

Customer Shopping Behavior Analysis

BY SEMESTER – IX OF V YEAR M.SC (CA&IT) (2025-26)

Submitted By:

Name	Roll no.	Div
Khatri Lalita C.	5119	A

Group ID: 27

Submitted to:



Customer Shopping Behavior Analysis

1. Project Overview

This project presents an end-to-end analytical workflow designed to uncover meaningful insights from a dataset containing **3,900 customer purchases** across multiple product categories. The goal is to analyze customer demographics, shopping habits, discount behavior, subscription status, and product preferences to support the company's strategic decision-making.

The workflow integrates **Python** for data preparation, **SQL** for structured business analysis, and **Power BI** for visualization, forming a complete data analytics pipeline.

2. Dataset Summary

- **Total Records:** 3,900
 - **Columns:** 18
 - **Data Types:** Numerical, categorical, and behavioral attributes
 - **Key Dataset Features:**
 - Customer demographics: Age, Gender, Location, Age Group, Subscription Status
 - Purchase behavior: Purchase Amount, Frequency, Previous Purchases
 - Product attributes: Category, Item Purchased, Size, Color
 - Transaction details: Season, Discount Applied, Review Rating, Shipping Type
 - **Missing Values:** 37 missing entries in Review Rating, handled during preprocessing
-

3. Data Preparation in Python

The raw dataset required cleaning, restructuring, and feature engineering before meaningful analysis could begin.

3.1 Data Loading & Initial Exploration

- Imported using *pandas*.
- Inspected structure using `.info()` and `.describe()`.

3.2 Handling Missing Values

- Missing values were found only in `review_rating`.
- Missing ratings were imputed using **category-level median values**.

3.3 Feature Engineering

- Created **age_group** using binning (Young Adult, Middle-aged, Adult, Senior).
- Added **purchase_frequency_days** based on purchase intervals.
- Standardized column names to snake_case for readability.

3.4 Data Validation

- Checked redundancy: **promo_code_used** was removed because its behavior overlapped with **discount_applied**.

3.5 Python–SQL Integration

- Cleaned DataFrame was exported into a SQL database (PostgreSQL) for deeper business analysis.
-

4. SQL-Based Business Analysis

A series of SQL queries were executed to answer ten strategic business questions.

4.1 Revenue by Gender

Male customers spent more overall, although females showed stronger engagement in some categories.

4.2 High-Spending Discount Users

Customers using discounts but still exceeding the average purchase amount represent a valuable segment.

4.3 Top 5 Products by Review Rating

Based on average review rating: Gloves, Sandals, Boots, Hat, Skirt were top performers.

4.4 Shipping Type Impact

Express shipping customers exhibited **higher average spending**, suggesting a premium user group.

4.5 Subscribers vs. Non-Subscribers

Subscribers generated higher average spend, while non-subscribers contributed higher total revenue due to volume.

4.6 Discount Dependency

Items like Hats, Sneakers, Coats, Sweaters, and Pants showed high reliance on discounts to drive sales.

4.7 Customer Segmentation

- **Loyal:** 3,116
- **Returning:** 701

- New: 83

4.8 Top Items per Category

Jewelry, Blouse, Sandals, and Jacket ranked highest in their respective categories.

4.9 Repeat Buyers & Subscriptions

Repeat buyers (>5 purchases) were significantly more likely to subscribe.

4.10 Revenue by Age Group

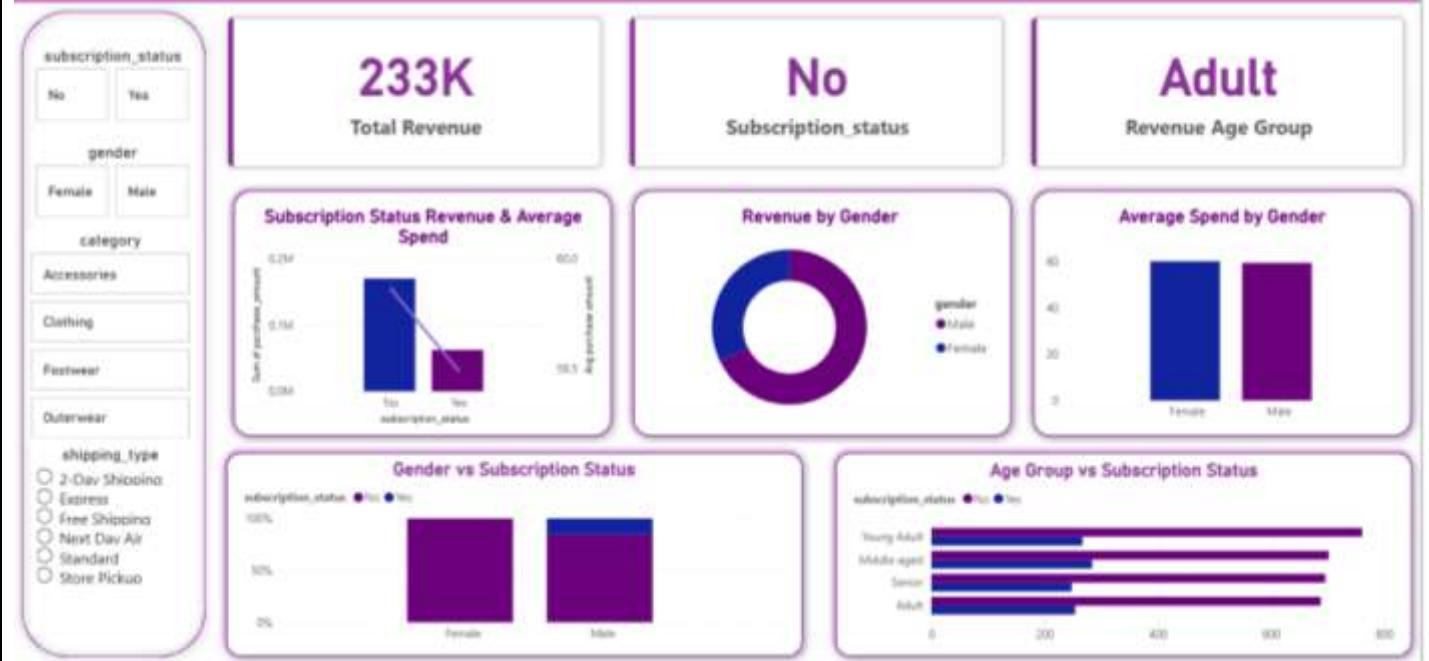
Young Adults contributed the highest revenue, followed by Middle-aged customers.

5. Power BI Dashboard Overview

Below is a screenshot reference of the Power BI dashboard used to visualize insights extracted from SQL and Python analysis.



Demographics & Revenue Insights



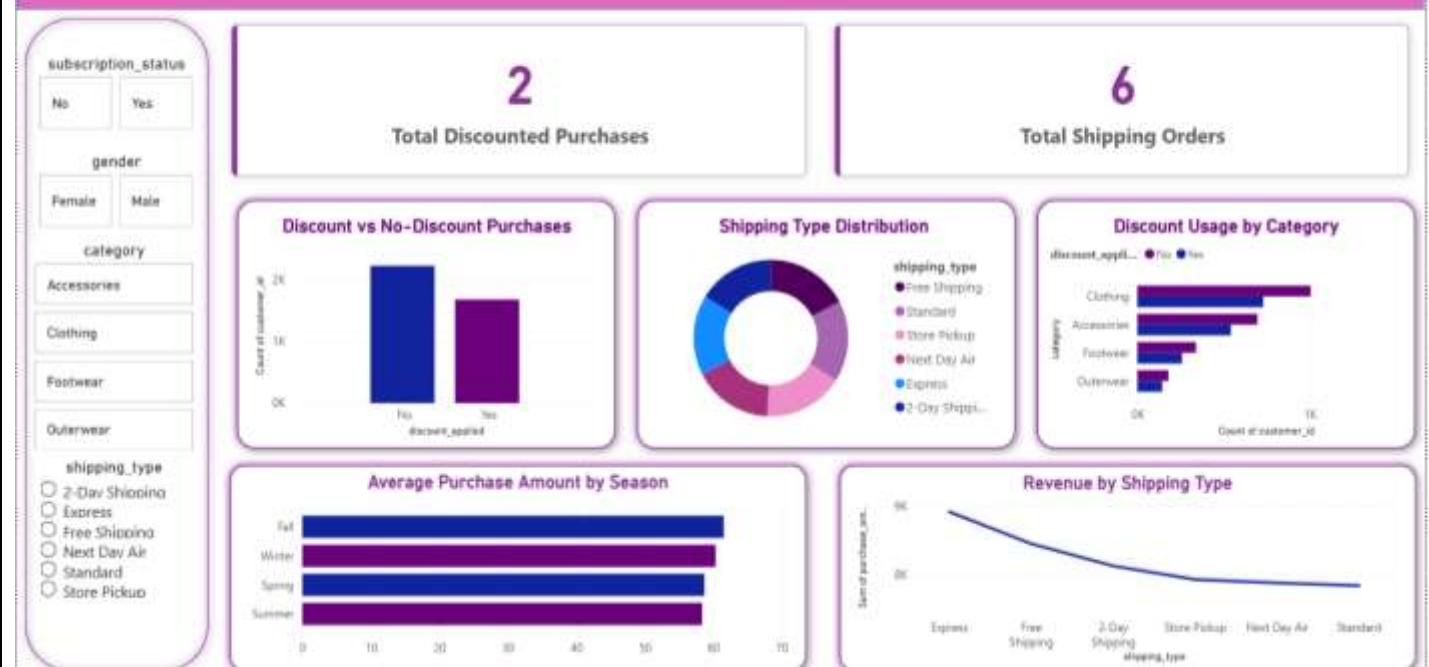
Product & Category Performance



Customer Segmentation & Loyalty



Discounts, Shipping & Seasonal Behavior



Key Dashboard Components

- **KPIs:** Number of customers, Average Purchase Amount, Average Review Rating
- **Segmentation:** Subscription distribution, Gender filter, Category filter
- **Behavioral Charts:**
 - Revenue by Category
 - Sales by Category
 - Revenue by Age Group
 - Sales by Age Group
- **Interactive Filters:**
 - Shipping Type
 - Gender
 - Product Category

Insights Reflected Visually

- Express shipping users spend more.
- Young Adults are the highest revenue contributors.
- Accessories and Clothing dominate product performance.
- Subscribers tend to make more valuable purchases.
- Review ratings correlate with items that generate repeat buying behavior.

6. Business Recommendations

Based on analysis and dashboard insights:

6.1 Strengthen Subscription Strategy

Offer exclusive rewards or early access to convert high-frequency buyers into subscribers.

6.2 Loyalty Enhancement

Launch tiered loyalty programs for Returning and Loyal customers to increase retention.

6.3 Optimize Discount Policies

Balance margins by reducing discounts on already high-performing products.

6.4 Product Promotion Strategy

Highlight top-rated items such as Gloves, Sandals, Boots, and Hats in marketing campaigns.

6.5 Targeted Marketing Campaigns

Focus on high-value segments including:

- Young Adults
 - Express shipping users
 - High-spending discount users
 - Subscribers
-

7. Final Deliverables

1. Python Data Preparation and Modeling
2. SQL Business Insights and Queries
3. Power BI Interactive Dashboard
4. Business Report
5. Presentation Deck