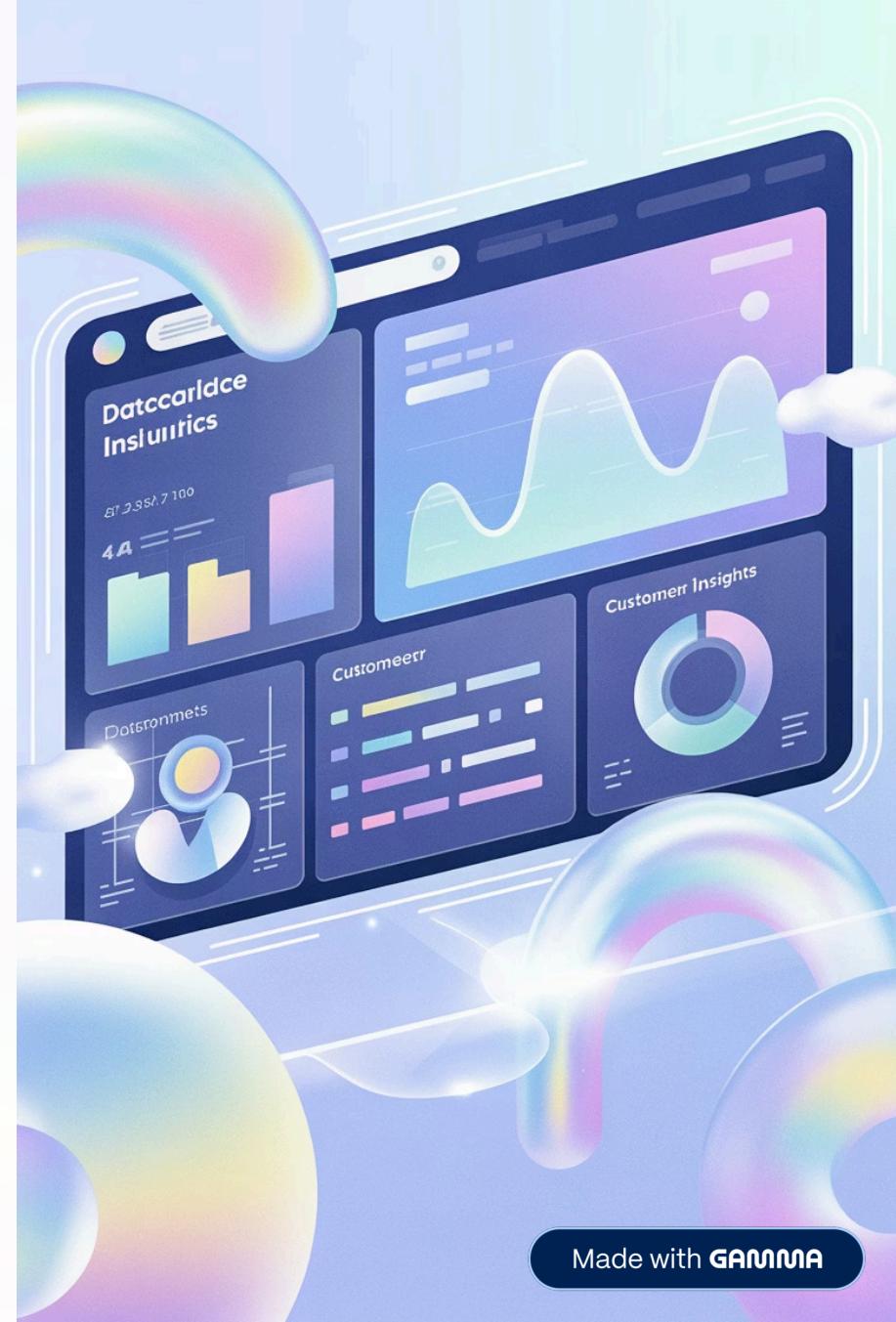


Customer Shopping Behavior Analysis

Uncovering insights from 3,900 purchases to drive strategic decisions



The Mission



Analyzing 3,900 purchases across product categories to uncover spending patterns, customer segments, and subscription behavior.

Moving beyond numbers to identify the "why" behind consumer choices—optimizing marketing and building long-term loyalty.

The Three-Pillar Approach

Data Engineering

Python (Pandas/NumPy) transformed raw data into analysis-ready format.
Handled 37 missing review ratings.

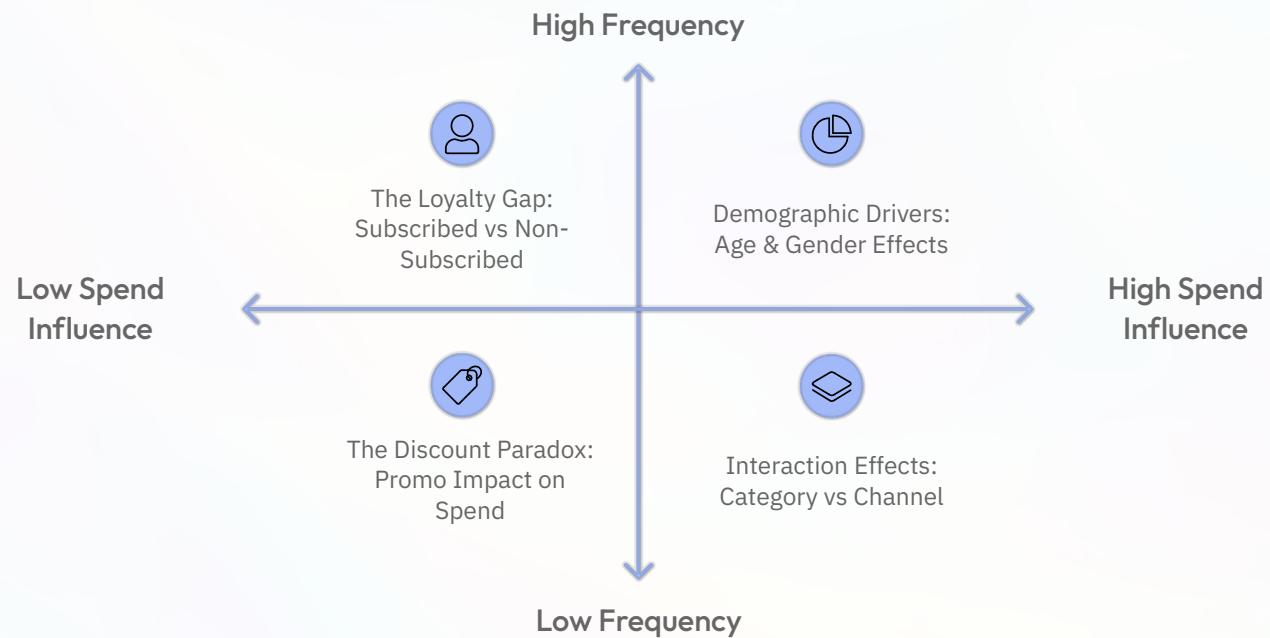
Investigative Analysis

SQL segmented customers and identified high-value purchase drivers—discounts, seasons, loyalty factors.

Visual Intelligence

Power BI dashboard communicates trends to stakeholders—location performance and subscriber behavior at a glance.

Why This Matters



The Loyalty Gap

Comparing purchase frequency: subscribed vs. non-subscribed customers

Demographic Drivers

How age and gender influence spending in Clothing vs. Footwear

The Discount Paradox

Do promo codes increase total spend or just reduce profit margins?

Dataset Summary

The Numbers

3,900

Purchase Records

18

Distinct Variables

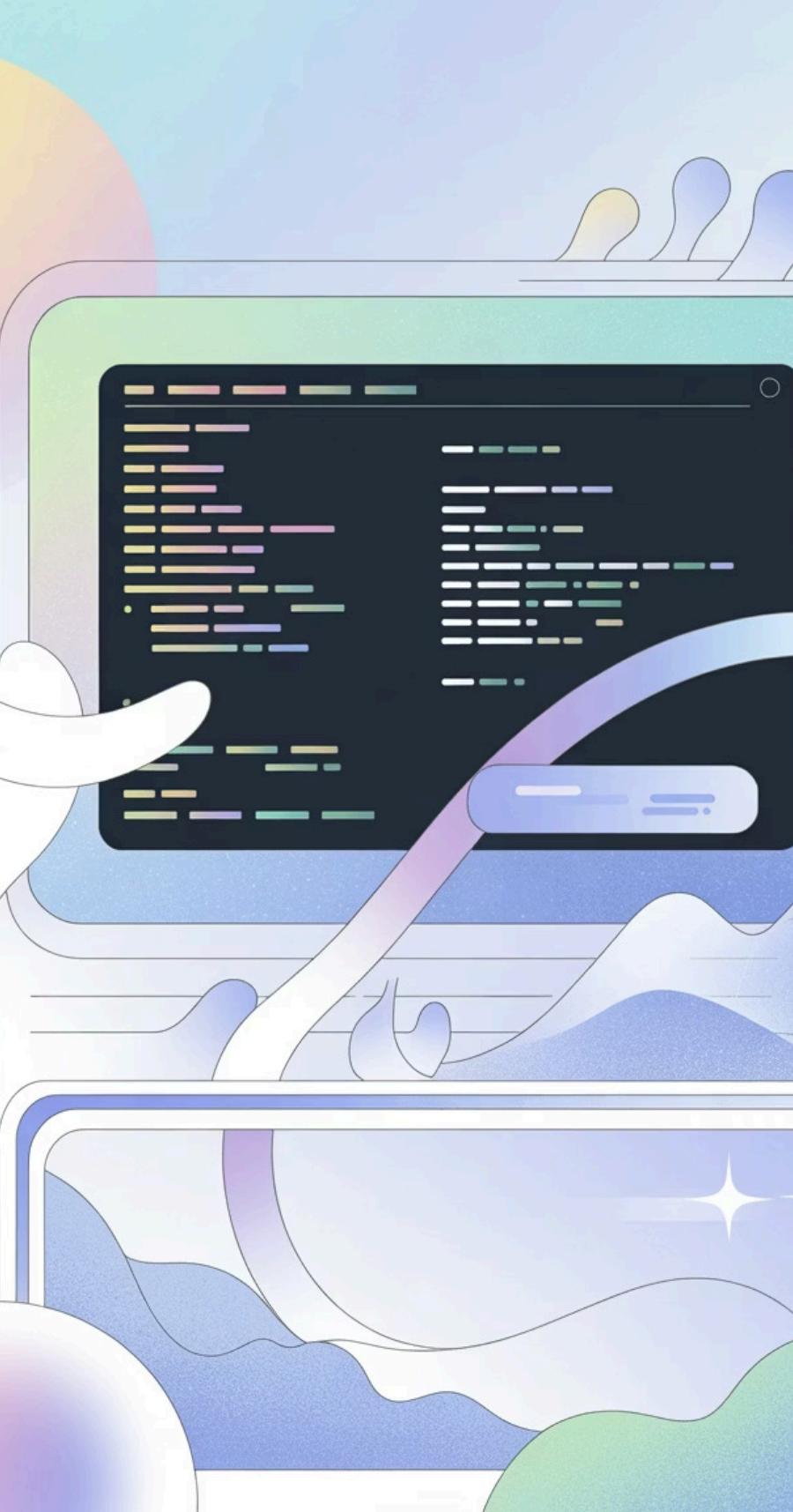
37

Missing Values

Review ratings imputed

Feature Breakdown

- **Customer Demographics:** Age, Gender, Location, Subscription Status
- **Purchase Details:** Item, Category, Amount (\$20–\$100), Season, Size, Color
- **Behavioral Indicators:** Discounts, Promo Codes, Previous Purchases, Frequency, Review Rating, Shipping Type



PHASE 1

Data Engineering & Cleaning

01

Data Profiling

Loaded 3,900 rows into Pandas. Audited data types, memory usage, statistical distributions.

02

Strategic Cleaning

Imputed 37 missing review ratings using median per category. Standardized to snake_case.

03

Feature Engineering

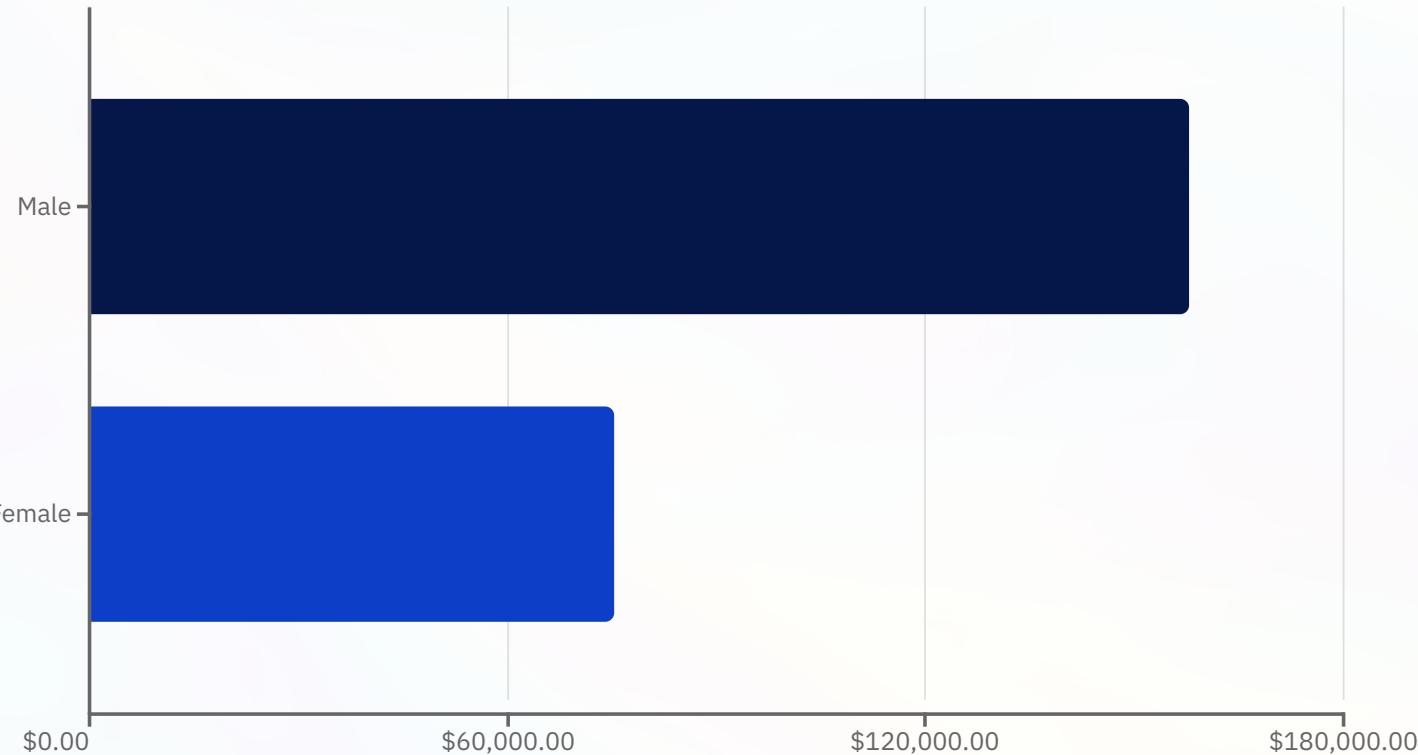
Created age_group bins and purchase_frequency_days for deeper insights.

04

Database Integration

Connected to PostgreSQL via SQLAlchemy. Injected clean DataFrame into structured table.

SQL Analysis: Key Findings



Revenue by Gender

Male customers generated 2x more revenue than female customers.

High-Spending Discount Users

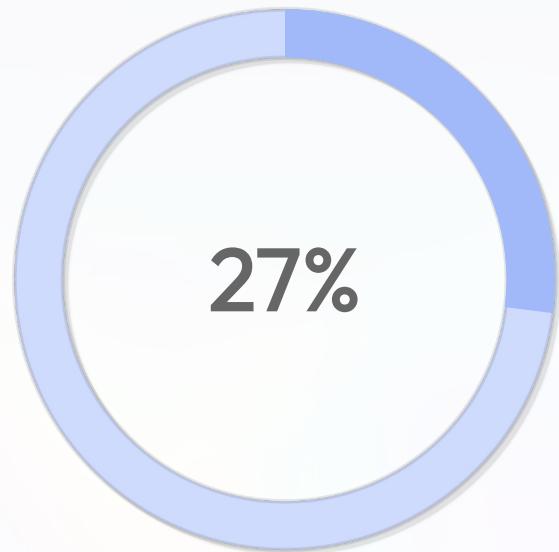
839 customers used discounts but still spent above average (\$59.76).

Top Products by Rating

Gloves (3.86), Sandals (3.84), Boots (3.82) lead in customer satisfaction.

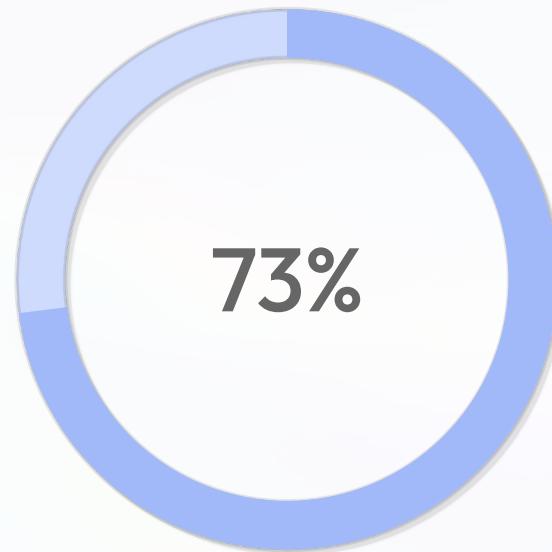
Customer Segmentation Insights

Subscription Status



Subscribed

1,053 customers



Non-Subscribed

2,847 customers

Customer Segments

- **Loyal:** 3,116 customers (80%)
- **Returning:** 701 customers (18%)
- **New:** 83 customers (2%)

Repeat Buyers & Subscriptions

958 subscribers have 5+ purchases vs. 2,518 non-subscribers—loyalty program shows promise.

Power BI Dashboard



The Road Ahead



Drive Subscriptions

Target customers with 5+ purchases for "Loyalty Upgrade" trial. Focus on exclusive perks, not just price.



Personalized Retention

Use customer segments to target "Returning" buyers with tiered rewards. Send "We Miss You" reminders based on purchase frequency.



Protect Margins

Recalibrate pricing on discount-dependent products. Offer free express upgrade on orders over \$80.



Strategic Marketing

Shift budgets to high-ROI segments (Montana, 18-70 age). Feature Top 5 rated products in campaigns.