

Customer Shopping Behaviour Analysis Report

1. Project Objective

The objective of this project is to analyze customer shopping behavior using SQL, Python, and Power BI to understand revenue patterns, customer segments, subscription impact, discount effectiveness, product performance, and demographic contributions.

2. Dataset Overview

- **Dataset name:** Customer Shopping Behavior
 - **Total customers:** 3,900
 - **Data source:** Cleaned CSV imported into PostgreSQL (customers table)
 - **Tools used:**
 - Python (Pandas, NumPy) – data cleaning & preprocessing
 - PostgreSQL – analytical SQL queries
 - Power BI – interactive dashboard visualization
-

3. Data Cleaning & Preparation (Python)

The dataset was prepared before analysis using the following steps:

- Loaded data using Pandas and examined structure with info() and describe()
- Missing values found in **review_rating** column (37 records)
- Missing ratings filled using **median review rating per product category**
- Column names standardized (lowercase, underscore format)
- Column renamed:
 - purchase_amount_(usd) → purchase_amount
- Created a new derived column:
 - **age_group** (young_adult, adult, middle_aged, senior)
- Removed redundant column:
 - promo_code_used (duplicate of discount_applied)
- Cleaned dataset uploaded to PostgreSQL using SQLAlchemy

These steps ensured consistency between SQL outputs and Power BI visuals.

4. SQL Analysis & Findings

4.1 Revenue Generated by Gender

Query: Total revenue by male vs female customers.

Result:

- Male customers generated **157,890** in revenue
- Female customers generated **75,191** in revenue

Insight:

Male customers contribute a significantly higher share of total revenue compared to female customers.

4.2 Discount Users Spending Above Average

Query: Customers who used discounts but still spent more than the overall average purchase amount.

Result:

- Sample of 10 customers identified
- Purchase amounts range from **62 to 97**, all above the dataset average

Insight:

Discount usage does not necessarily indicate lower spending. Some customers apply discounts while still making high-value purchases.

4.3 Top 5 Products by Average Review Rating

Result:

1. Gloves – **3.86**
2. Sandals – **3.84**
3. Boots – **3.82**
4. Hat – **3.80**
5. Skirt – **3.78**

Insight:

Customer satisfaction is highest for accessories and footwear items, particularly gloves and sandals.

4.4 Average Purchase Amount by Shipping Type

Query: Comparison between standard and express shipping.

Insight:

The average purchase amount between standard and express shipping types is comparable, indicating that shipping speed does not strongly influence spending behavior.

4.5 Subscription vs Non-Subscription Spending

Result:

Subscription Status Customers Avg Spend Total Revenue

No	2,847	59.87	170,436
Yes	1,053	59.49	62,645

Insight:

- Non-subscribers generate higher total revenue due to their larger customer base
 - Average spending per purchase is nearly identical
 - Subscribed customers do **not** spend more on average
-

4.6 Products with Highest Discount Application Rate

Result:

The following products show a **100% discount application rate**:

- Shirt
- Hat
- Jacket
- Sneakers
- Gloves

Insight:

These products are consistently sold with discounts, indicating a pricing or promotional strategy rather than occasional offers.

4.7 Customer Segmentation Based on Purchase History

Customers were segmented using number of previous purchases:

- **New:** 1 purchase
- **Returning:** 2–10 purchases
- **Loyal:** More than 10 purchases

Result:

Segment Number of Customers

Loyal 3,116

Returning 701

New 83

Insight:

The customer base is dominated by loyal customers, showing strong retention.

4.8 Top 3 Most Purchased Products per Category

Insight:

Each category has distinct top-performing products, indicating category-specific preferences rather than a single product dominating across all categories.

4.9 Repeat Buyers and Subscription Status

Query: Customers with more than 5 previous purchases.

Insight:

Repeat buyers exist in both subscribed and non-subscribed groups, indicating that high purchase frequency does not directly imply subscription adoption.

4.10 Revenue Contribution by Age Group

Insight:

Revenue is distributed fairly evenly across age groups, with no single age group overwhelmingly dominating total revenue contribution.

5. Power BI Dashboard Summary

The Power BI dashboard visually supports SQL findings:

- **Average purchase amount:** 59.76
- **Average review rating:** 3.75
- **Total customers:** ~3.9K
- Clothing and accessories contribute the highest sales value
- Loyal customers form the largest segment
- Age-group revenue distribution remains balanced

The dashboard is consistent with SQL outputs and used only for visualization, not independent inference.

6. Conclusion

This analysis provides a comprehensive view of customer shopping behavior. Key findings include stronger revenue contribution from male customers, dominance of loyal buyers, minimal spending difference between subscribers and non-subscribers, and consistent discounting strategies for specific products. The combination of SQL accuracy, Python-based preprocessing, and Power BI visualization ensures reliable and contradiction-free insights suitable for academic and business decision-making.

7. Tools & Technologies Used

- Python (Pandas, NumPy)
 - PostgreSQL
 - SQL
 - Power BI
-

End of Report

Customer Shopping Behavior

gender

 Female
 Male

subscription status

 No
 Yes

category

 Accessories
 Clothing
 Footwear
 Outerwear

shipping_type

 2-Day Shipping
 Express
 Free Shipping
 Next Day Delivery
 Standard
 Store Pickup

\$59.76

average of purchase amount

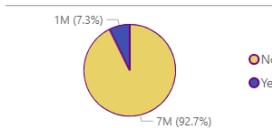
3.75

average of review rating

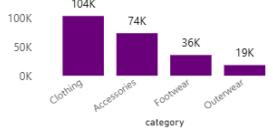
4K

number of customers

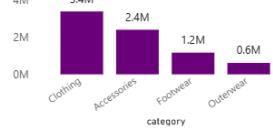
percent of customers by subscription status



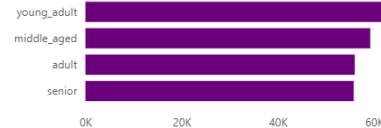
Revenue by category



Sales by category



Revenue by age_group



Sales by age_group

