



Customer Shopping Behavior Analysis

Analysis of 3,900 purchases to uncover spending patterns, segments, product preferences, and subscription behavior to inform strategic decisions.

Dataset at a Glance



Scope

Rows: 3,900 · Columns: 18

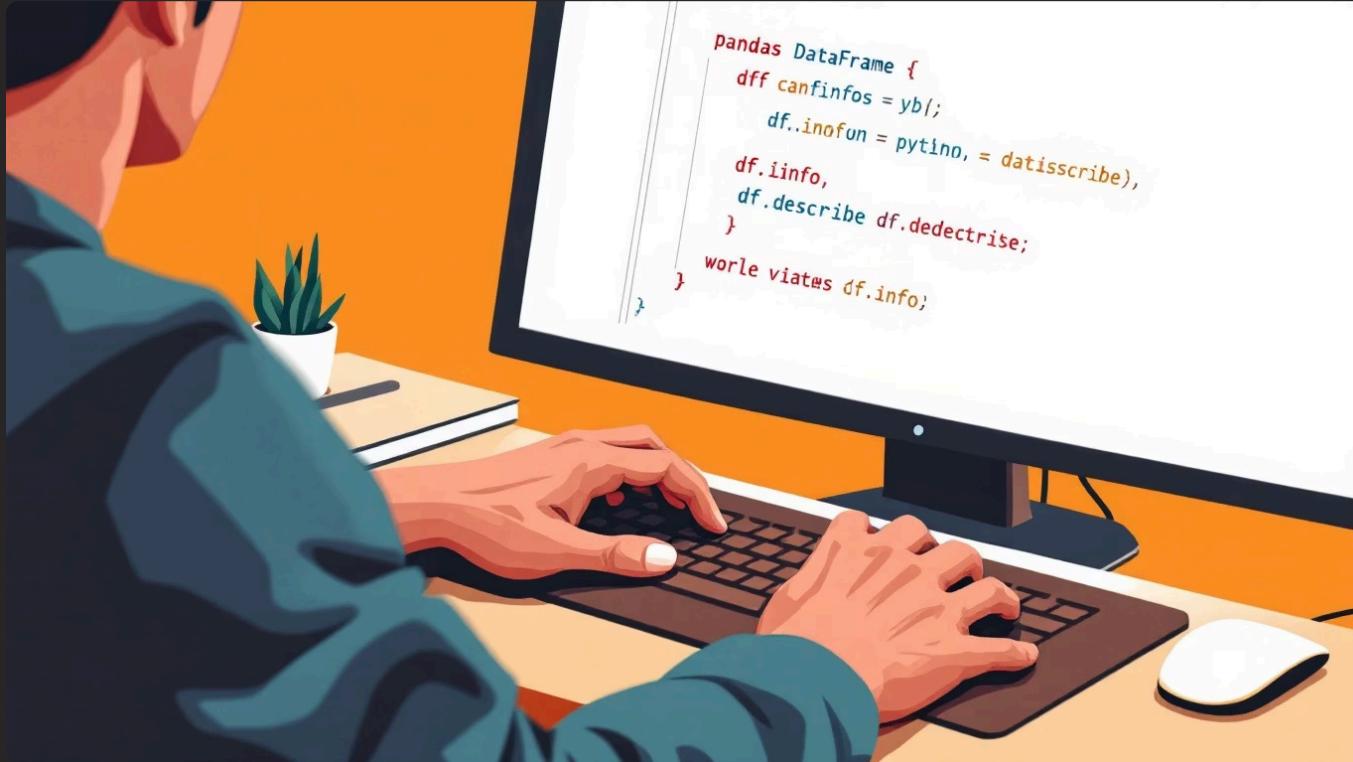
Key Features

Demographics, purchase details, behavior, subscription, shipping, ratings

Missing Data

37 nulls in Review Rating column

Python — Preparation & Cleaning



- **Data Loading**

Imported dataset with pandas;
explored structure and summary stats.

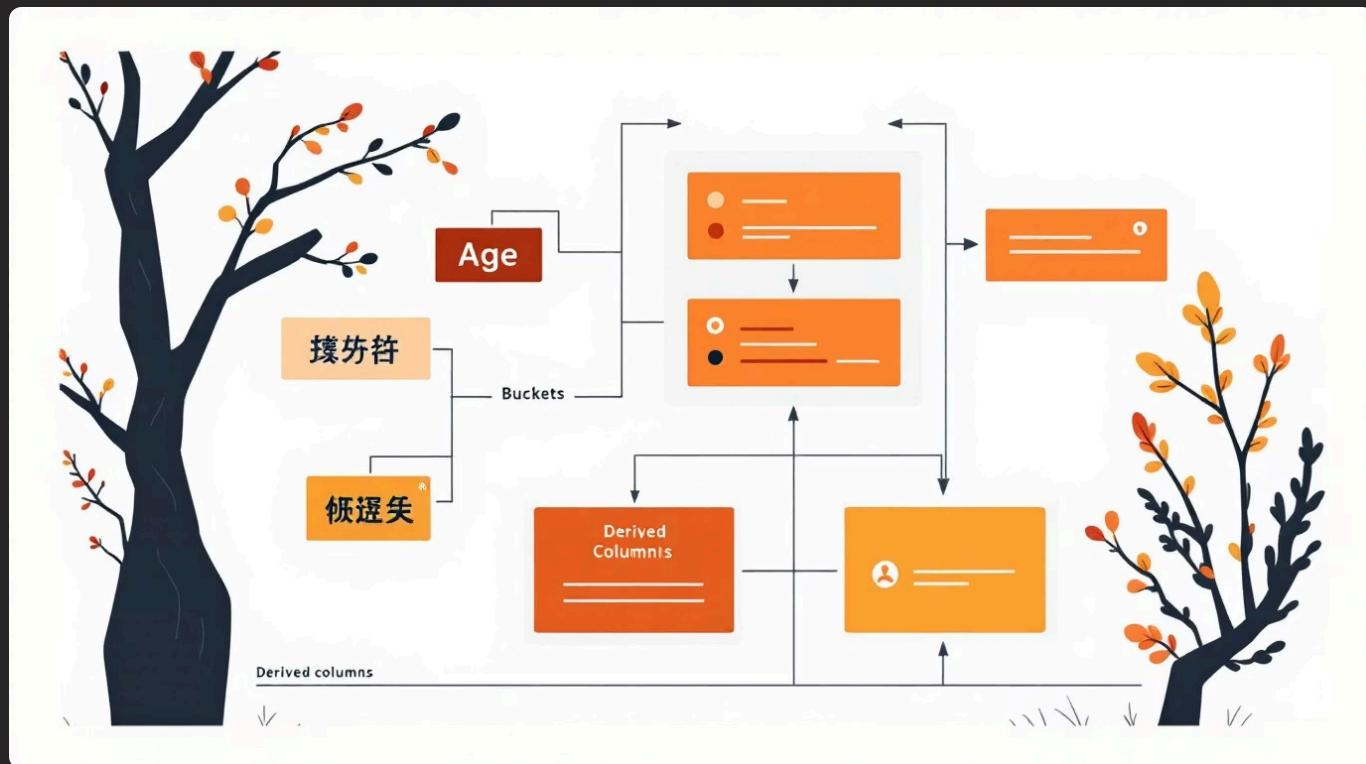
- **Missing Values**

Imputed Review Rating using median
per product category.

- **Standardization**

Renamed columns to snake_case for
readability.

Feature Engineering



Age Groups

Binned ages into age_group (Young Adult, Middle Aged, Adult, Senior).

Purchase Frequency

Derived purchase_frequency_days from transaction timestamps.

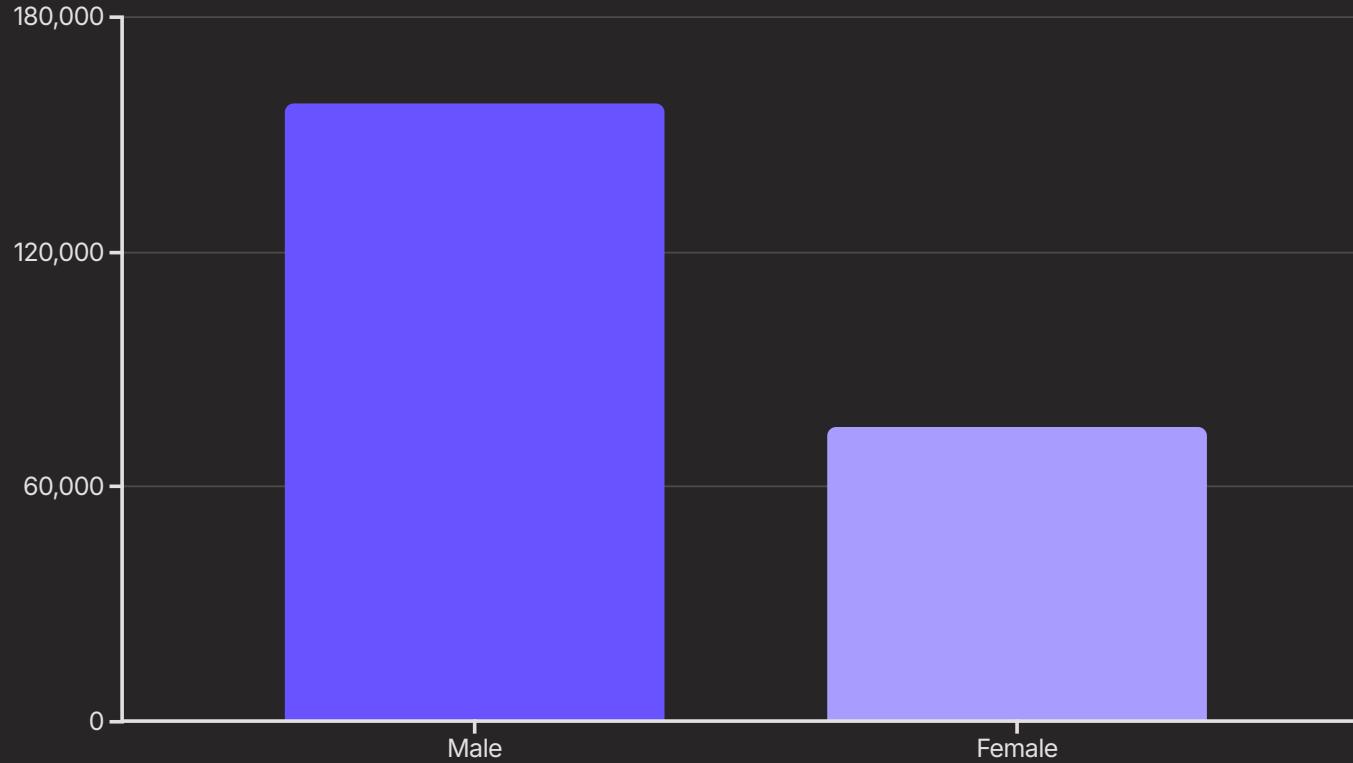
Redundancy Check

Dropped promo_code_used after confirming redundancy with discount_applied.

DB Integration

Loaded cleaned dataframe into PostgreSQL for SQL analysis.

Revenue by Gender (SQL Results)



The dataset shows higher total revenue from male customers (157,890) versus female customers (75,191). Use this insight to tailor campaigns and product mix.

High-Value Discount Users



Discount + Above-Average Spend

Several customers used discounts yet spent above average (examples: customers with purchases 60–97 USD).



Opportunity

Design targeted promotions that preserve margins while retaining high spenders.

Top Products & Discount Dependency



Top Rated

Gloves (3.86), Sandals (3.84), Boots (3.82), Hat (3.80), Skirt (3.78)

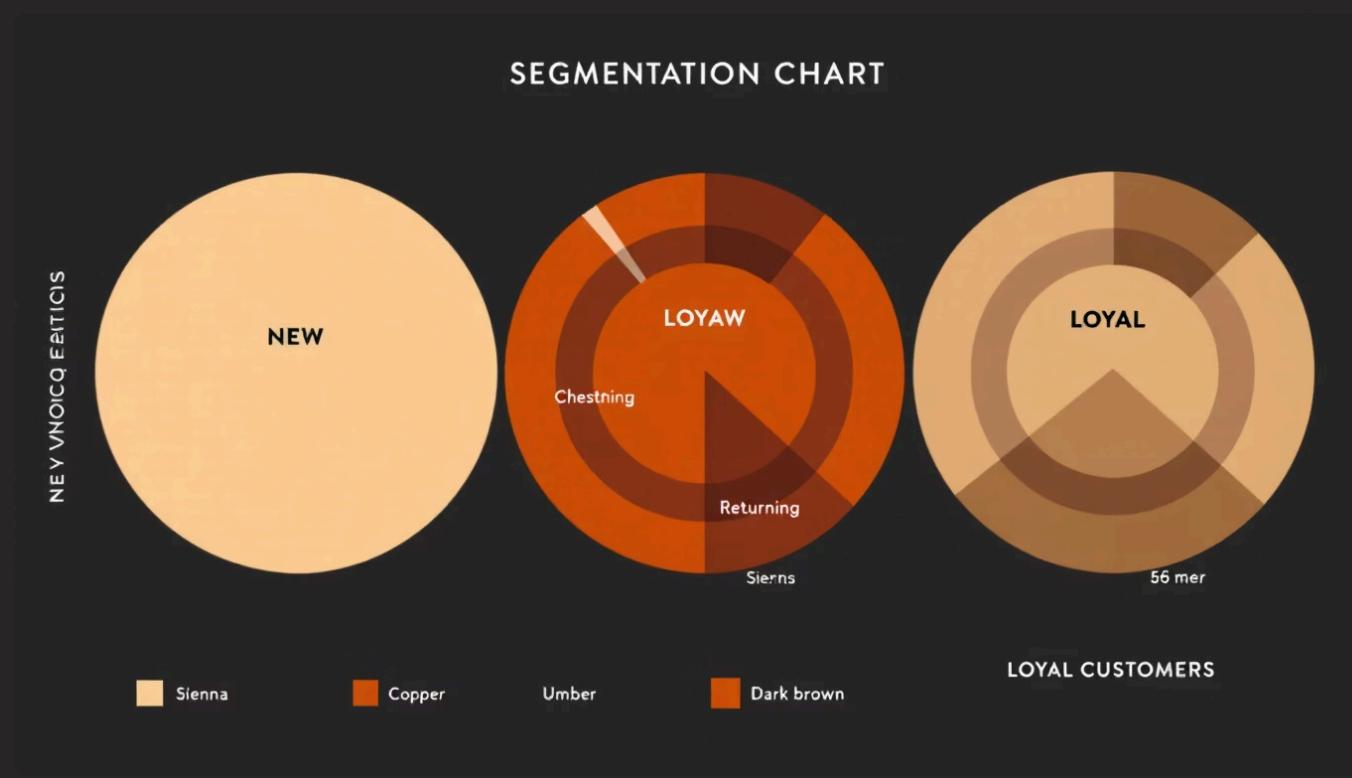


Discount-Dependent

Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), Pants (47.37%) — strong discount reliance.

Prioritize featuring top-rated items in marketing and review discount strategy for items with high discount dependence to protect margins.

Segments, Subscriptions & Loyalty



Customer Segments

Loyal: 3,116 · Returning: 701 · New: 83

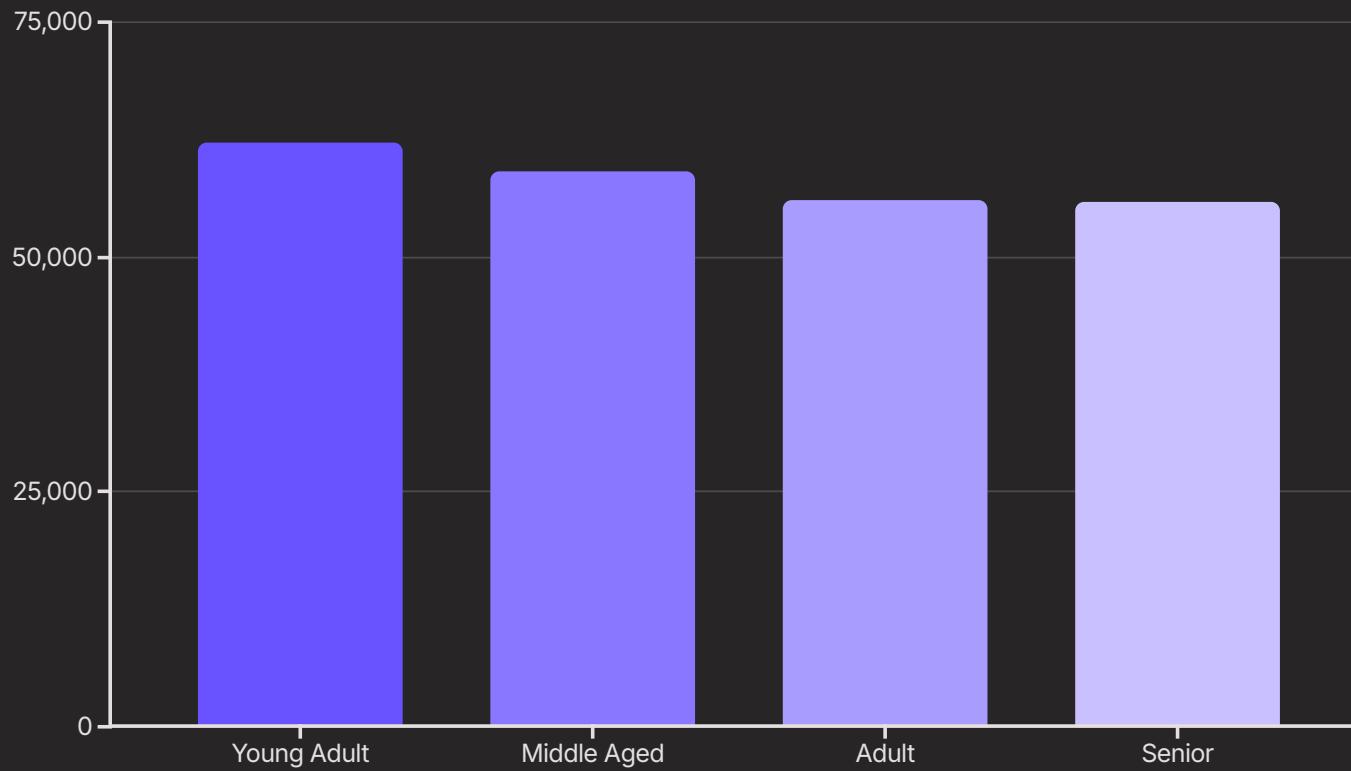
Subscriptions

Subscribers: 1,053 customers, Avg spend 59.49, Total revenue 62,645

Non-Subscribers

Non-subscribers: 2,847 customers, Avg spend 59.87, Total revenue 170,436

Age Groups & Revenue Distribution



Young Adults contribute the most revenue. Tailor targeted marketing and product positioning to high-revenue age groups; monitor sales vs orders per group for channel planning.

Recommendations & Next Steps

Boost Subscriptions

Promote exclusive subscriber benefits to increase recurring revenue.

Loyalty Programs

Reward repeat buyers to shift customers into the Loyal segment.

Review Discounts

Adjust discount policy for high-dependence products to protect margins.

Targeted Campaigns

Focus marketing on high-revenue age groups and express-shipping customers.

Dashboard & Monitoring

Use the Power BI dashboard for ongoing tracking and executive reporting.