

Customer Shopping Behavior Analysis – Portfolio Case Study

Project Overview

This project is an end-to-end customer behavior analysis designed as a portfolio case study for Data Analyst internships.

The goal was to transform raw transactional data into meaningful business insights using Python, SQL, and data visualization.

Problem Statement

Retail businesses often struggle to understand:

- Which customers drive the most revenue
- Which product categories perform best
- Whether subscriptions and discounts actually increase spending

This project answers these questions using real-world style customer data.

Dataset Summary

- ~3,900 customer records
- Average purchase value: \$59.76
- Average review rating: 3.75
- Data includes demographics, purchase behavior, subscriptions, discounts, and shipping preferences

Tools & Technologies

- Python (Pandas, NumPy) – data cleaning and analysis
- PostgreSQL – SQL-based analytics and segmentation
- Dashboard (Power BI) – KPIs and visual storytelling

Data Cleaning & Feature Engineering

Key steps performed:

- Standardized all column names to snake_case
- Renamed purchase_amount_(usd) to purchase_amount
- Handled missing ratings using median imputation

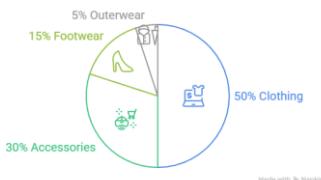
- Converted purchase frequency into numeric values
- Removed redundant columns after validation

These steps ensured clean, analysis-ready data.

Key Insights (SQL + Python)

1. Category Performance

Revenue Contribution by Clothing Category



- Clothing is the top revenue and sales-generating category
- Accessories and Footwear follow, while Outerwear underperforms

2. Customer Demographics

- Young Adults contribute the highest revenue
- Seniors show the lowest engagement

3. Subscription Impact

- Only 27% customers are subscribed
- Subscribers spend more on average than non-subscribers

4. Discount Analysis

- High-value customers remain high-value even when using discounts
- Certain products show heavy discount dependency

5. Customer Segmentation

- Loyal customers are fewer but contribute higher lifetime value

6. Revenue by Gender:

Revenue by Gender



Male customers generate higher revenue (**157,890**) compared to female customers (**75,191**), indicating stronger purchase activity or higher average spending among male shoppers.

Dashboard Highlights

The dashboard visually supports the analysis:

- KPI cards for customers, AOV, and ratings
- Revenue and sales by category
- Revenue and sales by age group
- Subscription vs non-subscription comparison
- Interactive filters for deeper exploration

Business Recommendations

- Focus inventory and promotions on Clothing
- Expand and promote subscription programs
- Target Young Adults with personalized marketing
- Use discounts strategically instead of blanket offers
- Improve engagement strategies for low-performing segments

Conclusion

This project followed a complete data analytics lifecycle, starting from raw data ingestion and cleaning, followed by feature engineering to make the data analysis-ready.

SQL was then used to answer key business questions related to revenue, customer segmentation, and purchasing behavior, while Python supported deeper exploratory analysis.

Finally, insights were communicated through an interactive dashboard and structured reporting, translating technical findings into clear business recommendations.

Overall, this case study demonstrates practical, real-world data analyst skills and showcases the ability to turn data into actionable decisions.