Customer Shopping Behavior Analysis

3,900 purchases • 18 features • goal: uncover spending, segments, preferences, subscriptions

Dataset Summary

Size
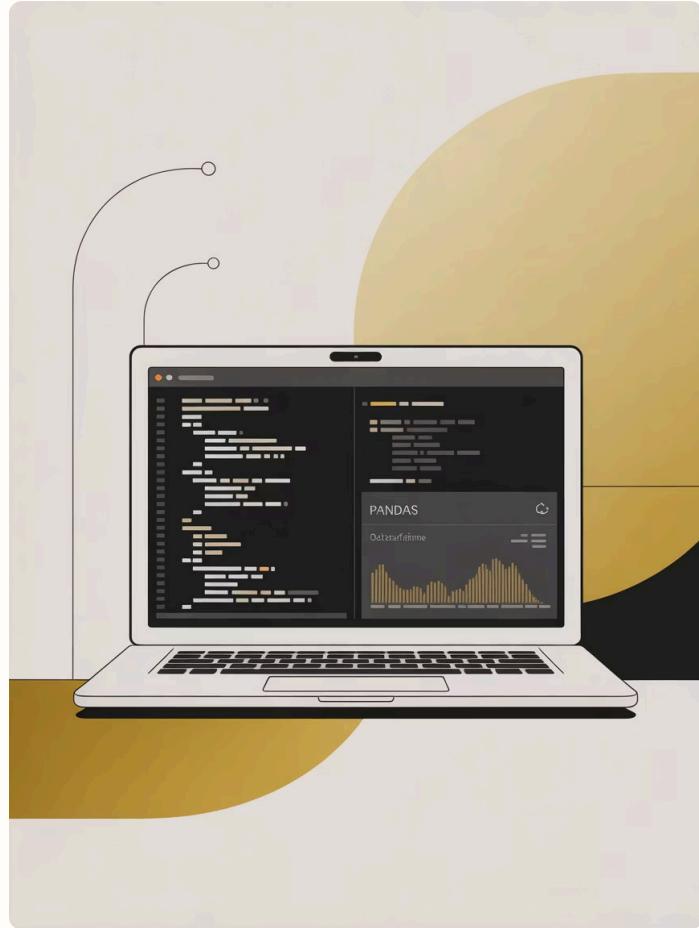
Rows: 3,900 • Columns: 18

Key Features

Demographics, purchase details, behavior, subscription

Missing Data
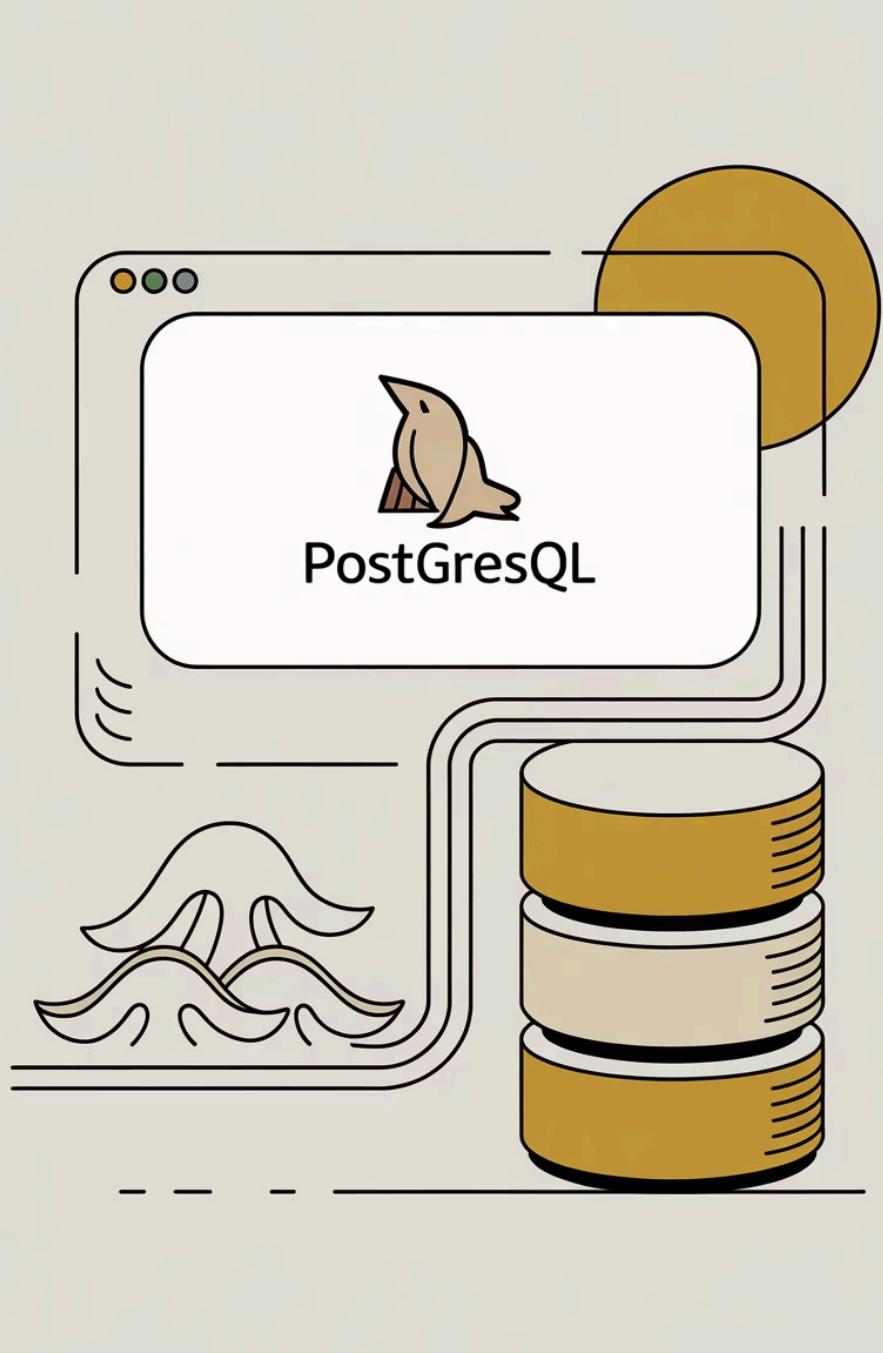
37 missing in Review Rating

Exploratory Analysis (Python)

Data load, df.info(), describe()

- Imputed review ratings by category median

- Renamed columns to snake_case

- Engineered age_group and purchase_frequency_days
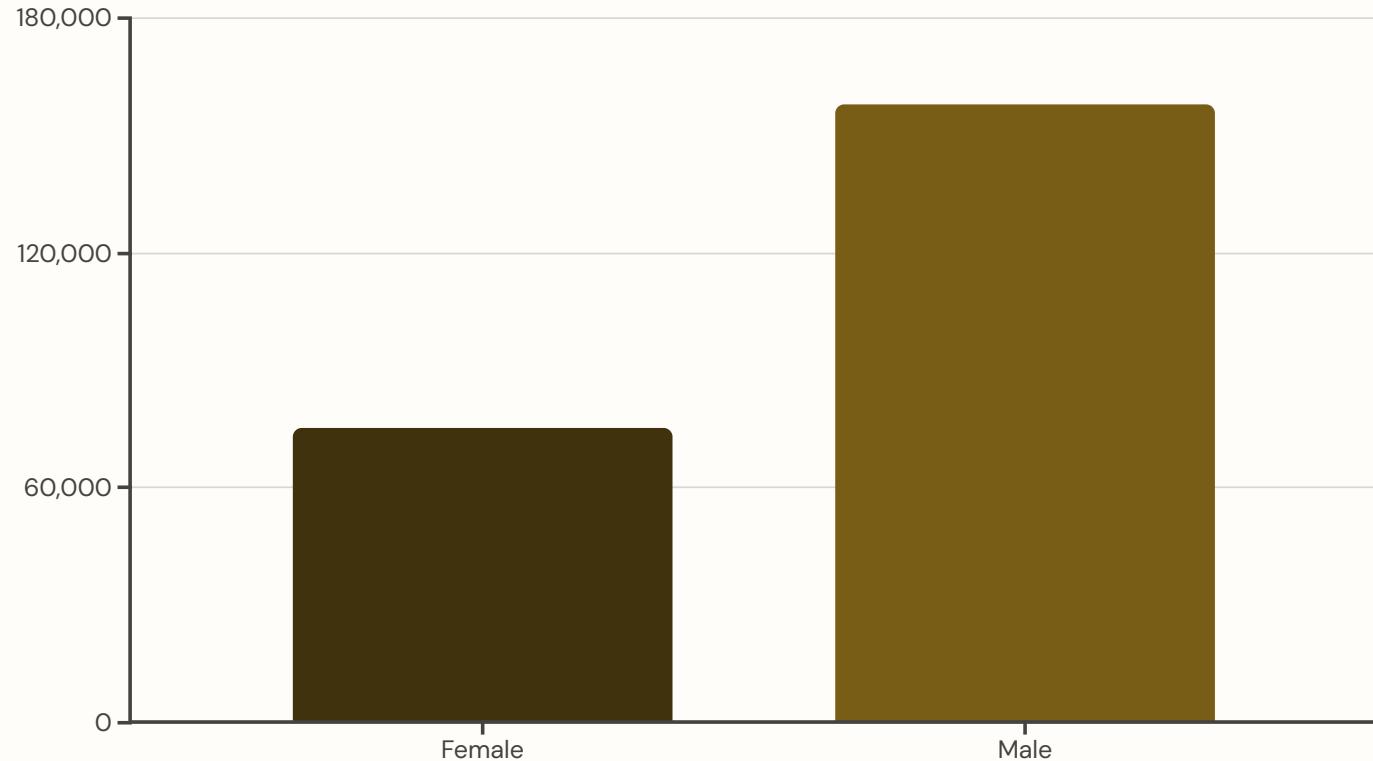
Database & Consistency

Checks

Verified discount_applied vs
promo_code_used; dropped
promo_code_used

Action

Loaded cleaned DataFrame
into PostgreSQL for SQL
analysis

Made with GAMMA

Revenue by Gender

Male revenue notably higher than female

Use targeted campaigns by gender

Discounts & High Spenders

Identified 839 rows of discount users; list of discount users who spent above average (examples: customer 4 spent 90, 9 spent 97)
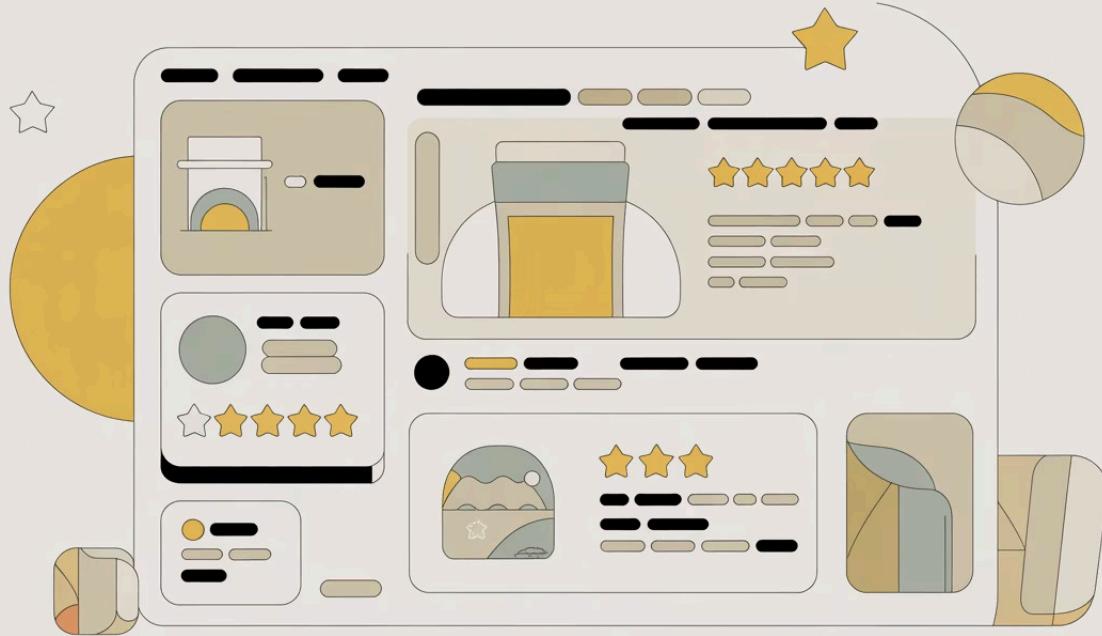
Discount-Dependent Items

Top: Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), Pants (47.37)

High-Spenders on Discount

Multiple customers used discounts yet exceeded avg purchase amount
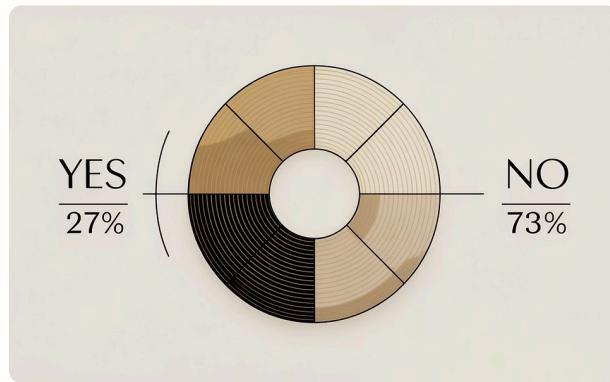
Top Products & Ratings

- Gloves — 3.86
- Sandals — 3.84
- Boots — 3.82
- Hat — 3.80
- Skirt — 3.78
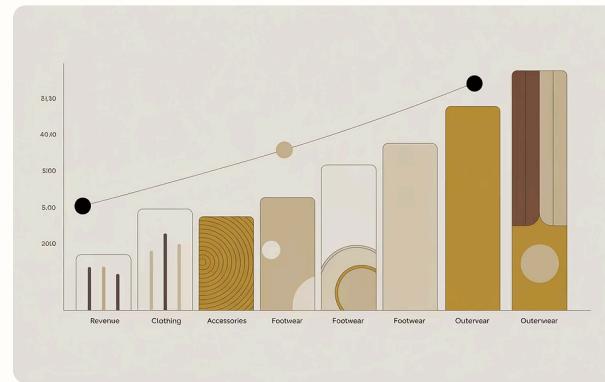
Prioritize top-rated items in promotions

Dashboard Highlights (Power BI)
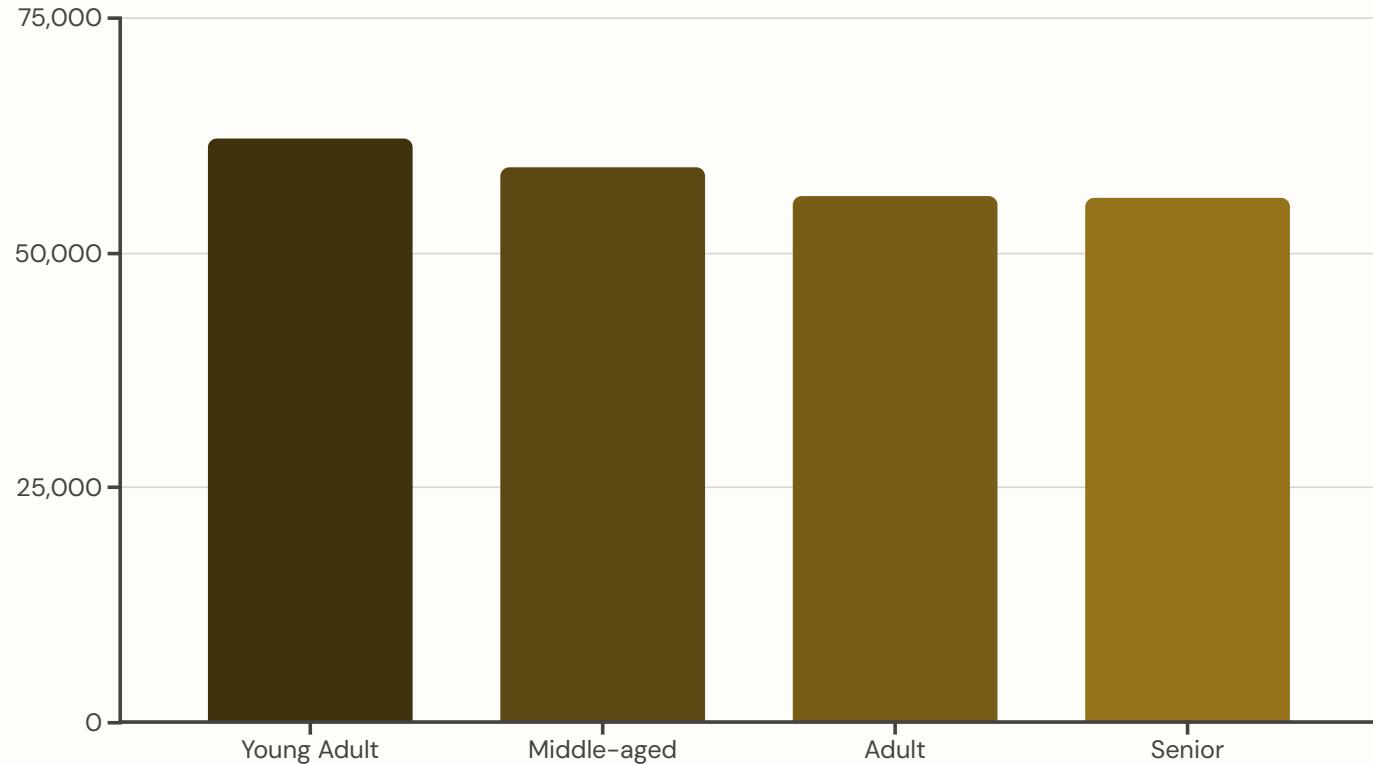
Subscriptions

Yes 27% • No 73%

Revenue by Category

Clothing $100K • Accessories $70K • Footwear $30K • Outerwear $20K

Age Groups & Revenue

Young Adult leads slightly; tailor marketing by age

Average purchase amount: $59.76 • Avg rating: 3.75

Business Recommendations

Boost Subscriptions
Promote exclusive subscriber benefits

Loyalty Programs
Reward repeat buyers to increase loyalty

Review Discounts
Balance sales volume with margins

Product Positioning
Feature top-rated, best-selling items

Targeted Marketing
Focus on high-revenue ages & express-shipping users

Power Bi Dashboard

interactive visualization bringing all insights together