

# Customer Shopping Behavior Analysis – Project Report

---

## 1. Project Overview

This project analyzes customer shopping patterns using transactional data from **3,900 purchases** across multiple product categories. The objective is to generate insights related to spending behavior, customer segments, product preferences, and subscription impact. These insights support data-driven decision-making for marketing, product planning, and customer retention strategies.

---

## 2. Dataset Summary

- **Rows:** 3,900
  - **Columns:** 18
  - **Major Feature Groups:**
    - **Customer Demographics:** Age, Gender, Location, Subscription Status
    - **Purchase Details:** Item Purchased, Category, Purchase Amount, Season, Size, Color
    - **Behavior Indicators:** Discount Applied, Promo Code Used, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type
  - **Missing Values:** 37 missing values in the *Review Rating* column
- 

## 3. Exploratory Data Analysis (Python)

### Data Preparation

- Loaded dataset using **pandas** and examined structure with `info()` and `describe()`.
- Handled missing values in **review\_rating** by imputing the **median rating per category**.
- Standardized column names to **snake\_case** for readability.

### Feature Engineering

- Created **age\_group** by binning customer ages.
- Derived **purchase\_frequency\_days** from timestamp data.

- Evaluated redundancy between **discount\_applied** and **promo\_code\_used** → removed *promo\_code\_used*.

## Database Integration

- Connected Python to **Oracle SQL**.
  - Loaded the cleaned DataFrame for advanced SQL-based business analysis.

## 4. SQL-Based Business Analysis (Oracle SQL 11g)

**Key business questions answered:**

1. **Revenue by Gender:** Compared total revenue from male vs. female customers.

	GENDER	REVENUE
1	Male	157890
2	Female	75191

2. **High-Spending Discount Users:** Identified customers who used discounts but still spent above average.

	CUSTOMER_ID	PURCHASE_AMOUNT
1	182	61
2	183	96
3	188	94
4	190	82
5	191	70
6	192	76
7	194	100

3. **Top 5 Products by Rating:** Ranked products with the highest average review ratings.

ITEM_PURCHASED	AVG_RATING
1 Gloves	3.86142857142857142857142857142857142857142857142857142857
2 Sandals	3.844375
3 Boots	3.81875
4 Hat	3.8012987012987012987012987012987012987012987012987
5 Skirt	3.78481012658227848101265822784810126582

4. **Shipping Type Comparison:** Compared average purchase amounts across Standard vs. Express shipping.

SHIPPING_TYPE	Avg_Amount
1 Express	60.48
2 Standard	58.46

5. **Subscribers vs. Non-Subscribers:** Analyzed revenue and average spend for both groups.

SUBSCRIPTION_STATUS	COUNT(CUSTOMER_ID)	Avg_Spend	Total_Revenue
1 Yes	1053	59.49	62645
2 No	2847	59.87	170436

6. **Discount-Dependent Products:** Identified top 5 products most frequently bought with discounts.

ITEM_PURCHASED	DISCOUNT_RATE
1 Hat	50
2 Sneakers	49.66
3 Coat	49.07
4 Sweater	48.17
5 Pants	47.37

7. **Customer Segmentation:** Grouped customers as *New*, *Returning*, or *Loyal* based on purchase history.

CUSTOMER_SEGMENT	Number of Customers
1 Returning	701
2 New	83
3 Loyal	3116

8. **Top 3 Products per Category:** Extracted best sellers within each product category.

Category	Item Purchased	Total Purchases	Item Rank
1 Accessories	Jewelry	171	1
2 Accessories	Belt	161	2
3 Accessories	Sunglasses	161	3
4 Clothing	Pants	171	1
5 Clothing	Blouse	171	2
6 Clothing	Shirt	169	3
7 Footwear	Sandals	160	1

9. **Repeat Buyers & Subscription:** Tested if customers with more than 5 purchases are more likely to subscribe.

Subscription Status	Repeat Buyers
1 Yes	980
2 No	2583

10. **Revenue by Age Group:** Evaluated which age segments contribute highest revenue.

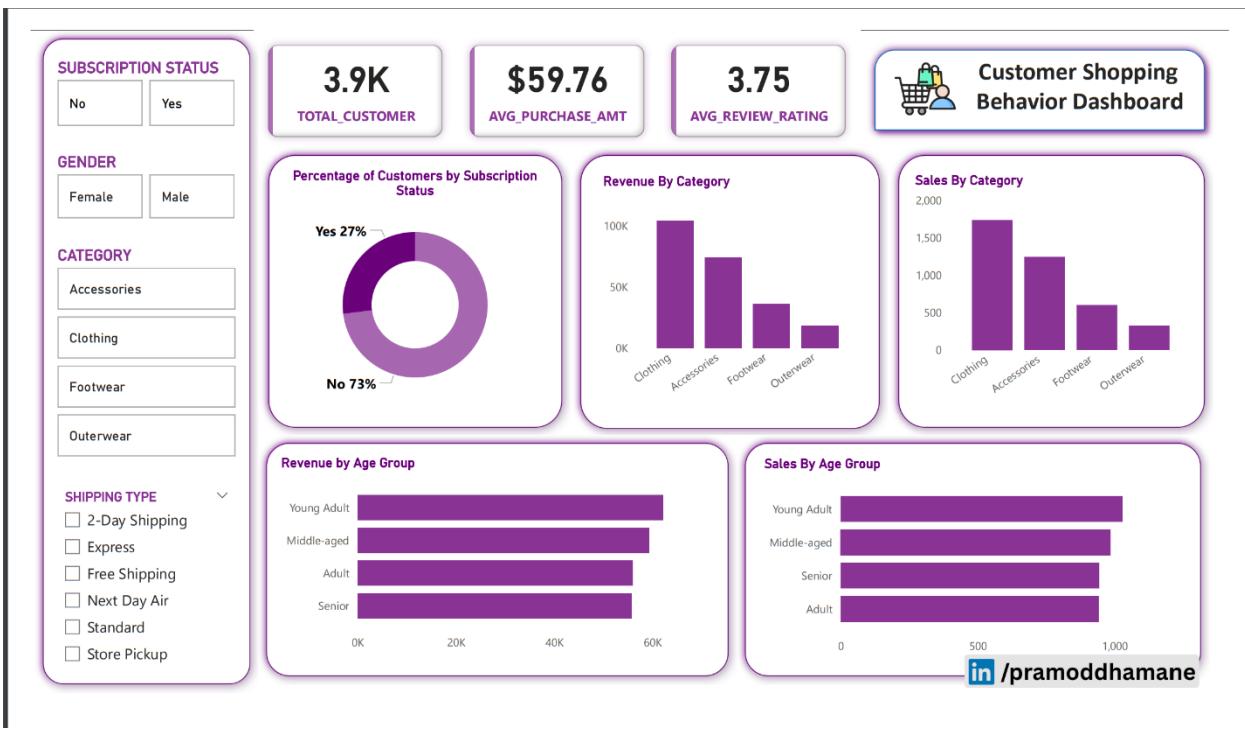
Age Group	Revenue Contribution
1 Young Adult	62143
2 Middle-aged	59197
3 Adult	55978
4 Senior	55763

## 5. Power BI Dashboard

Developed an **interactive Power BI dashboard** showcasing:

- Revenue by Gender, Age Group, Shipping Type
- Product Ratings & Top Categories
- Customer Segments
- Discount Usage Patterns
- Subscription Impact

The dashboard provides an intuitive, visual summary for decision-makers.



## 6. Business Recommendations

### 1. Boost Subscriptions

Promote exclusive subscriber benefits—early access, discounts, loyalty points.

### 2. Strengthen Loyalty Programs

Reward returning customers to convert them into *Loyal* customers.

### 3. Optimize Discount Strategy

Review discount-heavy product categories to maintain healthy profit margins.

### 4. Improve Product Positioning

Highlight top-rated and best-selling products in marketing campaigns.

### 5. Enable Targeted Marketing

- Focus on high-revenue age groups
- Prioritize Express shipping customers who show higher spending