

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)

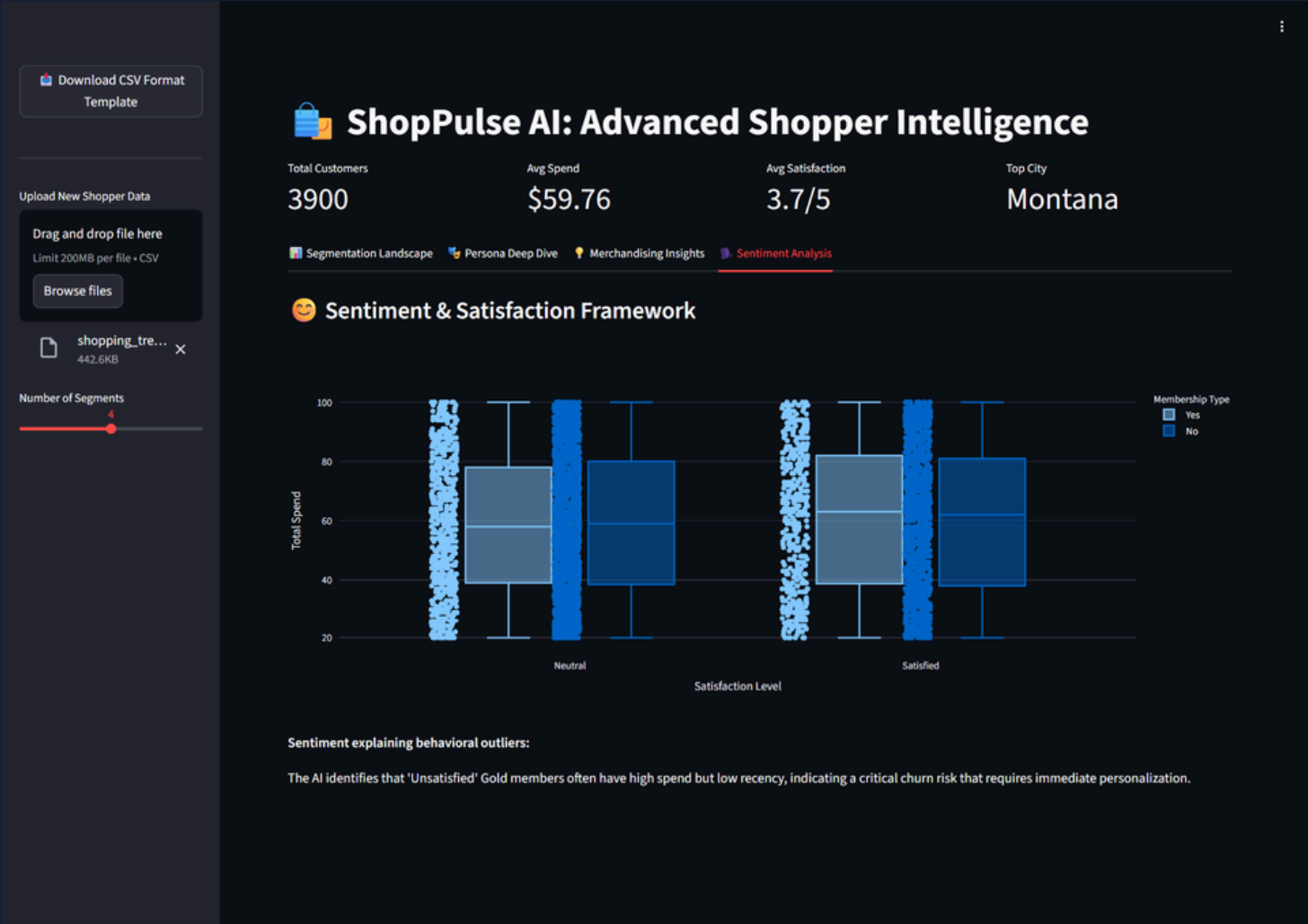


Fig 2: Sentiment Framework. Statistical Box Plot analysis correlating Satisfaction with Spend, detecting churn risks.

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.

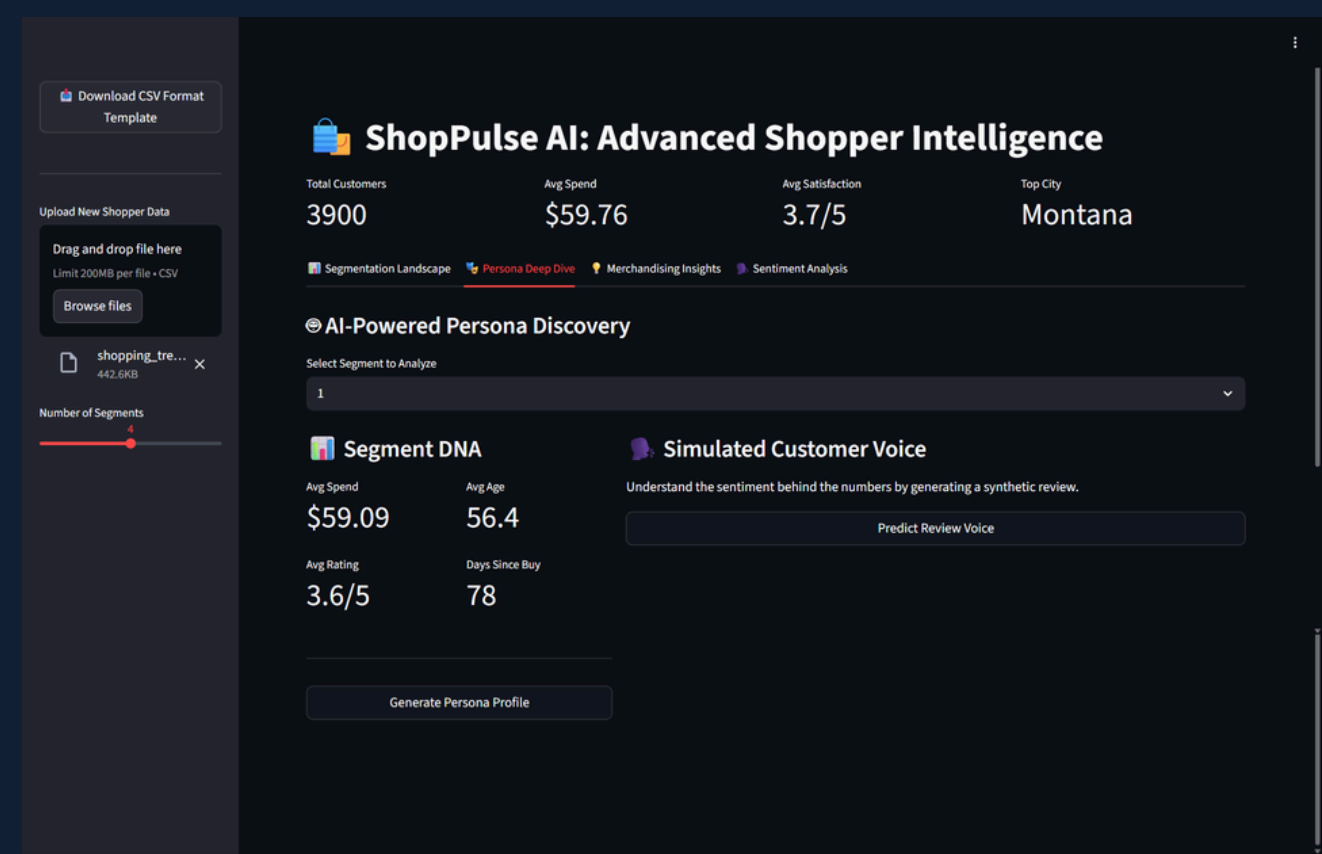


Fig 3: Persona Deep Dive. The 'Simulated Customer Voice' button calls Gemini 2.5 Flash to generate a synthetic review.

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.

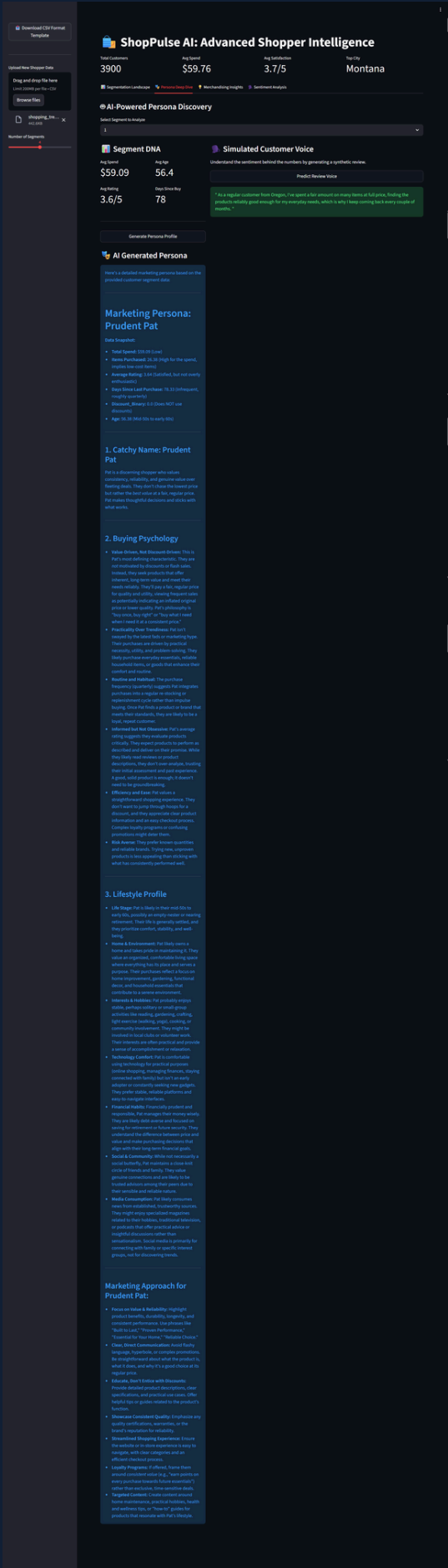


Fig 4: AI-Generated Persona 'Prudent Pat'. Note the Buying Psychology and Lifestyle Profile sections.

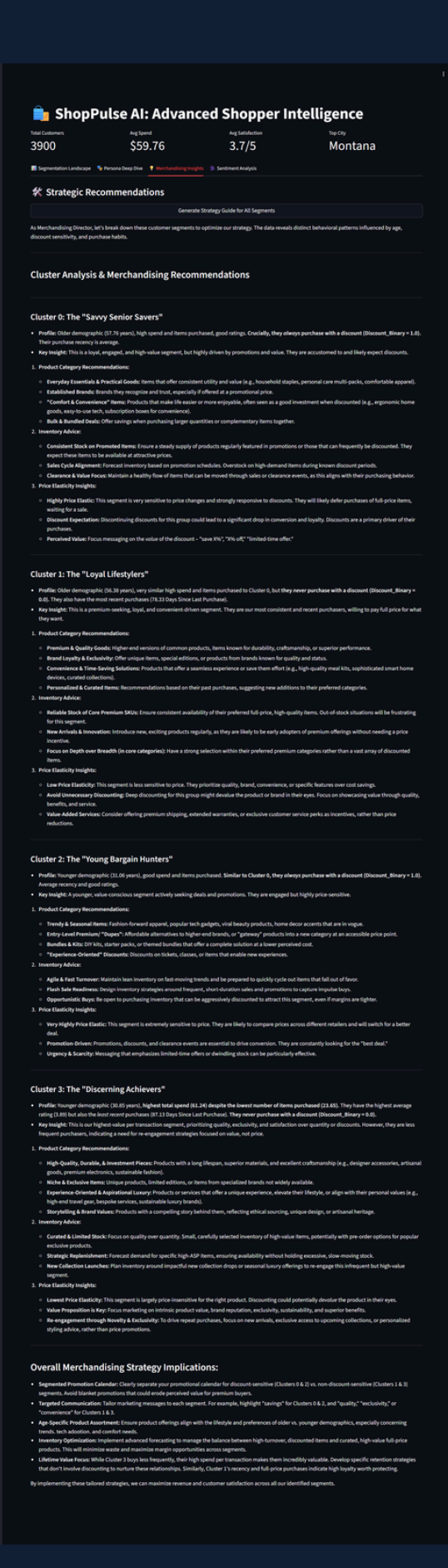


Fig 5: Strategic Recommendations. Detailed inventory and pricing advice derived from the cluster DNA.

7. Business Feasibility & Market

TargetMarket: 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).