Customer Shopping Behavior Analysis

Uncovering strategic insights from 3,900 customer transactions to drive smarter business decisions in retail and e-commerce

Project Overview

Our Mission

This comprehensive analysis examines customer shopping behavior using transactional data from 3,900 purchases across various product categories.

By combining Python data science, SQL business intelligence, and Power BI visualization, we uncover actionable insights into spending patterns, customer segments, and product preferences.

Strategic Focus Areas

- Customer spending patterns and demographics
- Product category performance and preferences
- Subscription behavior and loyalty metrics
- Discount effectiveness and margin impact
- Revenue optimization opportunities

Dataset Foundation

Scale

3,900 rows of transaction data across 18 feature columns

Demographics

Age, gender, location, and subscription status tracking

Transactions

Purchase details including amount, category, season, size, and color

Behavior

Discounts, reviews, shipping preferences, and purchase frequency

Data quality note: 37 missing values identified in the Review Rating column, addressed through category-specific median imputation.

Data Preparation Process

01

Data Loading & Exploration

Imported dataset using pandas, examined structure with df.info() and generated summary statistics to understand distributions

02

Missing Data Handling

Identified null values and imputed missing Review Ratings using median values by product category for accuracy

03

Standardization

Renamed columns to snake_case format for consistency and improved code readability across analysis

04

Feature Engineering

Created age_group bins and purchase_frequency_days columns to enable deeper segmentation analysis

05
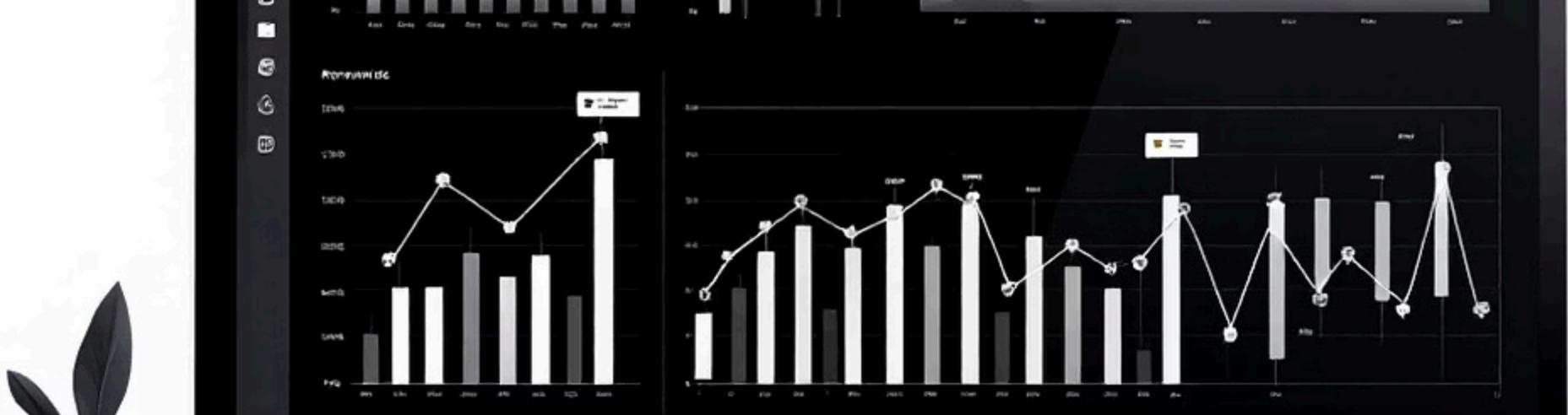
Data Consistency

Verified redundancy between discount_applied and promo_code_used fields, dropped duplicate column

06

Database Integration

Connected Python to PostgreSQL and loaded cleaned DataFrame for SQL-based business intelligence queries

Revenue Analysis: Key Findings

$233K
Male Customers

Total revenue generated

$206K
Female Customers

Total revenue generated

13%
Gender Gap

Revenue difference favoring male segment

Gender-based revenue analysis reveals male customers generate slightly higher total revenue, suggesting opportunities for targeted marketing campaigns to balance engagement across segments.

Shipping & Subscription Insights

Shipping Type Impact

Express shipping customers show a **$59.78 average purchase** compared to **$59.68 for Standard shipping**—nearly identical spending patterns across shipping preferences.

This minimal difference suggests shipping speed is a service preference rather than a spending indicator, allowing flexible fulfillment strategies.
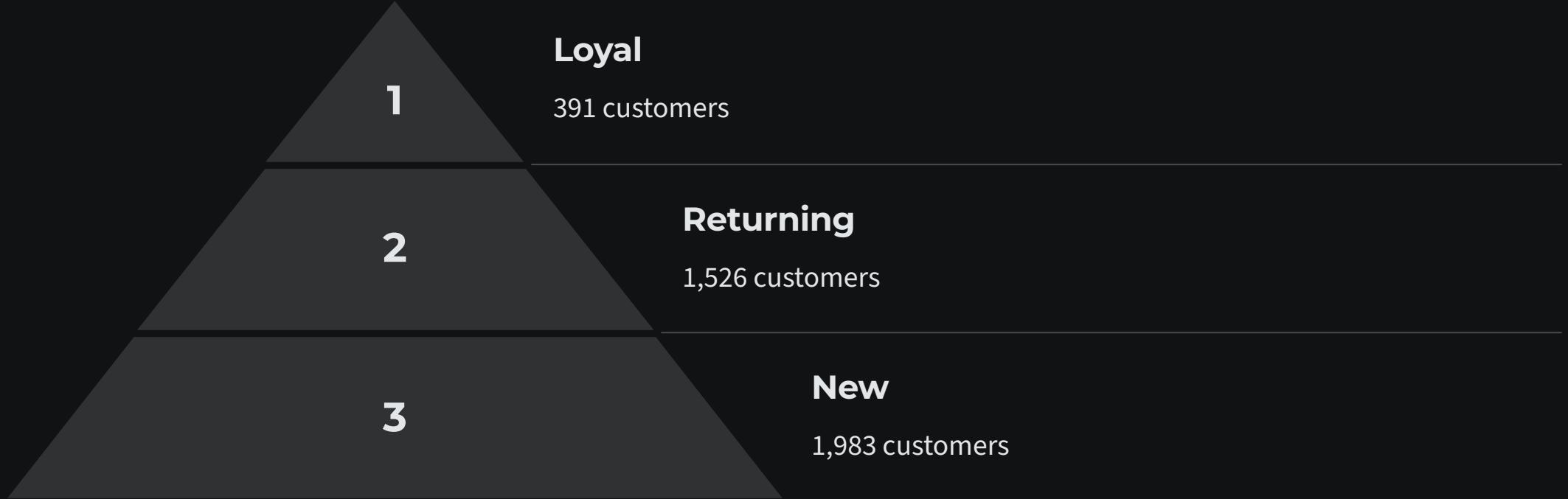
Subscription Value

Subscribers demonstrate significantly higher value with **$61.31 average spend** versus **$58.08 for non-subscribers**.

Total revenue follows the same pattern: subscribers contribute **$120K** while non-subscribers generate **$319K**, though the latter represents a larger customer base.

Customer Segmentation & Loyalty

1 — **Loyal**
391 customers

2 — **Returning**
1,526 customers

3 — **New**
1,983 customers

Customer classification based on purchase history reveals a healthy acquisition funnel, with approximately half being new customers. The challenge: converting the large returning segment (39%) into loyal advocates (10%).

Repeat buyers with 5+ purchases show 68% subscription rate—loyalty drives subscriptions, not the reverse.

Product Performance Matrix

⭐ Top-Rated Products

Jewelry (4.2 stars) leads ratings, followed by Handbag, Dress, Blouse, and Sweater—all averaging 4.0+ stars. Premium categories dominate quality perception.

🏷️ Discount-Dependent Items

Sunglasses tops the list with 60% of purchases using discounts. Dress, Sweater, Backpack, and Hoodie also show 58-60% discount dependency rates.
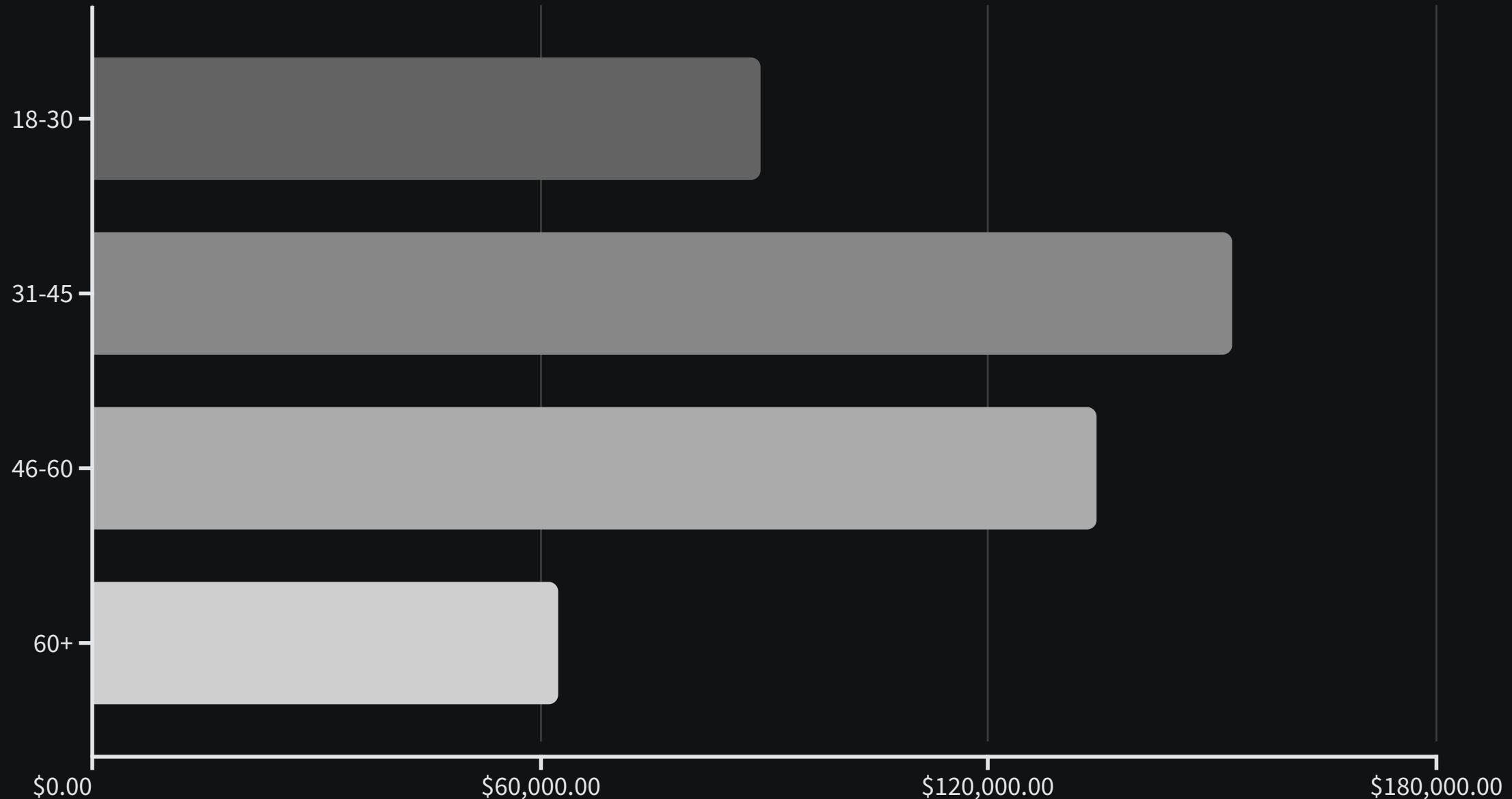
Category Leaders

Blouse dominates Clothing; Dress leads Footwear category; Jewelry excels in Accessories; and Backpack tops the Outerwear segment by purchase volume.

Age Group Revenue Distribution

Age Group	Revenue
18-30	
31-45	
46-60	
60+	

$0.00 $60,000.00 $120,000.00 $180,000.00

The 31-45 age group drives the highest revenue at $153K, representing 35% of total sales. This prime demographic combines purchasing power with digital engagement, making them ideal targets for premium product campaigns and loyalty programs.

Strategic Recommendations

Boost Subscription Adoption

Promote exclusive benefits and rewards for subscribers. Higher average spend ($61 vs $58) justifies investment in subscription incentives and marketing.

Enhance Loyalty Programs

Focus on converting the 1,526 returning customers into loyal advocates. Implement tiered rewards that activate after 3-5 purchases.

Optimize Discount Strategy

Review discount policies for items like Sunglasses (60% discount rate). Balance promotional effectiveness with margin protection through strategic targeting.

Targeted Marketing

Concentrate marketing efforts on high-revenue 31-45 age group and leverage top-rated products (Jewelry, Handbags) in campaigns to maximize ROI.