

# ShopPulse AI

## The Intelligent Customer Segmentation Engine

Democratizing Data Science for E-Commerce via Generative AI

**Praxis 2.0 Hackathon Submission**

Team: CodeCrusher India

Repository: <https://github.com/codecrusherindia/ShopPulse-AI>

Live Link: <https://shoppulse-ai-my12.onrender.com>

# Table of Contents

1. Executive Summary
2. The Problem Landscape
3. The Solution
4. Technical Architecture
5. Data Schema & Input
6. AI Implementation Strategy
7. Business Feasibility & Market
8. Ethical Considerations
9. Future Roadmap

# 1. Executive Summary

ShopPulse AI is a next-generation business intelligence tool designed to bridge the gap between raw e-commerce transaction data and actionable marketing strategy. While traditional analytics tools show what happened (descriptive analytics), ShopPulse AI explains who your customers are and what to do next (prescriptive analytics).

By fusing **Unsupervised Machine Learning (K-Means Clustering)** with **Generative AI (Google Gemini 2.5 Flash)**, the platform serves as an automated Chief Marketing Officer (CMO) for Small and Medium Enterprises (SMEs). It automatically segments user bases, generates psychological personas, and drafts strategic advice, requiring zero code or data science knowledge from the user.

# 2. The Problem Landscape

## The SME Data Gap

Small to mid-sized e-commerce businesses (Shopify, WooCommerce, Etsy) sit on mountains of customer data but lack the expertise to unlock its value.

- **The 'Black Box' of Data:** A typical merchant has access to thousands of rows of transaction data (CSV exports) but no way to interpret it beyond basic 'Total Sales' line charts.
- **One-Size-Fits-None Marketing:** Without segmentation, merchants send the same generic discount emails to High-Spending VIPs (wasting margin) and Churned Customers (irrelevant messaging).
- **Lack of Strategic Empathy:** Stakeholders often view customers as just 'User IDs' rather than people with specific needs. They lack the budget for enterprise tools like Salesforce Einstein.

# 3. The Solution

ShopPulseAI automates the entire data science workflow in a four-step pipeline:

1. **Ingestion:** Instantly processes raw CSV transaction files using intelligent column mapping.
2. **Segmentation (The Math):** Uses K-Means Clustering to mathematically group customers based on Recency, Frequency, Monetary value (RFM), and Satisfaction.
3. **Personification (The AI):** Uses Google Gemini 2.5 Flash to analyze the mathematical centroids and create detailed 'Buyer Personas' with psychological depth.
4. **Prescription:** Provides concrete, actionable advice on inventory, pricing, and marketing channels for each specific segment.

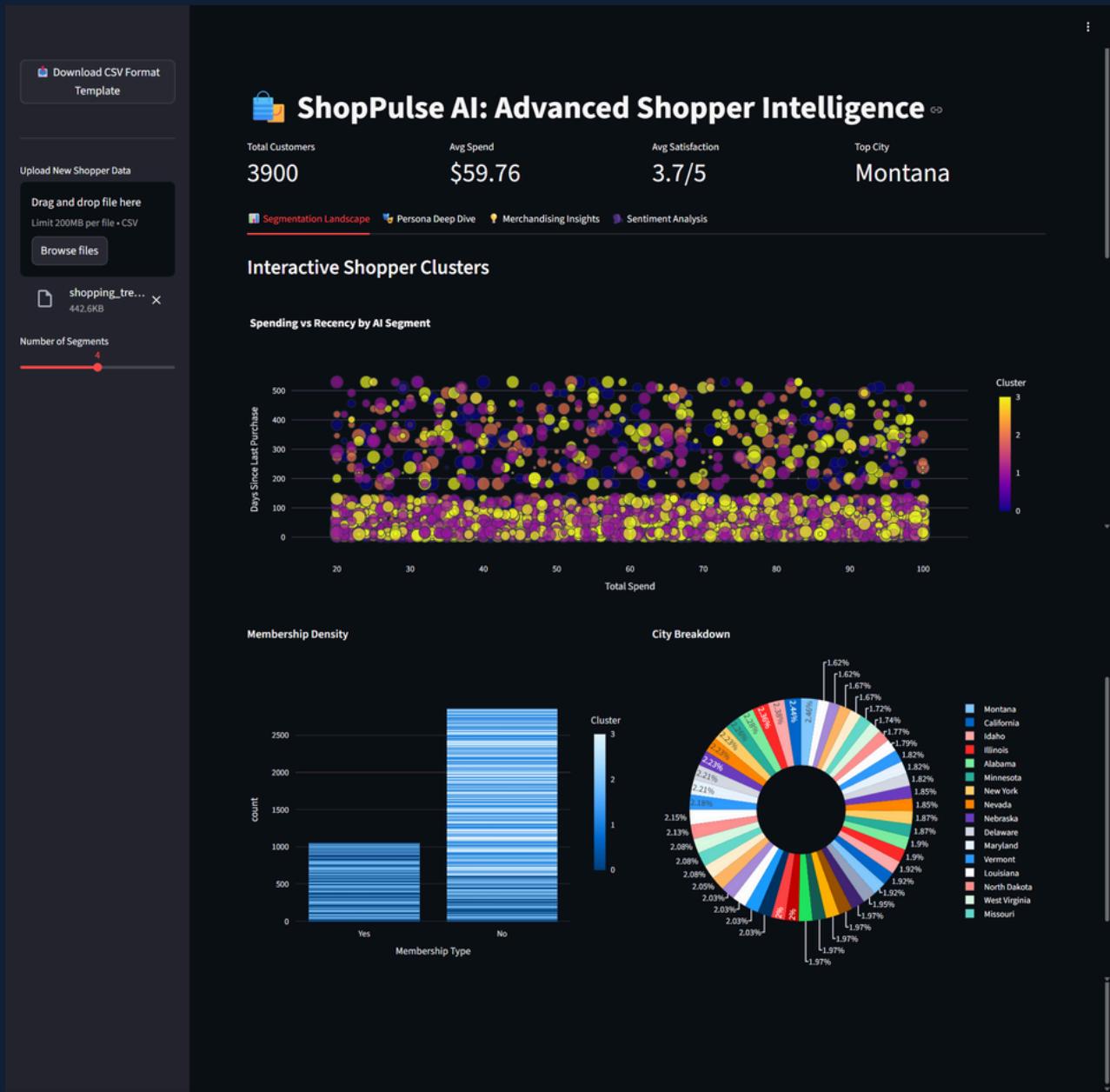


Fig 1: The Command Center. Real-time metrics show active customers and multi-dimensional K-Means Scatter Plot clustering users by Spend vs. Recency.

# 4. Technical Architecture

The system is built on a Streamlit frontend for rapid interactivity, backed by a robust Python data processing pipeline.

## A. The Machine Learning Engine (Unsupervised Learning)

Unlike simple rule-based sorting (e.g., 'Spend > \$100'), we use Scikit-Learn to uncover hidden, multi-dimensional patterns.

- **Feature Engineering:** We extract five critical dimensions: Total Spend (Revenue), Purchase Frequency (Loyalty), Recency (Churn risk), Average Rating (Sentiment), and Discount Sensitivity.
- **Preprocessing:** We apply StandardScaler to normalize metrics. This prevents variables with large ranges (like Spend) from overpowering small ranges (like Rating).
- **Clustering Algorithm:** We implement K-Means Clustering. The system dynamically evaluates the data to form distinct behavioral groups (typically k=4 to 6).

## B. Tech Stack

- **Language:** Python 3.10+
- **Frontend:** Streamlit (chosen for reactive data dashboards and caching)
- **ML Library:** Scikit-Learn (K-Means, StandardScaler, PCA)
- **AI Model:** Google Gemini 2.5 Flash (via google-generativeai SDK)
- **Visualization:** Plotly (Interactive 3D scatter plots) & Matplotlib
- **Deployment:** Containerized on Render (Cloud-agnostic)

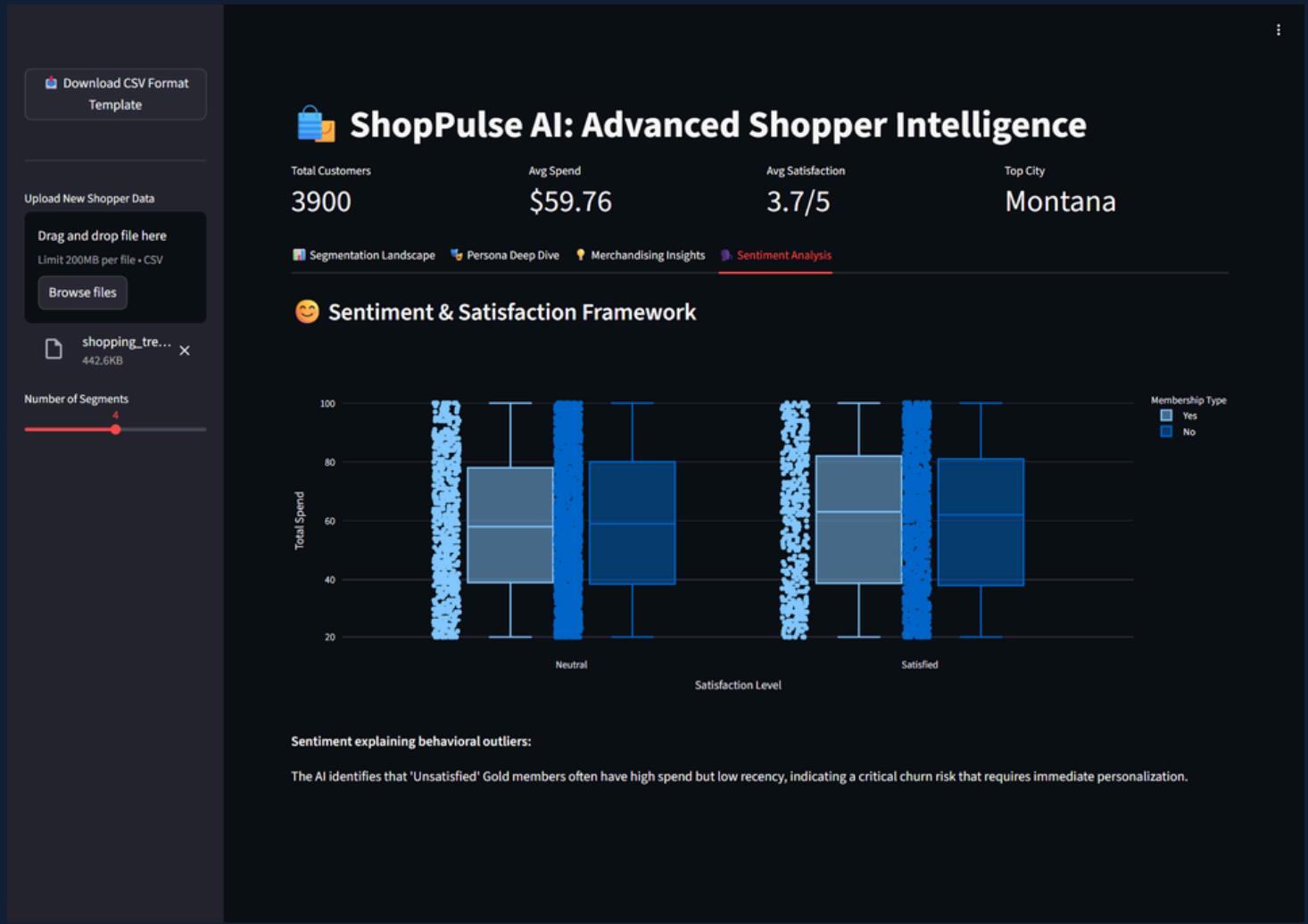


Fig 2: Sentiment Framework. Statistical Box Plot analysis correlating Satisfaction with Spend, detecting churn risks.

## 5. Data Schema & Input

To ensure flexibility, the application accepts standard CSV exports. The algorithm looks for the following behavioral signals:

Column Name	Data Type	Usage
Customer ID	String	Unique Identifier (anonymized)
Purchase Amount	Float	Used to calculate Monetary Value
Purchase Date	Date	Used to calculate Recency & Frequency
Review Rating	Int (1-5)	Used to gauge Satisfaction
Location	String	Used for Geographic distribution maps

# 6. AI Implementation Strategy

The integration of GenerativeAI is the core innovation of ShopPulse AI. We do not simply 'chat' with the data; we use a structured RAG-lite (Retrieval Augmented Generation) approach.

## The Context Injection

We treat the Cluster Centroids (the mathematical average of each group) as the context.  
Input to AI: {'Cluster 0': {'Avg\_Spend': \$950, 'Avg\_Rating': 2.1, 'Recency': 12 days}}  
Interpretation: The AI recognizes this pattern as a 'High Value but Dissatisfied' customer.

## Synthetic 'Voice of Customer'

The AI is also tasked with generating a Synthetic Review for each segment. This helps the merchant 'hear' the tone of that customer group.

The screenshot shows the ShopPulse AI platform's 'Advanced Shopper Intelligence' dashboard. On the left, there's a sidebar with options to 'Download CSV Format Template', 'Upload New Shopper Data' (with a file upload area for CSV files up to 200MB), and a slider for 'Number of Segments' set to 4. The main content area has a title 'ShopPulse AI: Advanced Shopper Intelligence' with a shopping bag icon. It displays summary statistics: Total Customers (3900), Avg Spend (\$59.76), Avg Satisfaction (3.7/5), and Top City (Montana). Below this, the 'AI-Powered Persona Discovery' section shows a dropdown menu for 'Select Segment to Analyze' with option '1'. It includes two cards: 'Segment DNA' (Avg Spend: \$59.09, Avg Age: 56.4, Avg Rating: 3.6/5, Days Since Buy: 78) and 'Simulated Customer Voice' (button to 'Predict Review Voice'). At the bottom is a 'Generate Persona Profile' button.

Fig 3: Persona Deep Dive. The 'Simulated Customer Voice' button calls Gemini 2.5 Flash to generate a synthetic review.

## AI-Generated Personas & Strategy

Gemini 2.5 Flash analyzes the centroid data to create a full psychographic profile (Prudent Pat, Savvy Senior, etc.) moving beyond simple 'Cluster 1' labels.

**ShopPulse AI: Advanced Shopper Intelligence**

Total Customers: 3900 Avg Spend: \$59.76 Avg Satisfaction: 3.7/5 Top City: Montana

Upload New Shopper Data  
Drag and drop file here  
Limit 200MB per file - CSV  
Browse files

Segmentation Landscape Persona Deep Dive Merchandising Insights Sentiment Analysis

**AI-Powered Persona Discovery**  
Select Segment to Analyze: 1

**Segment DNA**  
Avg Spend: \$59.09 Avg Age: 56.4  
Avg Rating: 3.6/5 Days Since Last Purchase: 78  
Days Since Last Purchase: 25.0 (High for the spend, medium rating, low days since last purchase)  
Average Rating: 3.64 (Satisfied, but not overly enthusiastic)  
Days Since Last Purchase: 78.33 (Inrequent, medium rating, high days since last purchase)  
Discount\_Binary: 0 (Does NOT use discounts)  
Age: 56.38 (Older 50s to early 60s)

**Simulated Customer Voice**  
Understand the sentiment behind the numbers by generating a synthetic review.  
Predict Review Voice  
\* As a regular customer from Oregon, I've spent a fair amount on many items full price, finding the products relatively good enough for my everyday needs, which is why I keep coming back every couple of months.\*

**Generate Persona Profile**

**AI Generated Persona**  
Here's a detailed marketing persona based on the provided customer segment data.

**Marketing Persona: Prudent Pat**  
Data Snapshot:  
• Total Spend: \$59.09 (Low)  
• Items Purchased: 25.0 (High for the spend, medium rating, low days since last purchase)  
• Average Rating: 3.64 (Satisfied, but not overly enthusiastic)  
• Days Since Last Purchase: 78.33 (Inrequent, medium rating, high days since last purchase)  
• Discount\_Binary: 0 (Does NOT use discounts)  
• Age: 56.38 (Older 50s to early 60s)

**1. Catchy Name: Prudent Pat**  
Pat is a discerning shopper who values consistency, reliability, and gets great value over fleeting deals. They don't chase the lowest price but are willing to pay a little extra if it's worth it. Pat makes thoughtful decisions and sticks with what works.

**2. Buying Psychology**

- Value First: For Prudent Pat, this is his most defining characteristic. They are not motivated by discounts or fast sales. Instead, they prioritize quality and value over convenience, reliability, and price. They seek out products that offer consistent quality and value, making frequent sales or potentially reducing an inflated original price less attractive.
- Consistency & Loyalty: "Buy once, buy again" is their motto. They are loyal to brands they trust and repeat purchases.
- Practicality Over Freedoms: Pat isn't swayed by the latest fads or marketing hype. They focus on items that provide practical necessity, utility, and problem solving. They may prioritize items that enhance their household items, or goods that enhance their comfort and routine.
- Reliable Supplier: The purchase frequency (quarterly) suggests Pat integrates purchases into a regular routine. They understand the value of consistent, reliable buying. Once Pat finds a product or brand that suits their needs, they stick with it, becoming a loyal, repeat customer.
- Informed but Not Obsessive: Pat's average purchase count is moderate, indicating criticality. They expect products to perform as intended and are willing to pay for quality. While they likely need incentives to purchase, they aren't ever price敏感. Their focus is on getting the best value for their money, ensuring a good, solid product is enough; it doesn't need to be the latest or most expensive.
- Efficiency and Ease: The value is straightforward: getting more bang for their buck. They are open to discounts, but are also willing to pay full price if it means getting a better deal. They appreciate older product offerings that have stood the test of time. Complex loyalty programs or confusing promotional offers are off-putting.
- Risk Aversion: They are conservative with their purchases, seeking quality and reliable brands. Trying new, unproven companies or brands can bring them back to what has consistently performed well.

**3. Lifestyle Profile**

- Life Stage: Pat is likely to be in their mid-50s to early 60s, prioritizing stability and security over personal growth. They are likely financially settled, and they prioritize comfort, stability, and well-being.
- Home & Environment: Pat owns a home and spends considerable time there. They value comfort, functionality, and durability over aesthetic trends. These items provide a sense of security and stability, contributing to a sense of accomplishment or relaxation.
- Interests & Hobbies: Pat probably enjoys reading, puzzles, and other quiet activities like reading, gardening, cooking, light exercise (walking, yoga), or crossword puzzles. These might be interests in local clubs or volunteer work. These hobbies likely provide a sense of accomplishment or relaxation.
- Technology: Comfortable with technology, Pat uses it for practical purposes like online shopping, managing finances, staying connected with family, and staying up-to-date with news or community-related news. They prefer stable, reliable platforms and avoid complex or constantly changing ones.
- Financial Habits: Financially prudent and responsible, Pat likely prioritizes saving over spending. They are very debt-aware and focused on saving for retirement or future security. They consider the long-term value of an item and value and make purchasing decisions that align with their financial goals and priorities.
- Social & Community: While not necessarily a social butterfly, Pat maintains a close-knit circle of friends and family. They prioritize genuine connections and are likely to be involved in their community, particularly for their service and related causes.
- Media Consumption: Pat likely consumes news and information through print media. They might enjoy specialized magazines related to their hobbies or interests, or possibly those that offer practical advice or insightful discussions rather than entertainment. They are also likely to be interested in connecting with family or specific interest groups, not for documenting trends.

**Marketing Approach for Prudent Pat:**

- Focus on Value & Reliability: Highlight product benefits, durability, longevity, and consistency. Emphasize how the product is built to last ("Proven Performance").
- Clean, Direct Communication: Avoid fancy language, hyperbole, or complex promotions. Be transparent about the product's value, what it does, and why it's a good choice at its original price.
- Reliable Brand Endorsements: Provide detailed product descriptions, clear care instructions, and honest reviews. Use helpful tips or guides related to the product's function.
- Emphasize Consistent Quality: Emphasize any quality certifications, warranties, or the vendor's reputation for reliability.
- Streamlined Shopping Experience: Ensure the website or in-store experience is easy to navigate, user-friendly, and an efficient checkout process.
- Loyalty Programs (if offered): Reward them for their loyalty and provide incentives on every purchase towards future essentials.
- Targeted Content: Create content around topics that align with their interests, health and wellness tips, and general guidance for products that resonate with their lifestyle.

Fig 4: AI-Generated Persona 'Prudent Pat'. Note the Buying Psychology and Lifestyle Profile sections.

**ShopPulse AI: Advanced Shopper Intelligence**

Total Customers: 3900 Avg Spend: \$59.76 Avg Satisfaction: 3.7/5 Top City: Montana

Upload New Shopper Data  
Drag and drop file here  
Limit 200MB per file - CSV  
Browse files

Segmentation Landscape Persona Deep Dive Merchandising Insights Sentiment Analysis

**Strategic Recommendations**  
Generate Strategy Guide for All Segments

As Merchandising Director, let's break down these customer segments to optimize our strategy. The data reveals distinct behavioral patterns influenced by age, discount sensitivity, and purchase habits.

**Cluster Analysis & Merchandising Recommendations**

**Cluster 0: The "Savvy Senior Savers"**

- Profile: Older demographic (57.76 years), high spend and items purchased, good ratings. Crucially, they always purchase with a discount (Discount\_Binary = 1).
- Key Insight: This is a logic, engaged, and high-value segment, but highly driven by promotions and value. They are accustomed to and likely expect discounts.
- Product Category Recommendations:
  - Everyday Essentials & Practical Goods: Items that offer consistent utility and value (e.g., household staples, personal care multi-packs, comfortable apparel).
  - Established Brands: Brands they recognize and trust, especially if offered at a promotional price.
  - "Comfort & Convenience" Items: Products that make life easier or more enjoyable, often seen as a good investment when discounted (e.g., ergonomic home equipment, travel tech, subscription boxes for convenience).
  - Bulk & Bundled Buys: Offer savings when purchasing larger quantities or complementary items together.
- Inventory Advice:
  - Consistent Stock on Promoted Items: Ensure a steady supply of products regularly featured in promotions or those that can frequently be discounted. They expect these items to be available at attractive prices.
  - Sales Cycle Alignment: Forecast inventory based on promotion schedules. Overstock on high-demand items during known discount periods.
  - Clearance & Value Focus: Maintain a healthy flow of items that can be moved through sales or clearance events, as this aligns with their purchasing behavior.
- Price Elasticity Insights:
  - Highly Price Elastic: This segment is very sensitive to price changes and strongly responsive to discounts. They will likely defer purchases of full-price items, waiting for a sale.
  - Discount Expectation: Discontinuing discounts for this group could lead to a significant drop in conversion and loyalty. Discounts are a primary driver of their purchases.
  - Promotional Value: Focus messaging on the value of the discount – "Save 10%," "10% off" "Limited time offer."

**Cluster 1: The "Loyal Lifestyler"**

- Profile: Older demographic (56.38 years), very similar high spend and items purchased to Cluster 0, but they never purchase with a discount (Discount\_Binary = 0).
- Key Insight: This is a premium-seeking, loyal, and convenient segment. They are our most consistent and recent purchasers, willing to pay full price for what they want.
- Product Category Recommendations:
  - Premium & Quality Goods: Higher-end versions of common products, items known for durability, craftsmanship, or superior performance.
  - Bundles & Loyalty Offers: Offer unique items, special editions, or products from brands known for quality and status.
  - Convenience & Quality Savings Solutions: Products that offer a seamless experience or save them effort (e.g., high-quality meal kits, sophisticated smart home devices, curated collections).
  - Personalized & Curated Items: Recommendations based on their past purchases, suggesting new additions to their preferred categories.
- Inventory Advice:
  - Reliable Stock of Core Premium SKUs: Ensure consistent availability of their preferred full-price, high-quality items. Out-of-stock situations will be frustrating for this segment.
  - New Arrivals & Innovation: Introduce new, exciting products regularly, as they are likely to be early adopters of premium offerings without needing a price incentive.
  - Focus on Depth over Breadth (in core categories): Have a strong selection within their preferred premium categories rather than a vast array of discounted items.
- Price Elasticity Insights:
  - Low Price Elasticity: This segment is less sensitive to price. They prioritize quality, brand, convenience, or specific features over cost savings.
  - Avoid Unnecessary Discounting: Deep discounting for this group might devalue the product or brand in their eyes. Focus on showcasing value through quality, benefits, and service.
  - Value-Added Services: Consider offering premium shipping, extended warranties, or exclusive customer service perks as incentives, rather than price reductions.

**Cluster 2: The "Young Bargain Hunters"**

- Profile: Younger demographic (33.06 years), good spend and items purchased. Similar to Cluster 0, they always purchase with a discount (Discount\_Binary = 1).
- Key Insight: A younger, value-conscious segment actively seeking deals and promotions. They are engaged but highly price-sensitive.
- Product Category Recommendations:
  - Trendy & Seasonal Items: Fashion-forward apparel, popular tech gadgets, viral beauty products, home decor accents that are in vogue.
  - Entry-Level Premium "Deals": Affordable alternatives to higher-end brands, or "gateway" products into a new category at an accessible price point.
  - Bundles & Kits: DIY kits, starter packs, or themed bundles that offer a complete solution at a lower perceived cost.
  - "Experience-Oriented" Discounts: Discounts on tickets, classes, or items that enable new experiences.
- Inventory Advice:
  - Agile & Fast Turnover: Maintain lean inventory on fast-moving trends as they are prepared to quickly cycle out items that fall out of favor.
  - Flash Sale Readiness: Design inventory strategies around frequent, short-duration sales and promotions to capture impulse buys.
  - Opportunistic Buys: Be open to purchasing inventory that can be aggressively discounted to attract this segment, even if margins are tight.
- Price Elasticity Insights:
  - Very High Price Elastic: This segment is extremely sensitive to price. They are likely to compare prices across different retailers and will switch for a better deal.
  - Promotion Driven: Promotions, discounts, and clearance events are essential to drive conversions. They are constantly looking for the "best deal."
  - Urgency & Scarcity: Messaging that emphasizes limited-time offers or depleting stock can be particularly effective.

**Cluster 3: The "Discerning Achievers"**

- Profile: Younger demographic (33.06 years), highest total spend (\$12.4) despite the lowest number of items purchased (23.65). They have the highest average rating (3.8) but also the best recent purchases (17.13 days since Last Purchase). They never purchase with a discount (Discount\_Binary = 0).
- Key Insight: A younger, value-conscious segment actively seeking deals and promotions. They are engaged but highly price-sensitive.
- Product Category Recommendations:
  - High-Quality, Durability, & Investment Pieces: Products with a long lifespan, superior materials, and excellent craftsmanship (e.g., designer accessories, artisanal goods, high-end apparel).
  - Niche & Exclusivity Items: Limited editions, or items from specialized brands not widely available.
  - Experience-Oriented & Aspirational Luxury: Products or services that offer a unique experience, elevate their lifestyle, or align with their personal values (e.g., high-end travel gear, bespoke services, sustainable luxury brands).
  - Storytelling & Brand Values: Products with a compelling story behind them, reflecting ethical sourcing, unique design, or artisanal heritage.
- Inventory Advice:
  - Curated & Limited Stock: Focus on quality over quantity. Carefully selected inventory of high-value items, potentially with pre-order options for popular exclusive products.
  - Strategic Replenishment: Forecast demand for specific high-ASP items, ensuring availability without holding excessive, slow-moving stock.
  - New Collection Launches: Plan inventory around impactful new collection drops or seasonal luxury offerings to re-engage this infrequent but high-value item category.
- Price Elasticity Insights:
  - Lowest Price Elasticity: This segment is largely price-insensitive for the right product. Discounting could potentially devalue the product in their eyes.
  - Value Proposition is Key: Focus marketing on intrinsic product value, brand reputation, exclusivity, sustainability, and superior benefits.
  - Re-engagement through Novelty & Exclusivity: To drive repeat purchases, focus on new arrivals, exclusive access to upcoming collections, or personalized styling advice, rather than price promotions.

**Overall Merchandising Strategy Implications:**

- Segmented Promotions Calendar: Clearly differentiate promotional calendar for discount-sensitive (Clusters 0 & 2) vs. non-discount-sensitive (Clusters 1 & 3) segments. Align market promotions that could erode perceived value for premium buyers.
- Targeted Communications: Tailor marketing messages to each segment. For example, highlight "savings" for Clusters 0 & 2, and "quality," "exclusivity," or "convenience" for Clusters 1 & 3.
- Age-Specific Product Assortment: Ensure product offerings align with the lifestyle and preferences of older vs. younger demographics, especially concerning trend, tech adoption, and comfort needs.
- Inventory Management: Implement dynamic forecasting to manage the balance between high-turnover, discounted items and curated, high-value full-price products. This will minimize waste and maximize margin opportunities across segments.
- Lifetime Value Focus: While Cluster 0 buys less frequently, their high spend per purchase makes them incredibly valuable. Develop specific retention strategies that don't involve discounting to nurture these relationships. Similarly, Cluster 1's reengagement and full-price purchases indicate high loyalty worth protecting.

By implementing these tailored strategies, we can maximize revenue and customer satisfaction across all our identified segments.

Fig 5: Strategic Recommendations. Detailed inventory and pricing advice derived from the cluster DNA.

## 7. Business Feasibility & Market

**Target Market:** Primary audience is independent Shopify/ WooCommerce merchants (estimated 4M+ globally).

**Revenue Model:** Freemium (Basic Charts) + SaaS Tier (\$29/mo) for Gemini 2.5 Flash insights.

**Competitive Advantage:** ShopPulse is prescriptive (tells you what to do) vs traditional descriptive tools. Inference costs are minimized (<\$0.01/report) using the efficient Flash model.

## 8. Ethical Considerations

**Privacy First Design :** We do not send Personally Identifiable Information (names, emails) to the AI. Only aggregated statistical averages are processed.

**Bias Mitigation:** We explicitly instruct the AI to avoid assigning gender, race, or cultural stereotypes to the generated personas unless explicitly supported by the geographic data.

## 9. Future Roadmap

- **Q3 2026:** Shopify App Store Integration (Direct API data sync).
- **Q4 2026:** Predictive Churn Modeling (Forecast who leaves next month).
- **2027:** Autonomous Marketing Agents (AI drafts & sends emails via Mailchimp).