

Customer Shopping Behavior Analysis

1. Project Overview

This project examines customer shopping behavior using transactional data from 3,900 purchases across multiple product categories. The objective is to identify insights related to spending patterns, customer segmentation, product preferences, and subscription behavior in order to support strategic business decision-making.

2. Dataset Summary

Total Rows: 3,900

Total Columns: 18

Key Features Include:

Customer demographics such as age, gender, location, and subscription status

Purchase-related details including item purchased, category, purchase amount, season, size, and color.

Shopping behavior indicators such as discount applied, promo code usage, previous purchases, purchase frequency, review rating, and shipping type

Missing Values: 37 records in the review rating column

3. Exploratory Data Analysis Using Python

The analysis began with data preparation and cleaning using Python:

Data Loading: The dataset was imported using the pandas library

Initial Exploration: The dataset structure was examined using `df.info()`, and summary statistics were generated using `.describe()`

Missing Data Handling: Identified null values and addressed missing entries in the Review Rating column by imputing them with the median rating for each product category.

Column Standardization: Standardized column names by converting them to snake_case to improve readability and maintain consistent documentation.

Feature Engineering:

Generated an `age_group` column by categorizing customer ages into defined bins.

Created a `purchase_frequency_days` column derived from purchase data.

Data Consistency Check: Assessed the redundancy between `discount_applied` and `promo_code_used` and removed `promo_code_used` where overlap was identified.

Database Integration: Connected the Python script to PostgreSQL and loaded the cleaned DataFrame into the database to support SQL-based analysis.

4. Data Analysis using SQL (Business Transactions)

A structured analysis was conducted in PostgreSQL to address key business questions:

1. **Revenue by Gender** – Analyzed and compared the total revenue generated by male and female customers.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

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

	customer_id bigint	purchase_amount bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
Total rows: 839		Query complete 00:00

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

	item_purchased text	Average Product Rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.79

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

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

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

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

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

	item_purchased text	discount_rate numeric
1	Hat	50.00
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.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

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

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessori...	Jewelry	171
2	2	Accessori...	Sunglasses	161
3	3	Accessori...	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160

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

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

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

	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

11. Do discounts increase revenue or not(Business question) ?

	discount_applied text	total_orders bigint	avg_purchase_amount numeric	total_revenue numeric
1	No	2223	60.13	133670.00
2	Yes	1677	59.28	99411.00

12. High-Value Customer Identification

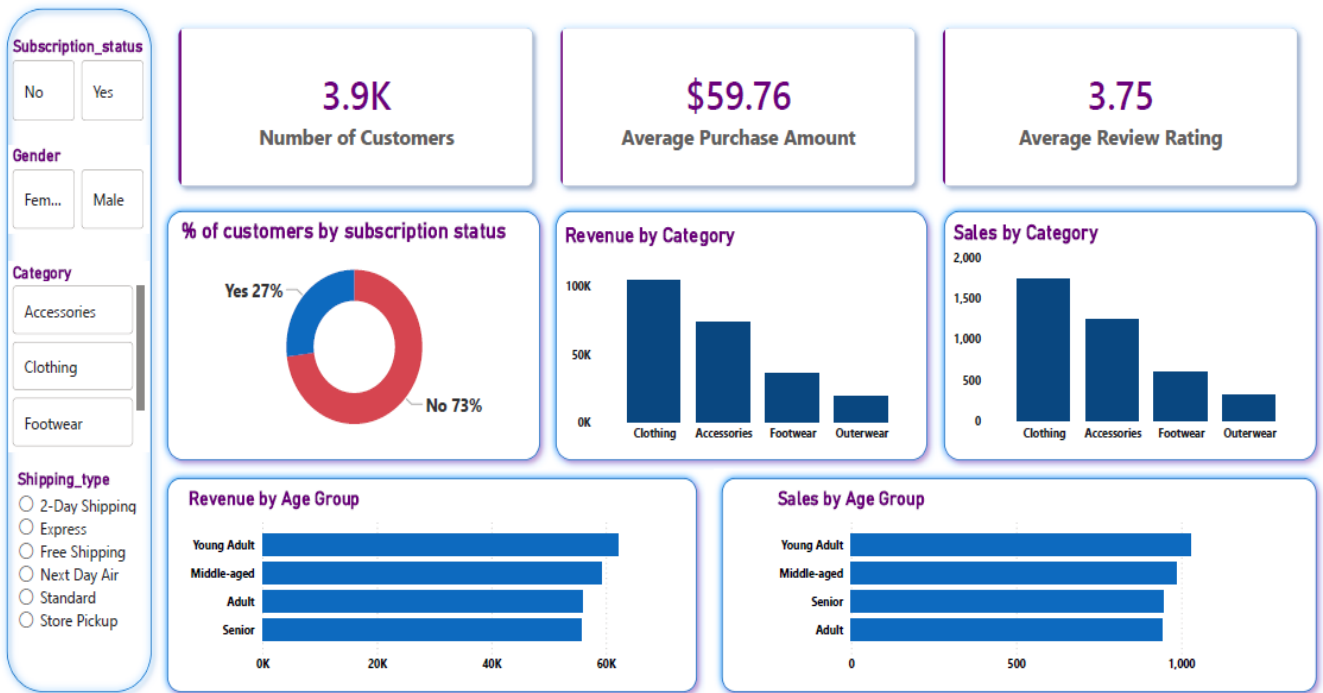
	customer_id bigint	clv_proxy numeric
1	249	5214.29
2	1480	5214.29
3	2215	5214.29
4	1848	5214.29
5	3477	5162.14
6	2975	5162.14
7	993	5162.14

Total rows: 10 Query complete 00:00:00.548

5. Dashboard in Power BI :-

Finally, we built an interactive dashboard in **Power BI** to present insights visually

Customer 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.