Customer Purchase Behavior Analysis

1. Project Overview

The objective of this project is to analyze customer shopping behavior using transactional and demographic data in order to understand buying patterns, customer segments, and factors that influence sales.

The insights generated from this analysis are intended to help businesses improve marketing strategies, optimize product offerings, and increase customer retention.

2. Dataset Summary

The dataset consists of customer demographic details and their shopping transactions.

Main attributes in the dataset

- Customer attributes
- Purchase details
- Marketing & pricing

Data preprocessing performed

Missing values in Review Rating were filled using the median rating of the corresponding product category

Column names were standardized to lowercase and snake_case, Redundant columns were removed

A new column age_group was created by dividing customers into quartiles:

- Young Adult
- Adult
- Middle Aged
- Senior

A new numeric column purchase_frequency_days was created by mapping frequency text (e.g., “Weekly”, “Monthly”) into number of days.

3. Exploratory Data Analysis (EDA) Using Python

EDA was performed using Pandas to understand data distribution, quality, and customer patterns.

Key EDA steps

- df.info() → checked data types and non-null values
- df.describe() → analyzed numerical and categorical distributions
- df.isnull().sum() → identified missing values
- value_counts() and unique() → analyzed frequency of customer behavior variables

Major insights from EDA

- ✓ Customers show uneven purchasing frequency, with certain frequencies (e.g., Monthly or Weekly) dominating
- ✓ Review ratings vary significantly across product categories
- ✓ Discount and promo code usage are highly correlated, leading to removal of one redundant column
- ✓ Age is well distributed, allowing meaningful age-based segmentation.

4. Data Analysis Using SQL

After cleaning and transforming the dataset in Python, it was exported to a relational database using SQLAlchemy and analyzed using SQL queries.

Structured analysis is performed in PostgreSQL to answer key business questions.

- Revenue by Gender – Analyzed total sales contribution across gender groups to identify high-value customer segments.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

- 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
8	16	81
9	20	90

Total rows: 839 of 839 Query complete 00:00:00.116

- Top 5 Products by Rating – Ranked products based on average customer review ratings to find top-performing items.

	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	3.78

- Shipping Type Comparison – Compared average spending between Standard and Express shipping customers.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

- Subscribers vs Non-Subscribers – Evaluated customer count, average spend, and revenue across subscription types.

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645
2	No	2847	59.87	170436

- Discount-Dependent Products – Identified products with the highest proportion of discounted purchases.

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.00
3	Coat	49.00
4	Sweater	48.00
5	Pants	47.00

- Customer Segmentation – Segmented customers into New, Returning, and Loyal groups based on purchase history.

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

- Top 3 Products per Category – Identified the three most purchased products in each category.

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

Total rows: 11 of 11 Query complete 00:00:00.080

- Repeat Buyers & Subscriptions – Analyzed subscription behavior among customers with more than five previous purchases.

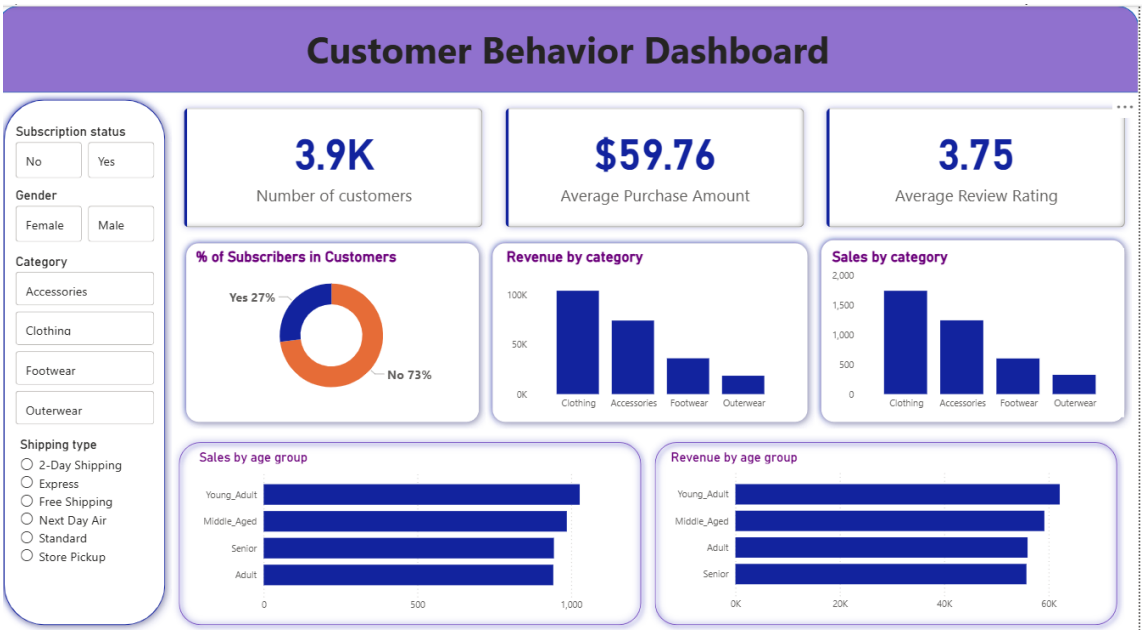
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

- Revenue by Age Group – Measured revenue contribution across different customer age groups.

	age_group text	revenue numeric
1	Young_Adult	62143
2	Middle_Aged	59197
3	Adult	55978
4	Senior	55763

5. Dashboard in PowerBI:

The cleaned and SQL-processed data was connected to Power BI for visualization and business analysis



Dashboards included

- Sales Overview
- Customer Segmentation
- Promotion Analysis
- Product Performance

Key value of Power BI

- ✓ Interactive filtering by category, age group, and purchase frequency
- ✓ Clear visualization of trends and outliers
- ✓ Business-ready reporting for decision makers

6. Business Recommendations

Based on the analysis, the following strategic insights can be derived:

- **Customer targeting** – Focus marketing campaigns on high-spending age groups, create personalized offers for frequent buyers
- **Pricing & promotions** – Discounts significantly influence purchase volume, target discounts toward high-value categories instead of blanket offers

- **Product strategy** – Invest more in top-performing categories, improve or discontinue low-rated or low-selling products
- **Customer retention** – Use purchase frequency to identify loyal customers, offer rewards and loyalty benefits to high-frequency buyers