Customer Shopping Behavior Analysis

A comprehensive data-driven business case study analyzing 3,900 customer purchases across multiple product categories. This analysis uncovers critical insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

*By: Pavan Meti | **Tools:** Python, SQL, Power BI | **Date:** October 2025*

# Project Overview & Dataset

Dataset Scope

This project analyzes transactional data from 3,900 purchases across various product categories. The dataset contains 18 columns capturing comprehensive customer and purchase information.

- *Total Rows: 3,900 transactions*
- *Total Columns: 18 features*
- *Missing Data: 37 values in Review Rating*

Key Data Features

The dataset encompasses customer demographics, purchase details, and shopping behavior metrics.

- *Demographics: Age, Gender, Location, Subscription Status*
- *Purchase Details: Item, Category, Amount, Season, Size, Color*
- *Behavior: Discounts, Promo Codes, Previous Purchases, Frequency, Ratings, Shipping Type*
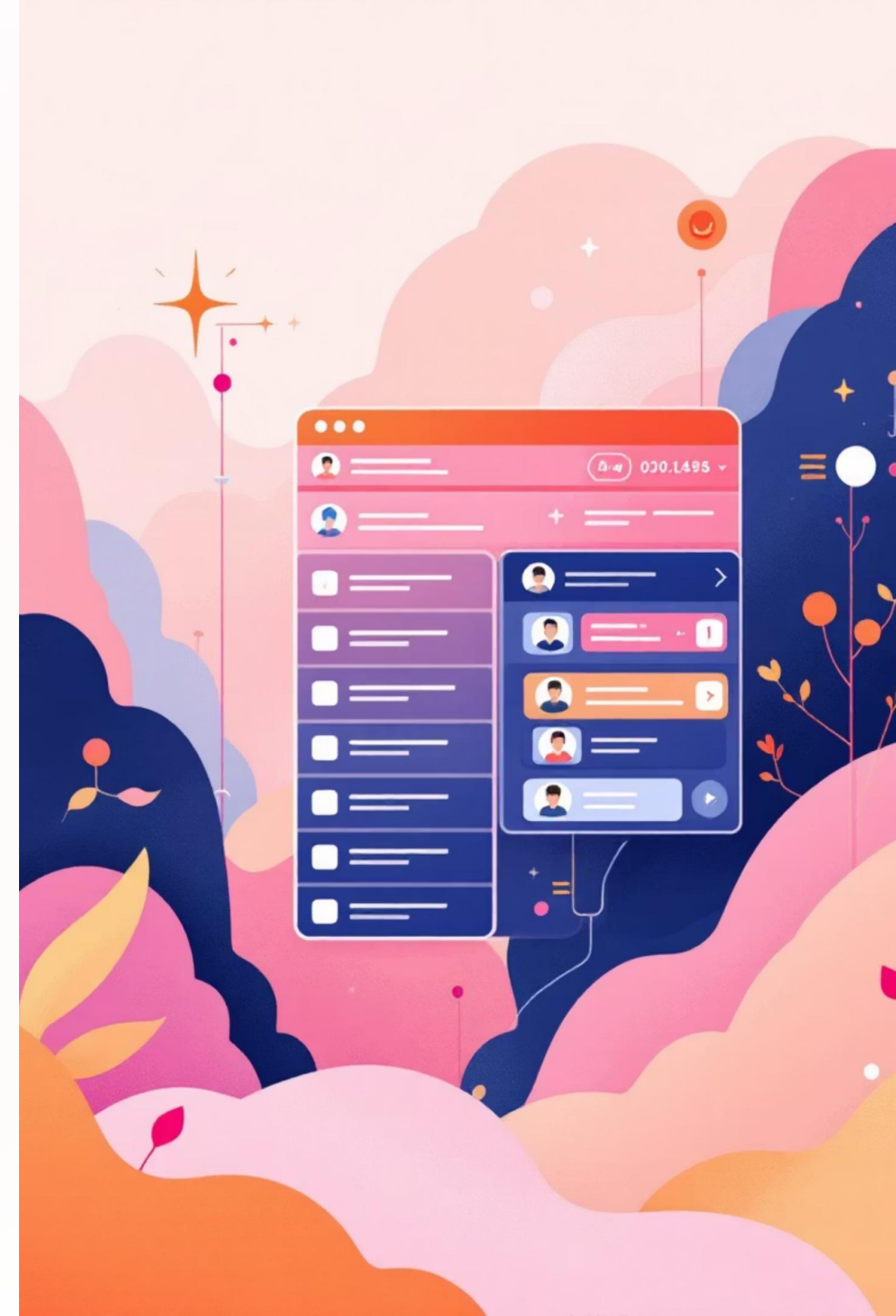
Data Preparation & Cleaning

We began with comprehensive data preparation and cleaning in Python to ensure data quality and consistency:

01

Data Loading & Exploration

Imported dataset using pandas and checked structure with df.info() and summary statistics using .describe().

02

Missing Data Handling

Identified 37 null values in Review Rating column and imputed using median rating by product category.

03

Column Standardization

Renamed all columns to snake case for improved readability and consistent documentation.

04

Feature Engineering

Created age_group column through binning and purchase_frequency_days from transaction dates.

05

Data Consistency & Integration

Verified redundancy between discount_applied and promo_code_used, dropped latter. Connected to PostgreSQL for SQL analysis.

Revenue Insights by Demographics

SQL analysis revealed critical revenue patterns across customer demographics and subscription status:

Revenue by Gender

Compared total revenue generated by male versus female customers to identify gender-based spending patterns.

Age Group Performance

Calculated total revenue contribution of each age group to identify highest-value customer segments.

Subscription Impact

Compared average spend and total revenue between subscribers and non-subscribers to quantify subscription value.
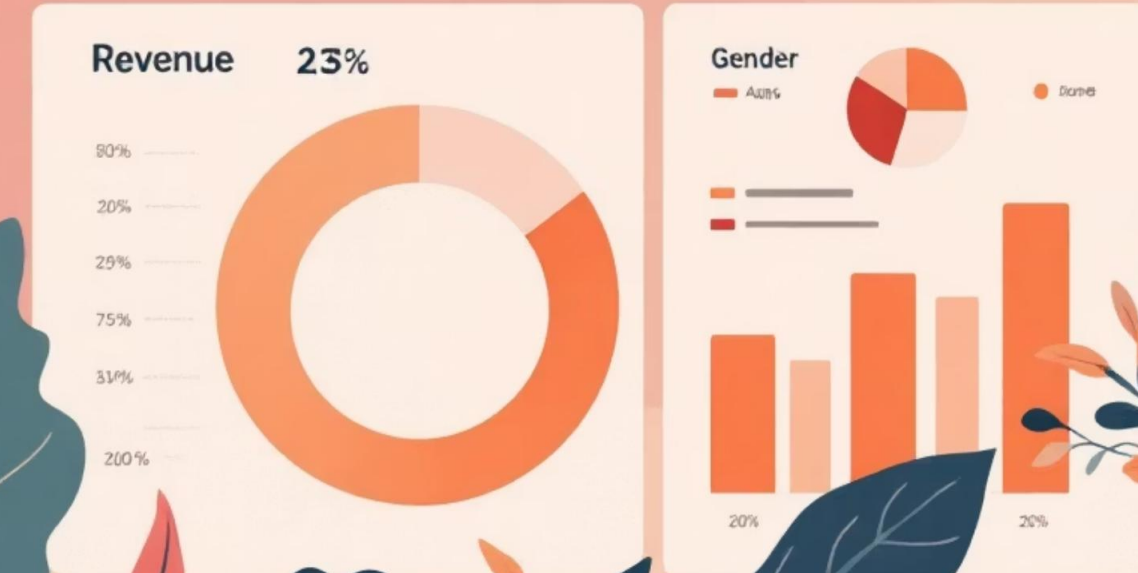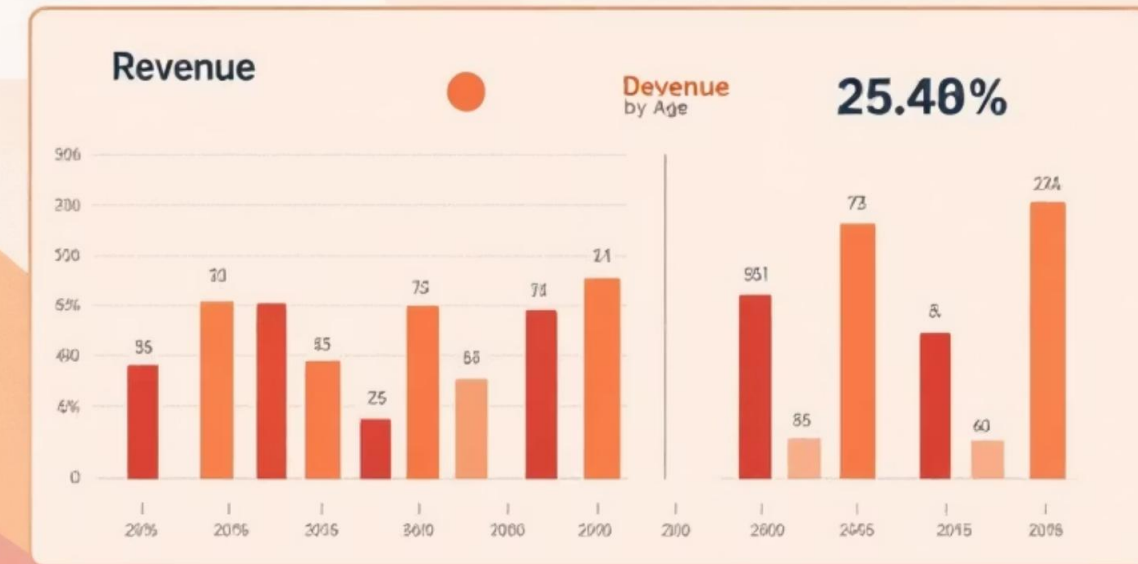
Customer Segmentation & Behavior

*Advanced SQL queries segmented customers into distinct behavioral groups and identified key purchasing patterns:*

Customer Segments

Classified customers into three segments based on purchase history:

- **New:** *First-time buyers*
- **Returning:** *Multiple purchases*
- **Loyal:** *Frequent repeat buyers*

Key Findings

Customers with more than 5 purchases show significantly higher subscription rates, indicating strong correlation between repeat purchases and subscription adoption.
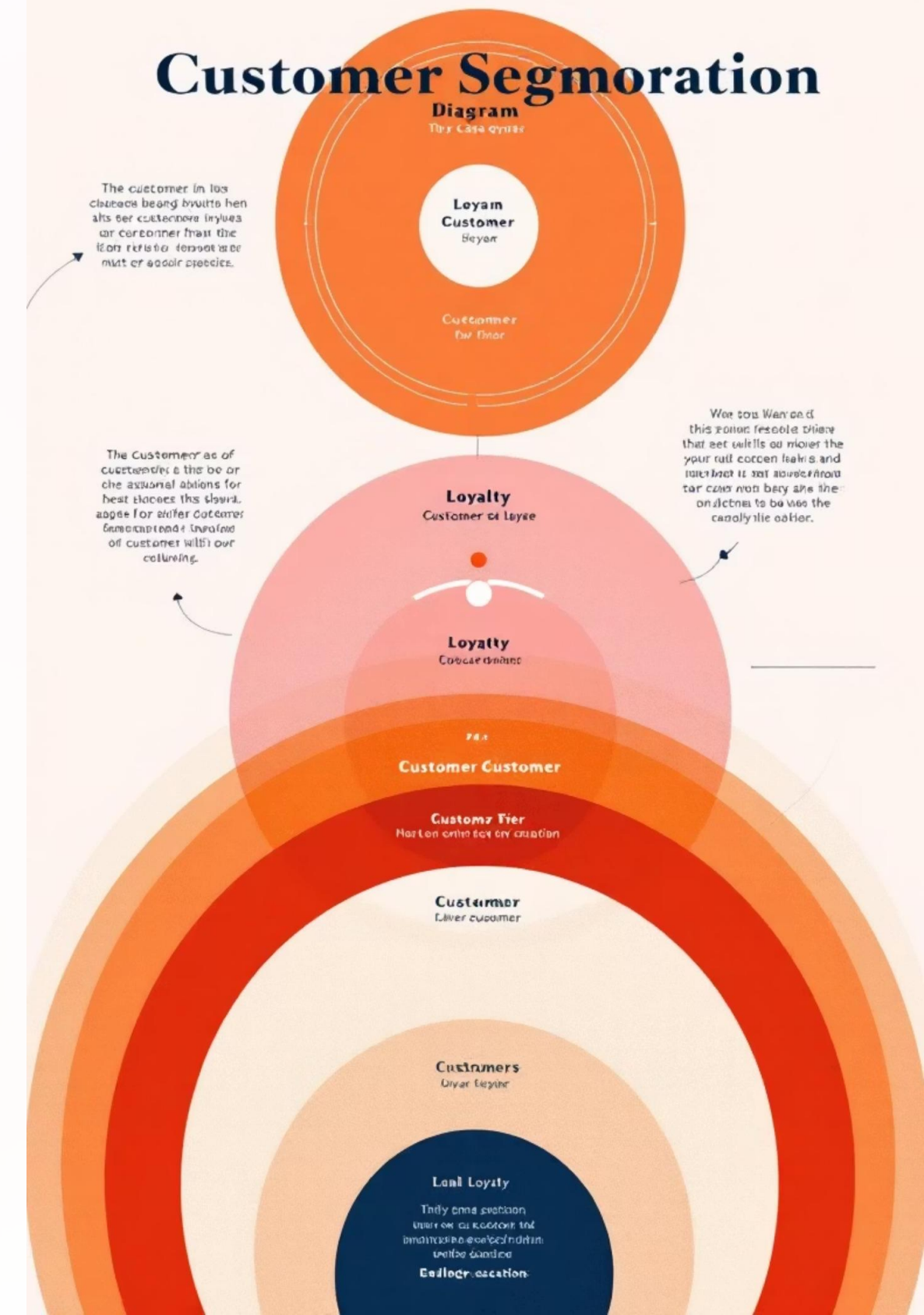
Product Performance Analysis

Comprehensive product analysis identified top performers and discount-dependent items across all categories:

1 **Top 5 Products by Rating**

Identified products with highest average review ratings, indicating customer satisfaction leaders.

2 **Top 3 Products per Category**

Listed most purchased products within each category to understand category-specific preferences.

3 **Discount-Dependent Products**

Found 5 products with highest percentage of discounted purchases, revealing margin pressure points.

4 **High-Spending Discount Users**

Identified customers who used discounts but still spent above average, showing discount elasticity.

Shipping & Purchase Behavior

Analysis of shipping preferences and purchase patterns revealed important operational insights:

Shipping Type Comparison

Compared average purchase amounts between Standard and Express shipping options. Express shipping users demonstrate higher average transaction values, suggesting premium customer segment.

Discount Strategy Impact

Examined relationship between discount application and customer behavior. Balance between sales boosts and margin control is critical for sustainable growth.

Power BI Dashboard

An interactive dashboard was built in Power BI to present all insights visually and enable real-time business monitoring. The dashboard consolidates key metrics, customer segments, product performance, and revenue trends into a comprehensive business intelligence platform for stakeholder decision-making.

Strategic Business Recommendations

Based on comprehensive data analysis, we recommend the following strategic initiatives:

Boost Subscriptions

Promote exclusive benefits and premium features for subscribers to increase adoption rates and customer lifetime value.

Loyalty Programs

Reward repeat buyers with tiered benefits to move customers from Returning to Loyal segments and increase retention.

Review Discount Policy

Balance promotional discounts with margin control by identifying discount-dependent products and adjusting strategy accordingly.

Product Positioning

Highlight top-rated and best-selling products prominently in marketing campaigns and store displays.

Targeted Marketing

Focus marketing efforts on high-revenue age groups and express-shipping users who demonstrate premium purchasing behavior.

Key Takeaways & Next Steps

This data-driven analysis provides actionable insights to optimize customer experience and drive business growth. The identified customer segments, product performance metrics, and behavioral patterns form the foundation for strategic decision-making.

Immediate Actions

Launch subscription promotion campaign targeting high-value customers and implement loyalty program for repeat buyers.

Medium-Term Goals

Refine discount strategy based on product-level analysis and optimize marketing spend toward high-revenue demographics.

Continuous Improvement

Monitor dashboard metrics regularly, update customer segments quarterly, and iterate on recommendations based on performance data.