

Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: 3,900
- Columns: 18
- Key Features:
 - ◆ Customer demographics (Age, Gender, Location, Subscription Status)
 - ◆ Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
 - ◆ Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
 - ◆ Missing Data: 37 values in Review Rating column

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- **Data Loading:** Imported the dataset using pandas.
- **Initial Exploration:** Used `df.info()` to check structure and `.describe()` for summary statistics.
- **Missing Data Handling:** Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.
- **Column Standardization:** Renamed columns to snake case for better readability and documentation.
- **Feature Engineering:**
 - Created `age_group` column by binning customer ages.
 - Created `purchase_frequency_days` column from purchase data.
- **Data Consistency Check:** Verified if `discount_applied` and `promo_code_used` were redundant; dropped `promo_code_used`.

4. Data Analysis using SQL (Business Transactions)

We performed structured analysis in BigQuery to answer key business questions:

1. **Revenue by Gender** – Compared total revenue generated by male vs. female customers.

gender ▼	revenue ▼
Male	157890
Female	75191

2. High-Spending Discount Users – Identified customers who used discounts but still spent above the average purchase amount.

Row	customer_id ▼	purchase_amount ▼
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68

Results per page: 50 ▼ 1 – 50 of 839 |< < > >|

3. Top 5 Products by Rating – Found products with the highest average review ratings.

Row	item_purchased ▼	average_product...
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.8
5	Skirt	3.78

4. Shipping Type Comparison – Compared average purchase amounts between Standard and Express shipping.

Row	shipping_type ▼	f0_ ▼
1	Standard	58.46
2	Express	60.48

5. Subscribers vs. Non-Subscribers – Compared average spend and total revenue across subscription status.

Row	subscription...	total_customers ▼	avg_spend ▼	total_revenue ▼
1	true	1053	59.49	62645.0
2	false	2847	59.87	170436.0

6. Discount-Dependent Products – Identified 5 products with the highest percentage of discounted purchases.

Row	item_purchased ▼	discount_rate ▼
1	Hat	50.0
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. Customer Segmentation – Classified customers into New, Returning, and Loyal segments based on purchase history.

Row	customer_segment ▼	number_of_custo... ▼
1	returning	701
2	loyal	3116
3	new	83

8. Top 3 Products per Category – Listed the most purchased products within each category.

Row	item_rank ▼	category ▼	item_purchased ▼	total_
1	1	Accessories	Jewelry	
2	2	Accessories	Belt	
3	3	Accessories	Sunglasses	
4	1	Clothing	Pants	
5	2	Clothing	Blouse	
6	3	Clothing	Shirt	

Results per page: 50 ▼ 1 – 11 of 11 |< < > >|

9. Repeat Buyers & Subscriptions – Checked whether customers with >5 purchases are more likely to subscribe.

Row	subscription...	repeat_buyers ▼	
1	true	958	
2	false	2518	

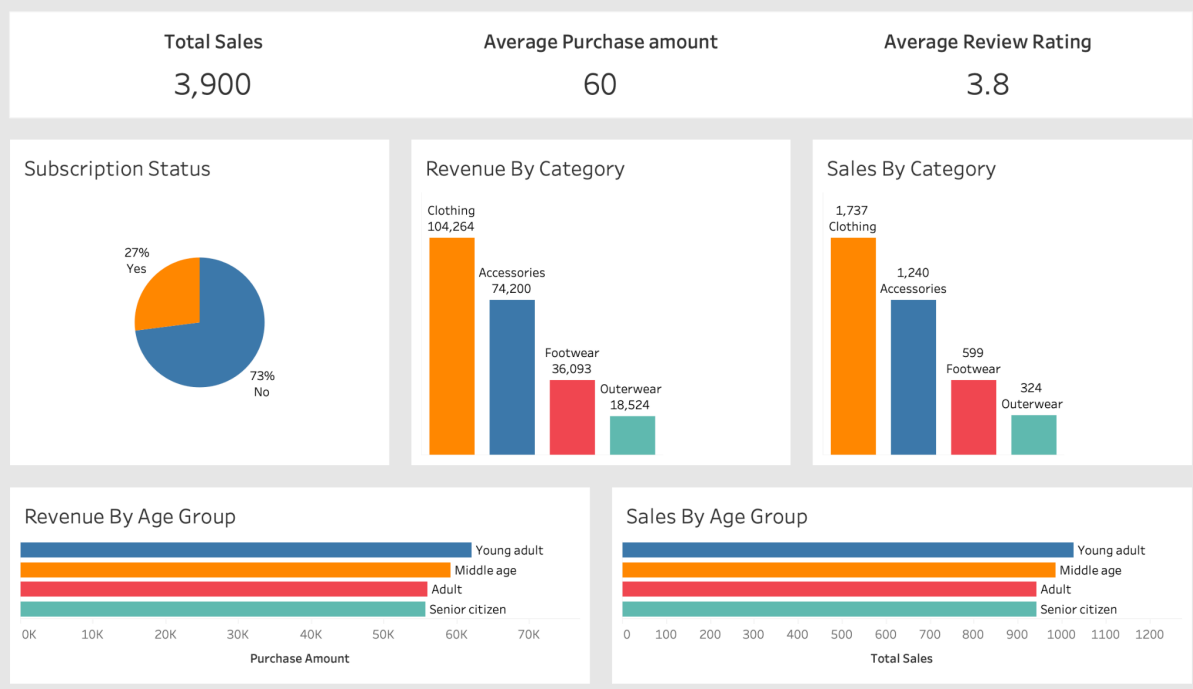
10. Revenue by Age Group – Calculated total revenue contribution of each age group.

Row	age_group ▼	total_revenue ▼	
1	Young adult	62143	
2	Middle age	59197	
3	Adult	55978	
4	Senior citizen	55763	

5. Dashboard in Tableau

Finally, we built an interactive dashboard in Tableau to present insights visually.

Customer Shopping Behavior Dashboard



6. Business Recommendations

- **Boost Subscriptions** – Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs** – Reward repeat buyers to move them into the “Loyal” segment.
- **Review Discount Policy** – Balance sales boosts with margin control.
- **Product Positioning** – Highlight top-rated and best-selling products in campaigns.
- **Targeted Marketing** – Focus efforts on high-revenue age groups and express-shipping users.