Product Categories and Trends Analysis Report

Author: Taksh Girdhar

Contact: takshgirdhar@gmail.com

1. Context & Objective

The goal of this analysis is to understand how a brand's apparel is performing across categories, SKUs, time, and channels, and to surface actionable recommendations that support merchandise planning, inventory allocation, and cross-category strategy.

Using transaction-level sales data and Tableau dashboards built on top of a SQL and Python pipeline, I focused on four core questions:

1. Which product categories and SKUs are driving revenue and profit?
2. How are sales and margins trending over time and across channels?
3. How do customers move across categories (category affinity)?
4. Where are the biggest levers for profit: high-contribution SKUs, loyal customers (RFM), and markdown strategy?

2. Dataset

Dataset Overview: Clothing Sales Transactions Dataset

The dataset leveraged for this analysis is sourced from the Kaggle repository of suryaprabha19 and is titled *“Clothing Sales Transactions Dataset”*. Below is a detailed breakdown of its structure, contents, and how it supports the analytical goals of this project.

Source & Context

- Published on Kaggle: link to dataset
- Scope: Each row represents a single transaction (sale) of a clothing item, capturing product, pricing, customer, and channel-level details.
- Relevance: Ideal for a product analytics / merchandising planning scenario because it contains both sales and cost information, allowing profitability, margin, and trend analysis.

Key Fields & Structure

Some of the principal columns in the dataset include:

- saleID -unique identifier for each transaction.
- saleDate -timestamp of the sale (used to derive year/month and trend metrics).
- productID, productName, productCategory - product identifiers and category metadata.
- quantity, unitPrice, costPrice, totalAmount, totalCost - transactional financials and the raw data needed to compute revenue, cost, profit, and margin.
- customerID, customerName, location -
- enables customer segmentation, RFM analysis, and channel behaviour.
- salespersonID, salespersonName, salesChannel -channel and personnel dimensions for deeper analysis.
- status -e.g., “Paid”, “Returned”, “Cancelled”

This variety of fields provides a mix of product, customer, channel and temporal data which is exactly what is needed to support the multiple analyses executed later on (category profitability, RFM, markdown simulation, category affinity, etc.).

Data Quality & Pre-Processing Notes

- Completeness: A number of rows (~10,000 transactions) which is sufficient for structural patterns and segmentation.
- Clean-up steps applied:
  - Conversion of saleDate to datetime type and extraction of year, month, year_month.
  - Computation of derived fields:
    - revenue = totalAmount
    - cost = totalCost
    - profit = revenue – cost

        - margin_pct = profit / revenue
- Assumptions made:
    - No currency was provided, so it is assumed that the monetary value is in United States Dollar (USD).
    - Cost values were provided (costPrice, totalCost), so margin analysis is real rather than simulated.
    - For markdown efficiency, a simulation layer adds hypothetical discounting.
- Limitations:
    - If customers make single-item purchases frequently, traditional basket analysis (same transaction) may be sparse.

This dataset provides a robust foundation for the analyses performed in this project. It captures the essential dimensions (product, transaction, customer, channel, cost) needed for deep merchandising and planning analytics. With proper pre-cleaning and derived metric construction, it serves as an excellent candidate for the kind of data used in real-world apparel planning teams.

3. Executive Overview

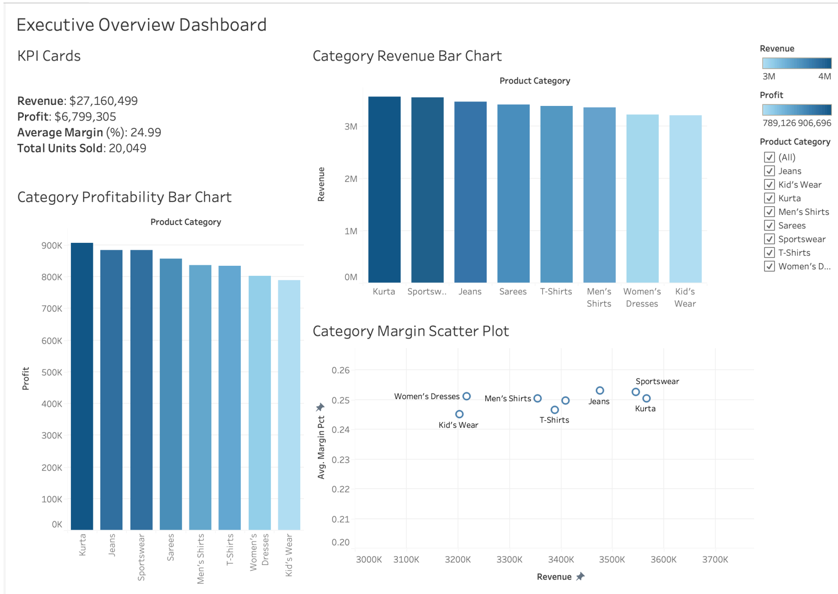
Key KPIs (full period)

- Revenue: ~$27,160,499 (USD)
- Profit: ~$6,799,305
- Average Margin: ~24.99%
- Total Units Sold: ~20,049

High-level takeaways

- Revenue and profit are well diversified across categories, but a small group of categories (Jeans, Kurtas, Sportswear, T-Shirts) account for a disproportionate share of sales.
- Margins are relatively stable by category, in a ~22–26% band, with some categories consistently above average.

> Exhibit 1:



4. Category Performance & Profitability

4.1. Revenue and Profit by Category

From the Category Revenue Bar Chart and Category Profitability Bar Chart:

- Top revenue categories are:
  - *Kurta - $3,566,053*
  - *Sportswear - $3,545,695*
  - *Jeans - $3,475,597*
  - *Sarees - $3,409,486*
- These same categories also lead in absolute profit, indicating that they are not just high-volume but also economically attractive.
- Kid’s Wear and Women’s Dresses sit at the lower end of revenue and profit, suggesting they may be more niche or under-represented segments at current levels.

Implication for planning

- These findings support prioritizing receipt and inventory allocation to Kurtas, Sportswear, Jeans, and T-Shirts, while running focused diagnostics on Kid’s Wear and Women’s Dresses to understand whether the opportunity is constrained by demand, pricing, or assortment.

4.2. Margin Profile by Category

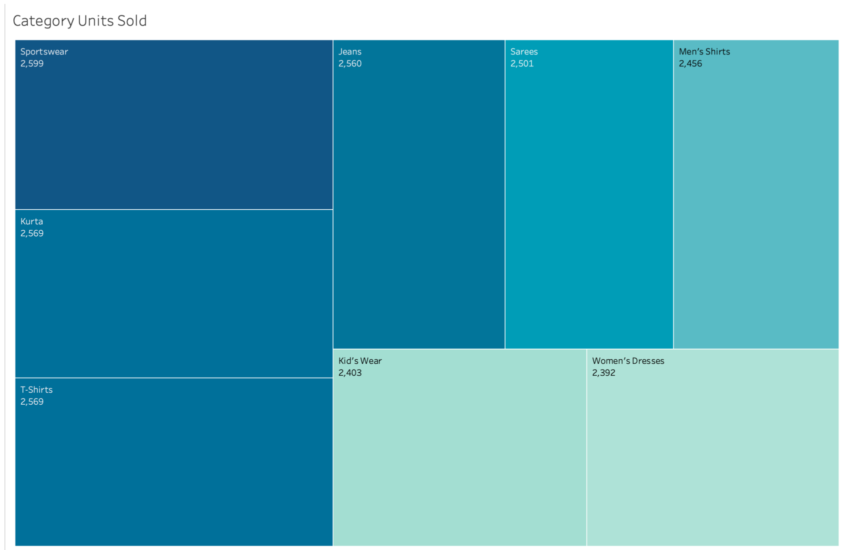
The Category Margin Scatter Plot (Revenue vs. Avg Margin %) shows:

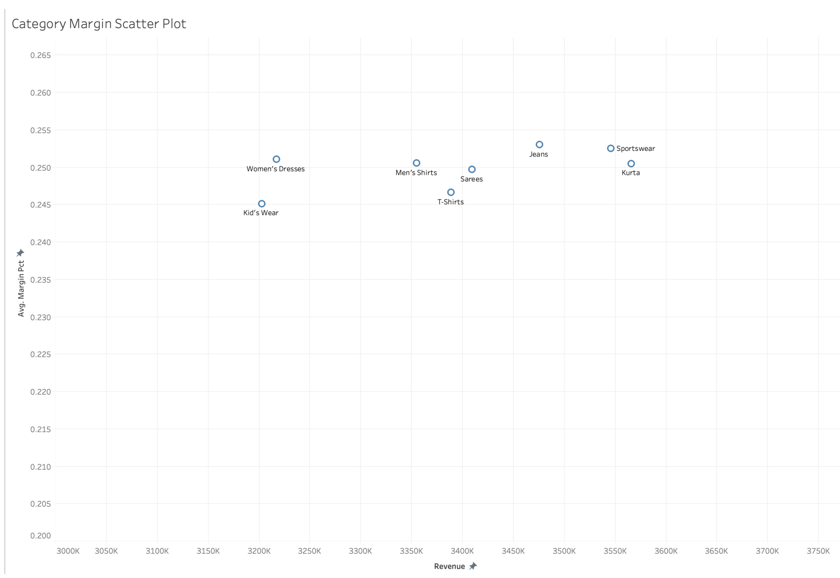
- Categories cluster tightly around an average margin of ~25%, with a few categories slightly above or below that line.
- For example, Jeans and Sportswear sit above the average margin line at similar revenue scale, indicating they are both high-volume and high-quality

Implication

- Because margins are fairly consistent, the primary lever at the category level is volume, not pricing. However, categories with structurally lower margins should be reviewed for cost, promotional reliance, or mix of SKUs.

> Exhibit 2:





5. Category Deep Dive: Volume, Mix & Margin Over Time

5.1. Units vs Revenue

The Category Units Sold treemap highlights:

- Sportswear, Kurtas, and T-Shirts are the highest-volume categories, aligning with their strong revenue contribution.

- Some categories (e.g., Women's Dresses, Sarees) have lower unit volume, reinforcing that they are smaller but potentially strategic segments (e.g., for occasion wear).

Implication

- Sportswear, Kurtas, and T-Shirts appear to be core categories, so it is critical to ensure adequate inventory in key sizes/colors.

5.2. Revenue vs Profit by Category

The Revenue & Profit Comparison chart (dual bars by category) shows:

- Most categories maintain a healthy profit contribution relative to revenue, reinforcing that pricing and cost structure are generally aligned.
- Note any categories where the gap between revenue and profit looks relatively weaker is a sign of lower margin or higher cost.

Implication

- For categories where revenue is strong but profit is relatively muted, there is an opportunity to revisit cost negotiations, pricing or markdown strategy.

5.3. Margin Heatmap by Month

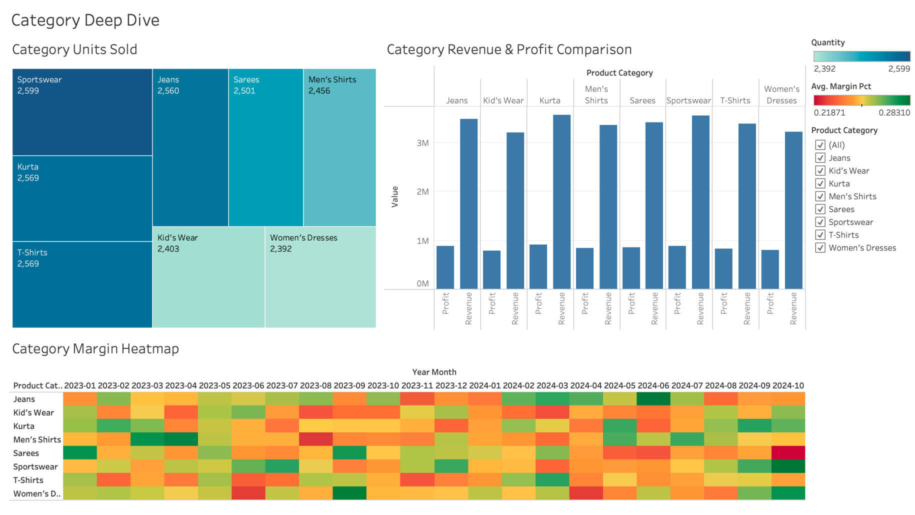
The Category Margin Heatmap across Year Month reveals:

- Margins fluctuate month-to-month by category, with distinct green (higher margin) and red (lower margin) periods.
- For example, Jeans tends to have stronger margins earlier in the year whereas Men's shirts tends to have stronger margins around the middle of the year.

Implication

- These patterns suggest seasonal or promotional dynamics which allows planning to:
  - Time markdowns in categories where margins naturally compress.
  - Protect margins in peak demand months by avoiding unnecessary discounts.

> Exhibit 3:



6. Product & SKU Analysis

6.1. Top and Bottom SKUs

From the Top 10 SKUs by Revenue and Bottom 10 SKUs by Revenue:

- Top performers are concentrated in a mix of categories, with a notable presence from Jeans, Kid’s Wear, Kurta, Sportswear, and T-Shirts.
- Bottom SKUs tend to come from smaller or more niche variants (e.g., specific Women’s Dresses or Sarees styles) with both low units and low revenue.

Implication

- The top SKUs form a clear “must-have” list that should be considered for:
  - repeat buys
  - expansion of color/size inventories
  - prominent placement across channels
- The bottom SKUs are candidates for:
  - markdowns,
  - assortment rationalization,
  - or testing different pricing / marketing if they are strategically important styles.

6.2. Profit vs Units Scatter

The Profit vs Units scatter displays a clear positive relationship: as units increase, profit tends to rise, with a few outlier SKUs generating very high profit at elevated unit volumes.

Implication

- Those high-unit/high-profit outliers should be explicitly tagged as “hero SKUs” and protected in planning (adequate inventory, prioritized in forecasting).

- Low-unit/low-profit SKUs create the long tail. Trimming the tail slightly would simplify the assortment without materially impacting profitability.

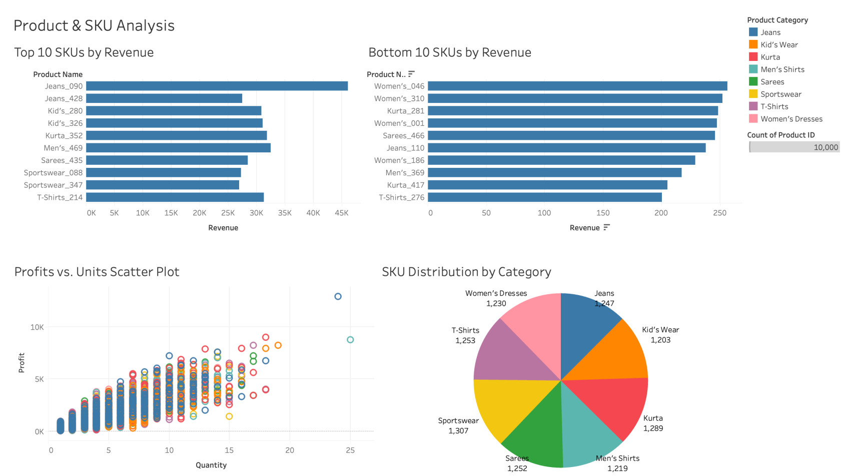
6.3. SKU Distribution by Category

The SKU Distribution pie shows that each category carries a relatively similar number of SKUs (roughly ~1.2–1.3K each).

Implication

- Since SKU counts are fairly balanced, performance differences are driven more by guest demand than by assortment depth.
- In underperforming categories, the answer is less about “more SKUs” and more about better-targeted product, more relevant styles, or price/value realignment.

> Exhibit 4:



7. Trends & Seasonality

From the Monthly Revenue, Units, and Profit Trend lines:

- Revenue, units, and profit show a gradual softening over the time horizon, with a pronounced drop in the final month (likely due to partial month data or end of dataset rather than true demand collapse).
- Earlier months show more stable performance, suggesting relatively consistent demand patterns with small seasonal shifts.

Implication

- The brand should be cautious about interpreting the final month as a trend break; instead, treat it as incomplete data unless confirmed otherwise.
- The relatively smooth curves are a good foundation for time-series forecasting to support future buy planning.

Channel Split

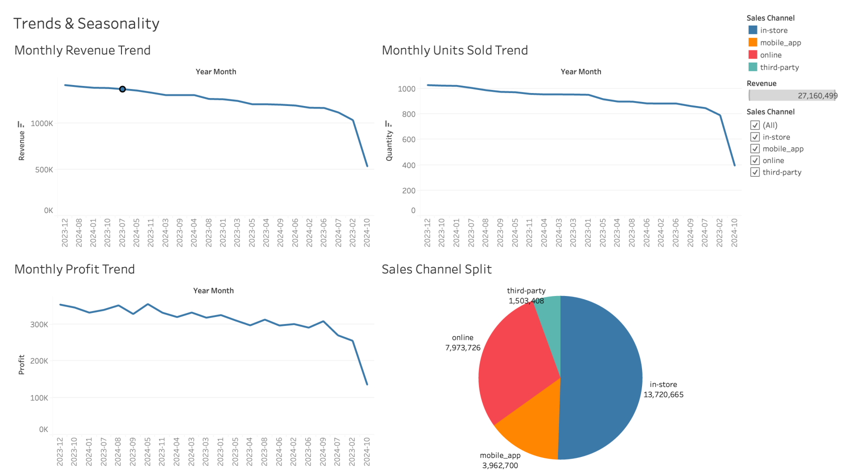
The Sales Channel Split pie chart indicates:

- In-store is the largest contributor (≈$13.7M, ~50% of revenue)
- Online is the second largest (≈$8.0M)
- Mobile app contributes a material mid-tier share (≈$4.0M)
- Third-party partners are a smaller but non-trivial channel (≈$1.5M)

Implication

- In-store is still the backbone of the business, but digital (online and mobile) collectively represents a substantial share and an important growth and experimentation surface.
- Category and SKU decisions should be made with channels in mind (for example, certain SKUs might be online only or mobile app first).

> Exhibit 5:



8. Margin Contribution

By aggregating profit at the SKU level and computing cumulative profit contribution:

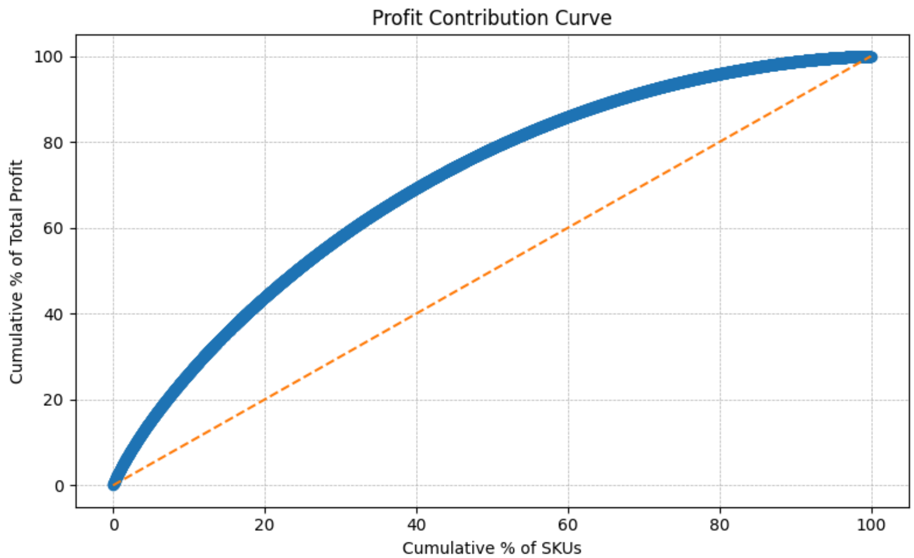
- The top ~20% of SKUs account for ~43.8% of total profit.
- These SKUs are concentrated in Kurta, Jeans, Sportswear.

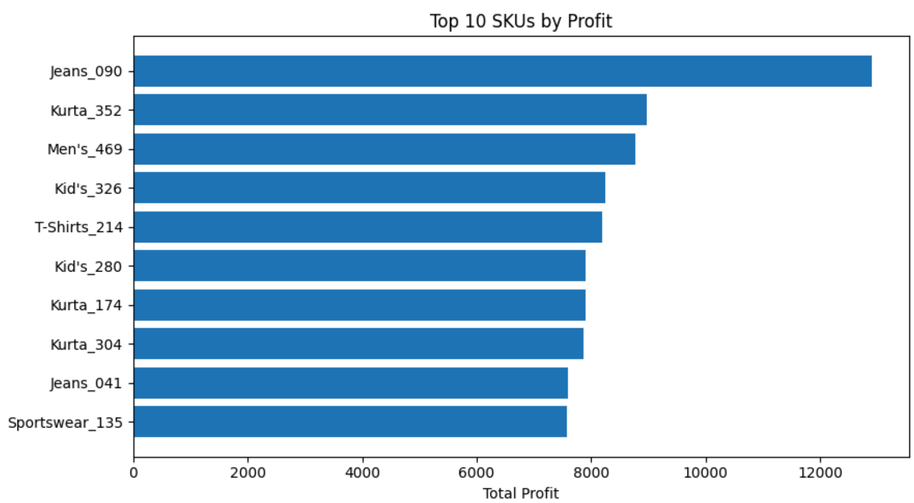
Implication

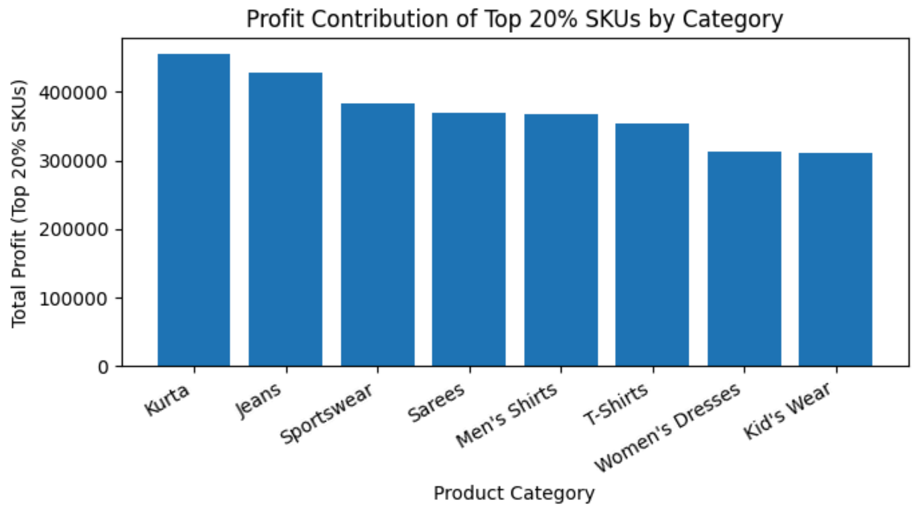
- These top SKUs should be treated as core to be:
  - prioritized in forecasting and allocation
  - protected from stockouts
  - reviewed for potential range extensions (new colors, fits)
- The remaining long-tail SKUs contribute comparatively little profit and can be candidates for:
  - SKU rationalization
  - Simplified assortments in smaller stores

- More targeted buys in future seasons

> Exhibit 7:







9. RFM-Based Customer Segmentation

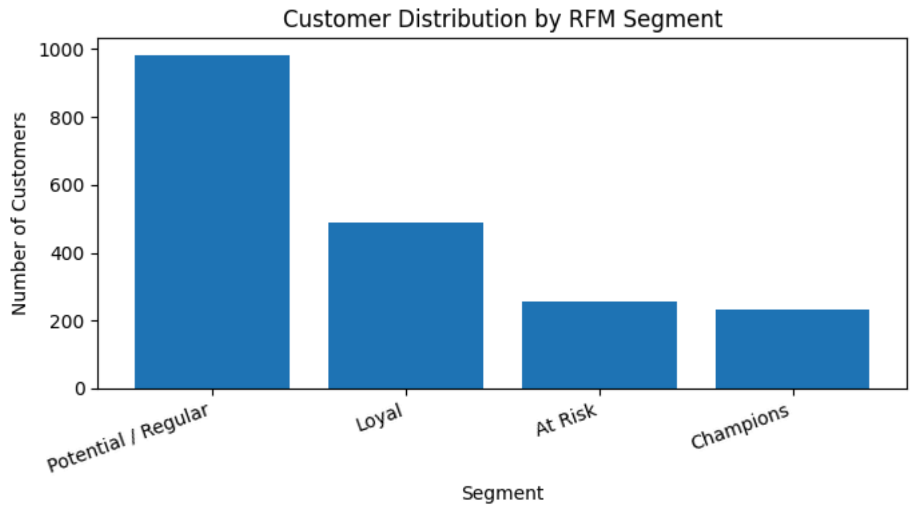
Using Recency, Frequency, Monetary (RFM) scoring at the customer level:

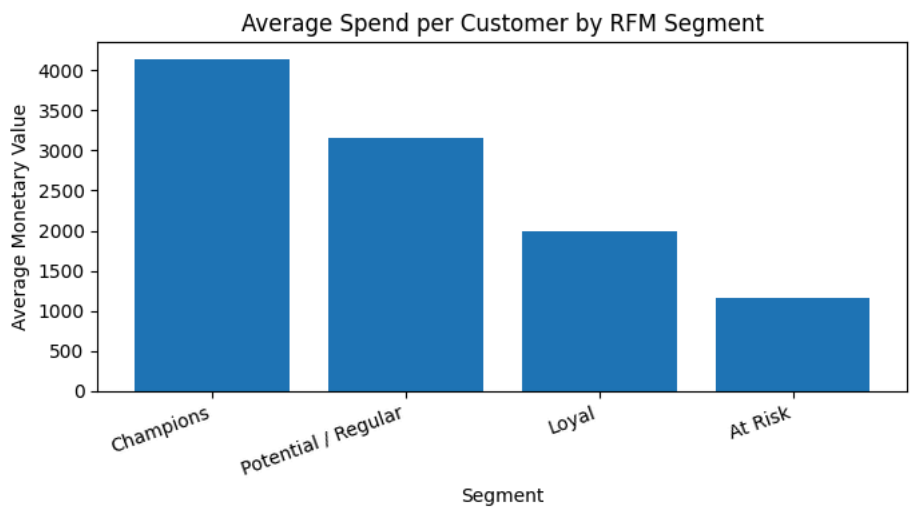
- Champions are high recency, frequency, and spend. They represent ~11.9% of customers but contribute ~39.6% of revenue.
- Loyal customers purchase regularly but may have slightly lower spend per order.
- At Risk customers show older last-purchase dates and lower frequency, contributing modest but non-negligible revenue share.
- Potential / Regular customers sit between these segments.

Implication

- For planning and marketing:
    - Champions & Loyal: focus on newness and exclusivity; they are good candidates for early access or cross-sell into adjacent categories.
    - At Risk: consider win-back campaigns (targeted offers, email/app push).
    - Potential: nudge towards higher frequency via bundles or outfit recommendations.

> Exhibit 8:





10. Markdown Efficiency (Simulated Scenarios)

To understand markdown sensitivity, I modelled 10%, 20%, and 30% discount scenarios at the category level, assuming constant units and stable cost:

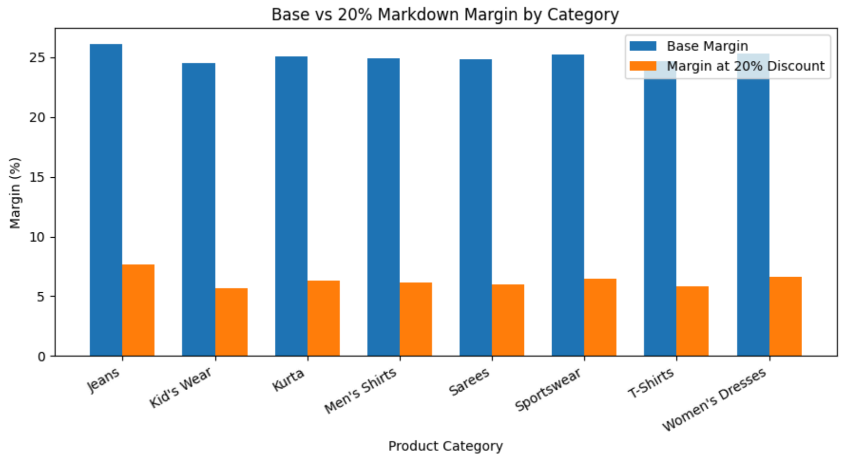
- At 10% markdown, most categories maintain healthy margins, with modest profit erosion.
- At 20% markdown, some categories such as Kurtas, Sportswear, Men’s Shirts and T-Shirts and still retain >5% margin, while others drop sharply.
- At 30% markdown, the decline in margin is pronounced.

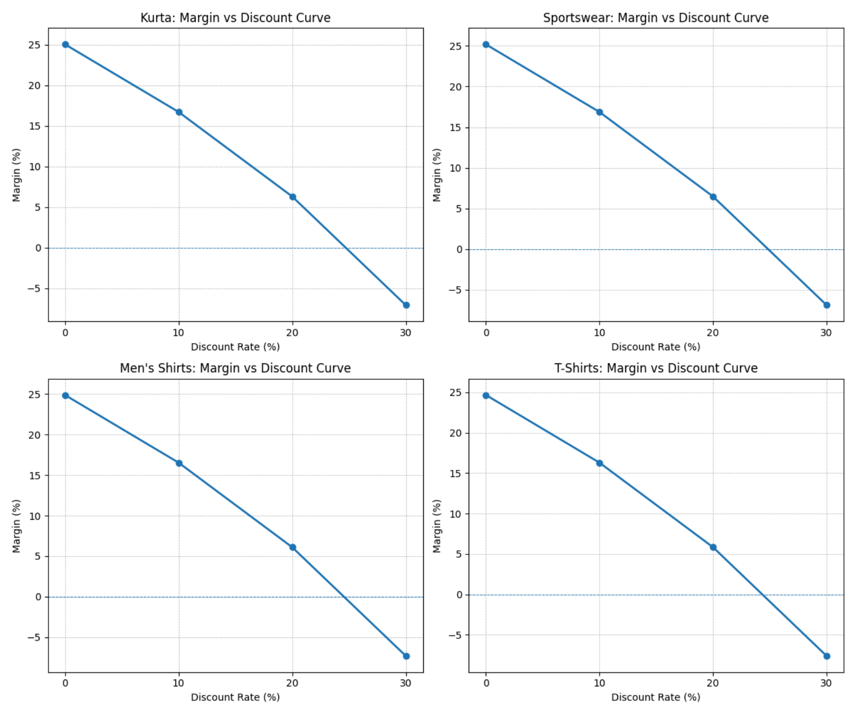
Implication

- Markdown strategy should be category-specific:
  - Categories that remain above acceptable margin thresholds even under 20% markdowns are good candidates for promotions to drive volume or clear inventory.

◦ Categories whose margins collapse under moderate markdowns should be managed with tighter buys upfront to minimize the need for heavy discounting.

> Exhibit 9:





11. Recommended Next Steps

From a planning and analytics perspective, the next actions I would recommend are:

1. Deep-dive on top-performing categories (Jeans, Kurta, Sportswear, T-Shirts)
    - Validate size curves, color mix, and store/channel coverage.
    - Ensure buy plans for upcoming seasons protect these core categories.

2. Targeted interventions for underperforming categories (e.g., Kid’s Wear, Women’s Dresses)
    - Diagnose whether issues stem from price, product fit, or lack of awareness.
    - Consider curated capsule assortments instead of broadly expanding SKUs.
3. Leverage category affinity in merchandising & digital
    - Align store adjacencies and mannequin stories with top affinity pairs.
    - Implement simple rule-based recommendations online/app (“Complete the look”).
4. Focus on high-contribution SKUs in planning
    - Treat top 20% SKUs as non-negotiable in buys.
5. Layer RFM and markdown insights into lifecycle planning
    - Use RFM segments to target markdowns smartly, not uniformly.
    - Pair win-back offers with categories that have high margin.