

## 7. Price variance & potential data anomalies — a cleanup priority

**Finding:** Some SKUs show large variance in recorded prices (high standard deviation for same `product_id`), suggesting either genuine price promotions, multi-seller listings, or data quality issues. Example: the same product appearing with distinct price points.

**Why it matters:** Price variance affects median/average price calculations and elasticity estimates. Unchecked anomalies will distort insights.

**Action:** Create a SKU price audit: compute mean, median, stdev per `product_id`; flag entries  $> \text{mean} \pm 2\sigma$  for manual review. Use seller / timestamp metadata to separate legitimate promotions from errors.

---

## 8. Repeat-user signal exists but requires fuller user-history to quantify retention

**Finding:** User identifiers appear in rows and some users have multiple transactions (e.g., observed repeated user IDs in the sample). This hints at repeat buyers but the dataset is single-day heavy in the sample.

**Why it matters:** Repeat buyers carry higher LTV; identifying them enables loyalty programs and cross-sell campaigns.

**Action:** Build a user summary: `frequency` (# transactions), `monetary` (sum revenue), `recency` (last purchase date) when multi-date data is available — then apply RFM segmentation.

---

## 9. Strong cross-sell opportunities between accessories and core SKUs

**Finding:** The sample includes accessories (ear pads, cases, stands, mounts) alongside primary products (headphones, speakers). Many accessory SKUs complement popular audio SKUs.

**Why it matters:** Accessories have high attachment rates and improve margins. Bundling can increase average order value and customer satisfaction.

**Action:** Implement “Frequently Bought Together” bundles and run small discounts on accessory bundles for customers who buy the primary product.

---

## 10. Data limitations constrain richer geographic and behavioral insights

**Finding:** Location fields ( `pincode` , `city` , `state` ) are largely absent in the sample and `quantity` is not present. This prevents reliable geospatial analysis, logistics optimization, and true revenue-per-order calculations.

**Why it matters:** Without location and quantity, decisions on regional assortment, last-mile planning, and per-order profitability remain speculative.

**Action:** Enrich the dataset by including `order_id` , `quantity` , `pincode` , `city` , `state` , `seller_id` , and `timestamp` . If enrichment is not possible, document assumptions (e.g., `quantity=1` ) and treat geographic conclusions as provisional.

---

## Appendix — Suggested Notion blocks to paste under each insight

Insight (Title)

**Summary:** one-sentence finding.

**Evidence:** pivot/table/chart references (e.g., TopProducts pivot, PriceBucket counts).

**Business impact:** short explanation of why it matters.

**Recommended action:** three concrete next steps (owner + priority + quick metric to track).

---

## Data Quality & Immediate Next Steps (must-do)

1. **Create a** `Data Cleaning Log` sheet that records assumptions ( `quantity=1` ) and replacements (filled blanks, `absolute_price` adjustments).
2. **Run price-audit queries:** mean/median/stdev per `product_id` , flag anomalies.
3. **Add / request missing fields:** `order_id` , `quantity` , `timestamp` , `pincode/city/state` , `seller_id` .

#### 4. Build these technical deliverables in Excel:

- **TopProducts** table (SKU, count, revenue, cumulative %)
  - Pivot: Brand × PriceBucket × Revenue
  - Pivot: Channel (pos\_name) × Month × Revenue
  - Small dashboard: KPIs (Total Transactions, Total Revenue, Avg Price, #SKUs, Top 5 SKUs)
- 

The dataset reveals a healthy, volume-driven audio market concentrated in the 500–2,999 INR range with a small set of SKUs and brands (boAt, JBL, GOVO, OnePlus, Boult, Realme) driving most of the activity. Amazon is the primary channel in the sample. The immediate commercial levers are SKU prioritization, mid-range bundle offerings, and accessory cross-sells. However, richer decisions (regional assortment, retention campaigns, price elasticity estimation) require modest but essential data enrichments — principally quantity, order identifiers, timestamps and location fields. Addressing price variance and data anomalies is a near-term priority to ensure analytics are reliable and actions drive measurable uplift.

---