Customer Shopping Behavior Analysis

A comprehensive analysis of 3,900 customer transactions revealing critical insights into spending patterns, customer segments, product preferences, and subscription behavior. This data-driven study combines Python analytics, SQL business intelligence, and Power BI visualization to guide strategic business decisions and unlock revenue growth opportunities.

Dataset Overview

Scale & Structure

Our analysis encompasses **3,900 rows** and **18 columns** of rich transactional data, providing a comprehensive view of customer behavior across multiple dimensions.

Key Data Categories

- **Customer Demographics:** Age, Gender, Location, Subscription Status

- **Purchase Details:** Item Purchased, Category, Purchase Amount, Season, Size, Color

- **Shopping Behavior:** Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type

Data Quality Note: Only 37 missing values detected in the Review Rating column, representing less than 1% of the dataset—ensuring high data integrity for our analysis.

Python Data Preparation & Engineering

Our analytical journey began with comprehensive data preparation using Python, ensuring data quality and creating meaningful features for deeper insights.

01

Data Loading & Exploration

Imported dataset using pandas, performed initial exploration with df.info() and .describe() to understand structure and summary statistics.

02

Missing Data Handling

Identified null values and imputed missing Review Rating values using median rating of each product category for accurate representation.

03

Column Standardization

Renamed all columns to snake_case format for improved readability, consistency, and better documentation practices.

04

Feature Engineering

Created age_group column by binning customer ages and purchase_frequency_days column from purchase data for segmentation analysis.

05

Data Consistency Check

Verified redundancy between discount_applied and promo_code_used fields; dropped promo_code_used to eliminate duplication.

06

Database Integration

Connected Python script to PostgreSQL and loaded cleaned DataFrame into database for advanced SQL analysis and querying.

Key Statistical Insights

3,900
Total Transactions
Complete purchase records analyzed across all customer segments

$59.76
Average Purchase
Mean transaction value across all categories and customer types

44
Median Age
Central customer age, ranging from 18 to 70 years old

3.75
Average Rating
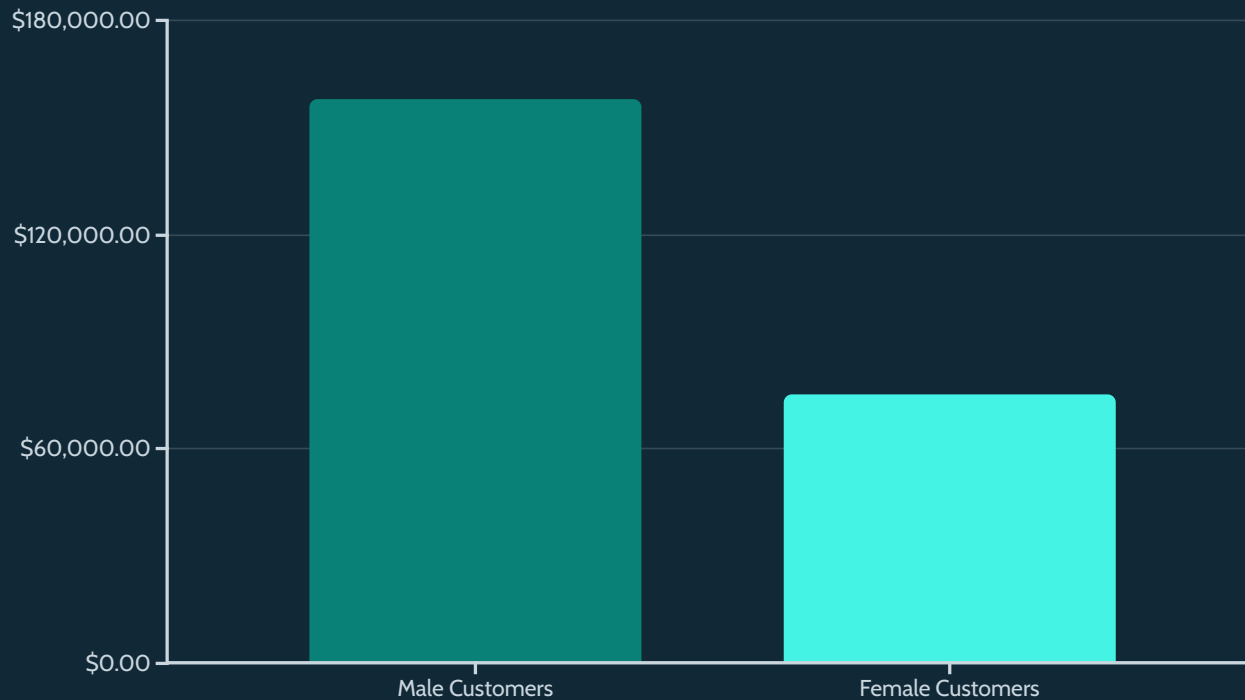Mean customer satisfaction score out of 5.0 stars

25
Previous Purchases
Median number of prior transactions per customer

50
Unique Locations
Geographic diversity across customer base

Made with GAMMA

Revenue Analysis: Gender & Spending Patterns

Male customers generate **$157,890** in total revenue compared to **$75,191** from female customers—representing more than double the revenue contribution. This significant disparity suggests opportunities for targeted marketing campaigns to increase female customer engagement and spending.

High-Value Discount Users

We identified **839 customers** who used discounts but still spent above the average purchase amount of $59.76. These savvy shoppers represent a valuable segment that responds to promotions while maintaining high transaction values.

Product Performance & Customer Preferences

⭐ Top-Rated Products

Gloves lead with 3.86 average rating, followed by **Sandals** (3.84), **Boots** (3.82), **Hat** (3.80), and **Skirt** (3.78). These high-performing items should be featured prominently in marketing campaigns.

🛵 Shipping Preferences

Express shipping customers spend an average of **$60.48** compared to **$58.46** for Standard shipping—a modest but meaningful difference indicating willingness to pay premium for faster delivery.

🏷️ Discount-Dependent Items

Hat (50% discount rate), **Sneakers** (49.66%), **Coat** (49.07%), **Sweater** (48.17%), and **Pants** (47.37%) show highest discount dependency—critical for pricing strategy optimization.

Category Leaders: Most Purchased Products

Accessories

1. Jewelry (171 orders)
2. Sunglasses (161)
3. Belt (161)

Clothing

1. Blouse (171 orders)
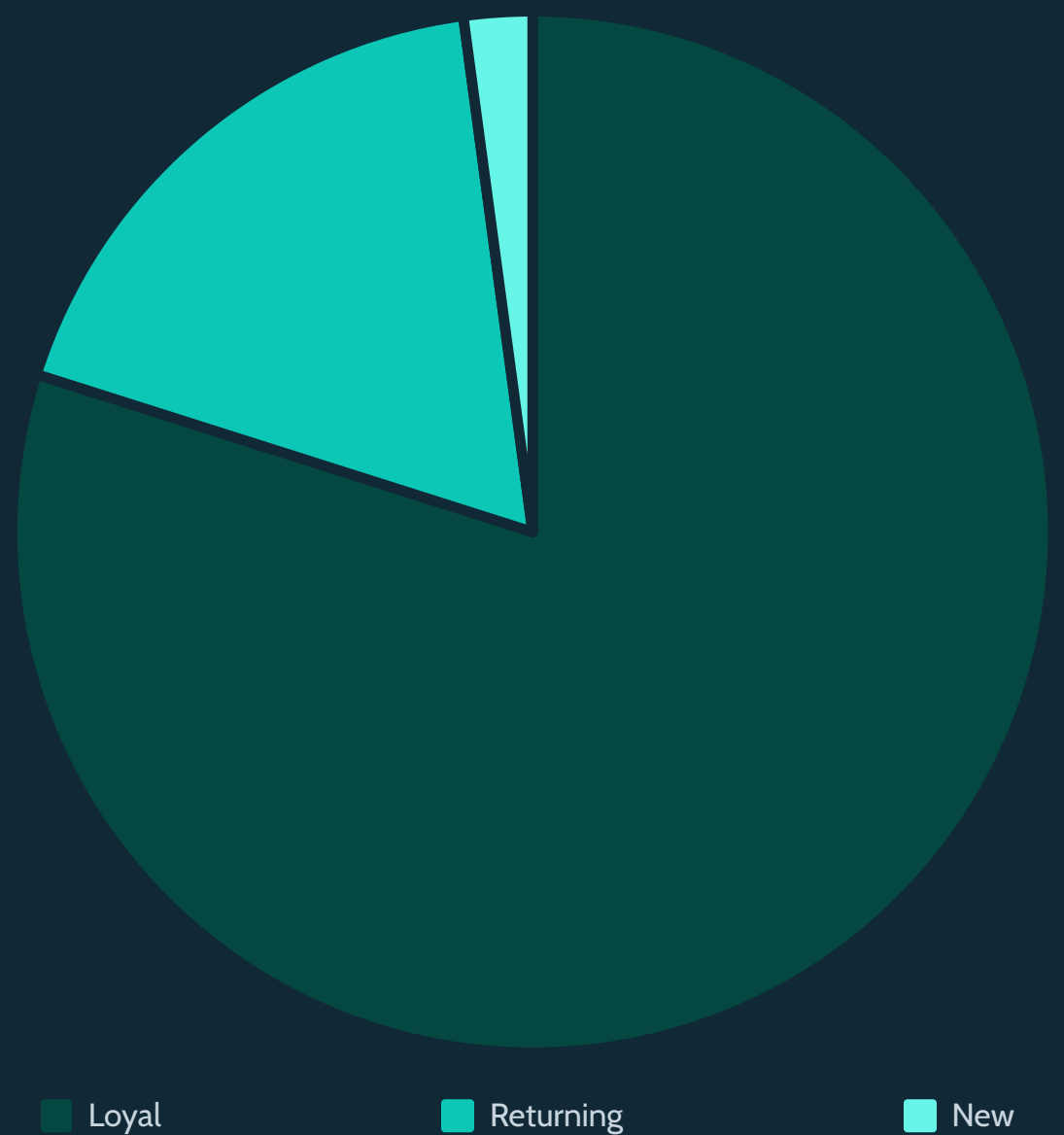2. Pants (171)
3. Shirt (169)

Footwear

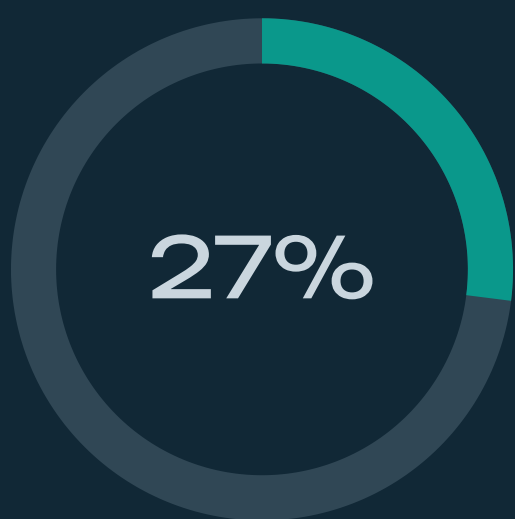1. Sandals (160 orders)
2. Shoes (150)
3. Sneakers (145)

Outerwear

1. Jacket (163 orders)
2. Coat (161)

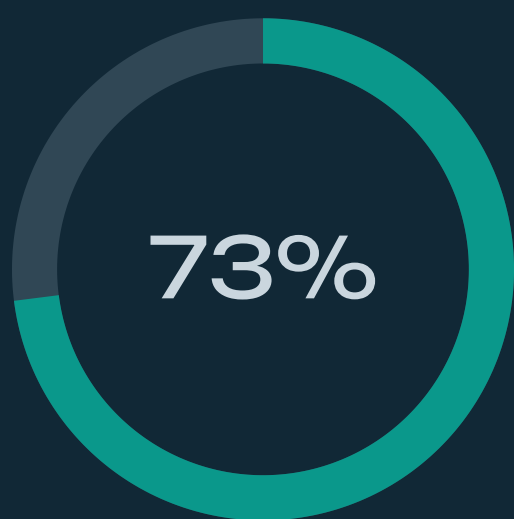# Customer Segmentation & Loyalty Insights

■ Loyal　　　■ Returning　　　■ New

Our customer base is dominated by **Loyal customers (79.9%)** with multiple repeat purchases, followed by **Returning customers (18.0%)** and **New customers (2.1%)**. This distribution indicates strong customer retention but highlights an opportunity to convert more first-time buyers into returning customers.

27%
Subscription Rate
1,053 subscribers out of 3,900 total customers

73%

(Non-Subscribers ring)
Non-Subscribers
2,847 customers without active subscriptions

31%
Repeat Buyer Subscription
958 of 3,476 repeat buyers (>5 purchases) have subscriptions
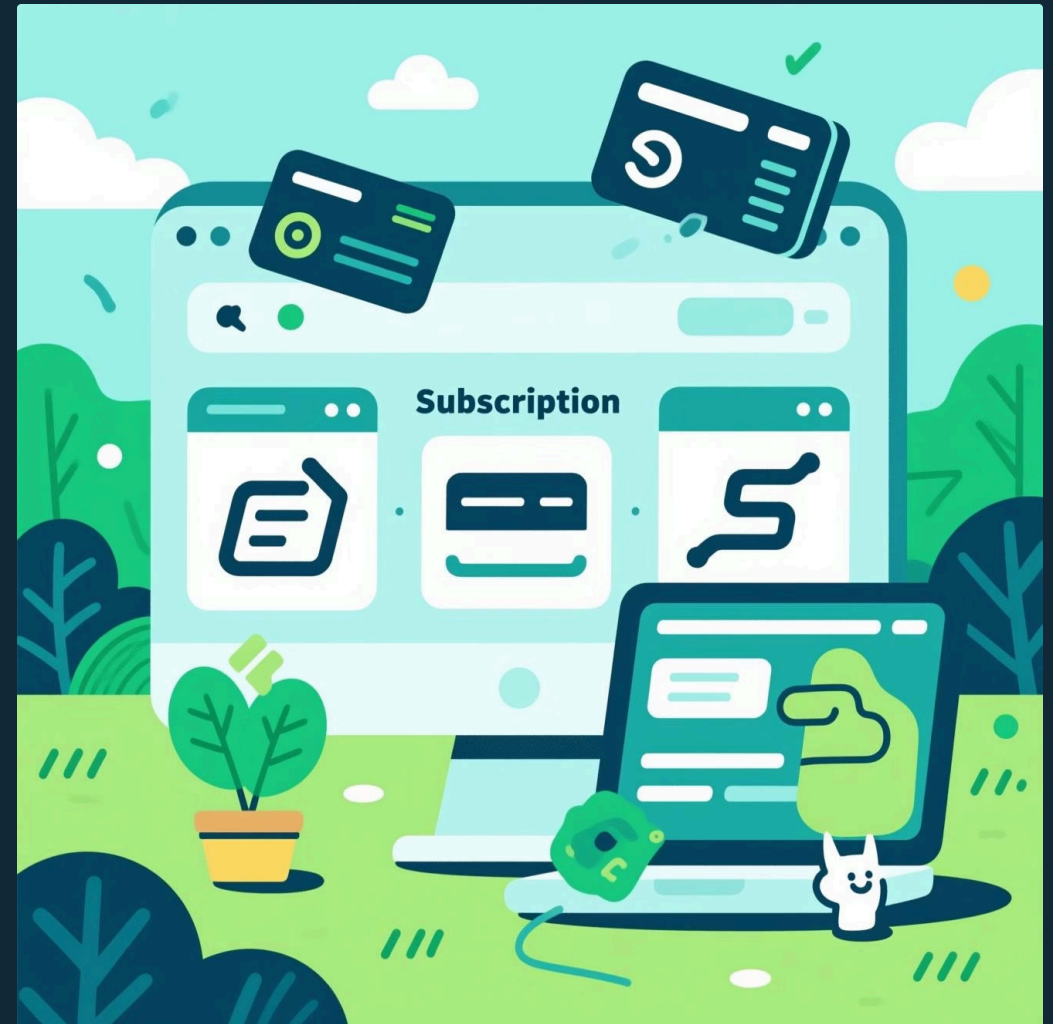
Made with **GAMMA**

Subscription Economics Analysis

Revenue Comparison

Despite having fewer customers, the subscription model shows interesting dynamics:

- **Subscribers:** 1,053 customers generating **$62,645** total revenue with **$59.49** average spend

- **Non-Subscribers:** 2,847 customers generating $170,436 total revenue with **$59.87** average spend

The minimal difference in average spend (**$0.38**) suggests subscription status doesn't significantly impact per-transaction value, but the sheer volume of non-subscribers drives higher total revenue.

Strategic Opportunity: With 73% of customers not subscribed and only 31% of repeat buyers having subscriptions, there's massive potential to grow subscription revenue through targeted conversion campaigns.

Age Group Revenue Distribution

Young Adults lead revenue generation at **$62,143**, followed closely by Middle-aged customers at **$59,197**. The relatively balanced distribution across age groups (ranging from $55,763 to $62,143) indicates broad market appeal, though Young Adults represent the highest-value demographic for targeted marketing investments.

Strategic Business Recommendations

1

Boost Subscription Adoption

With only 27% subscription rate and minimal spending difference between subscribers and non-subscribers, promote exclusive benefits, early access to products, and loyalty rewards to convert the 73% non-subscribed customer base—particularly targeting the 2,518 repeat buyers without subscriptions.

2

Implement Customer Loyalty Programs

Design tiered rewards to move customers from New (2.1%) to Returning (18%) and ultimately to Loyal (79.9%) segments. Focus on the 701 Returning customers with incentives that encourage repeat purchases and long-term engagement.

3

Optimize Discount Strategy

Review discount policy for high-dependency items (Hat, Sneakers, Coat at ~50% discount rates) to balance sales volume with margin protection. Consider strategic discounting for the 839 high-value discount users who maintain above-average spending despite promotions.

4

Enhance Product Positioning

Feature top-rated products (Gloves 3.86★, Sandals 3.84★, Boots 3.82★) and category leaders (Jewelry, Blouse, Sandals, Jacket) prominently in marketing campaigns and website placement to capitalize on proven customer preferences.

5

Execute Targeted Marketing

Focus marketing investments on Young Adults (highest revenue at $62,143), male customers (generating $157,890 vs. $75,191 from females), and Express shipping users (spending $60.48 vs. $58.46 for Standard) to maximize ROI on advertising spend.