Customer Purchasing Behavior Insights

1. Project Overview

This project explores patterns in customer purchasing behavior using a dataset containing 3,900 retail transactions across multiple product categories. The objective is to uncover meaningful insights related to spending habits, customer segments, product preferences, and subscription trends—ultimately supporting more informed, data-driven business decisions.

2. Dataset Summary

- **Total Rows:** 3,900
- **Total Columns:** 18

Key Data Attributes

- **Customer Demographics:** Age, Gender, Location, Subscription Status
- **Purchase Information:** Product Name, Category, Purchase Amount, Season, Size, Color
- **Behavioral Indicators:** Discount Applied, Promo Code Usage, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type

Missing Values:

- 37 missing entries in the *Review Rating* field

3. Exploratory Data Analysis using Python

The analysis began by preparing and cleaning the dataset using Python. Key steps included:

• Data Loading

Imported the dataset using pandas for exploration and preprocessing.

• Initial Inspection

Used functions like .info() and .describe() to understand the structure, datatype distribution, and statistical summary.

• Handling Missing Data

Checked for null values and replaced missing review ratings using the median rating within each product category.

• Standardizing Columns

Renamed all columns to *snake_case* to ensure consistency and easier reference during analysis.

• Feature Engineering

- Created an **age_group** field by categorizing customer ages.
- Derived a **purchase_frequency_days** feature to evaluate buying intervals.

• Consistency Validation

Reviewed redundancy between *discount_applied* and *promo_code_used*, ultimately removing the promo code field to avoid duplication.

• SQL Database Integration

Loaded the cleaned and enhanced dataset into PostgreSQL to support further business-focused SQL analysis.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discou Appli
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	39
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	22
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	Na
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	Na
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	Na
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	Na
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	Na
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	Na
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	Na

Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
3900	3900	3900.000000	3900	3900
2	2	NaN	6	7
No	No	NaN	PayPal	Every 3 Months
2223	2223	NaN	677	584
NaN	NaN	25.351538	NaN	NaN
NaN	NaN	14.447125	NaN	NaN
NaN	NaN	1.000000	NaN	NaN
NaN	NaN	13.000000	NaN	NaN
NaN	NaN	25.000000	NaN	NaN
NaN	NaN	38.000000	NaN	NaN
NaN	NaN	50.000000	NaN	NaN

4. Data Analysis using SQL (Business Transactions)

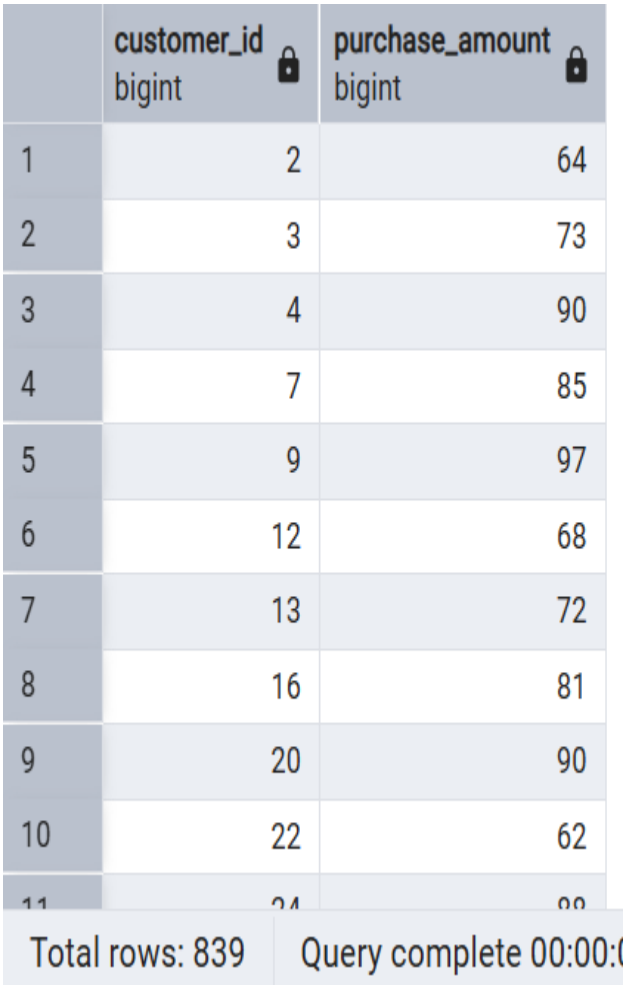
We performed structured analysis in PostgreSQL to answer key business questions:

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

3. **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
10	22	62
11	24	88

Total rows: 839 Query complete 00:00:(

4. **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	3.78

5. **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

6. **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

7. **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

8. **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

9. **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	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

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

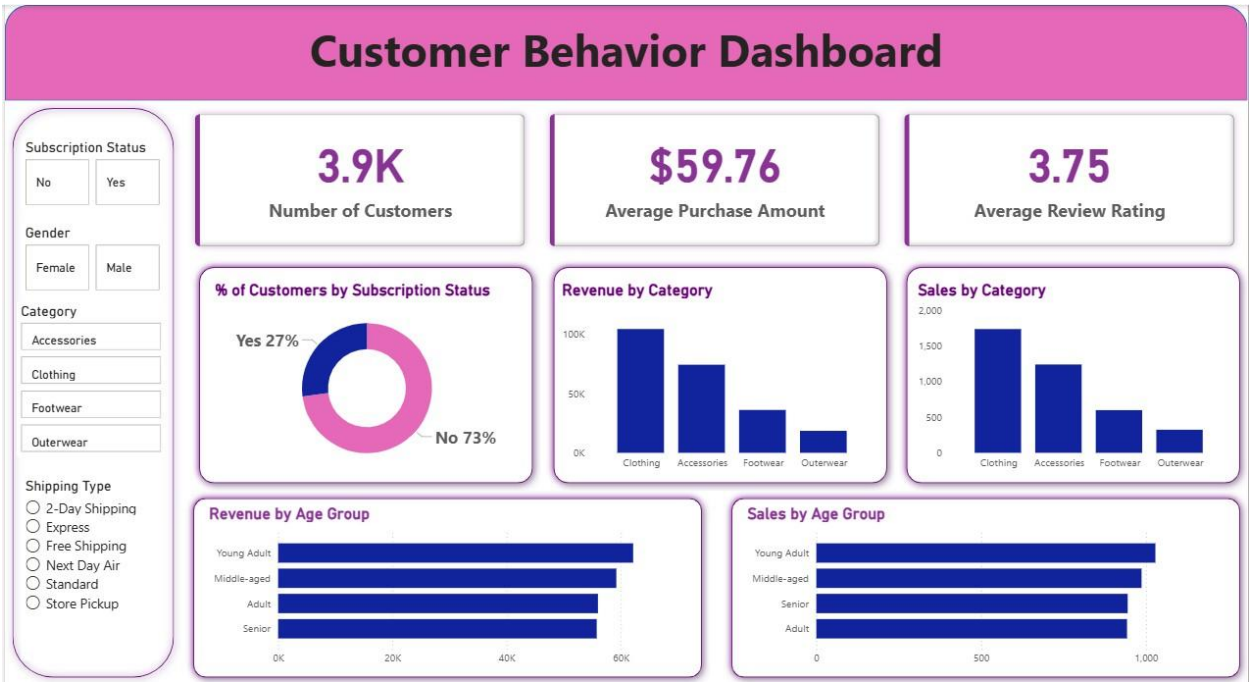
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

11. **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

5. Dashboard in Power BI

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



6. Business Recommendations

Based on the analysis, the following actions are recommended:

- **Expand Subscription Incentives**
  Offer additional perks to encourage more customers to subscribe.
- **Strengthen Loyalty Programs**
  Reward returning customers to convert them into loyal, high-value participants.

- **Evaluate Discount Strategy**
  Ensure promotional activity boosts sales without harming profit margins.
- **Optimize Product Positioning**
  Promote high-rated and top-selling products more prominently in marketing campaigns.
- **Target High-Value Segments**
  Focus marketing efforts on strong-spending demographics and express shipping use.