Customers shopping behavior analysis

1. **Project overview**

This project focuses on understanding customer purchasing patterns and key drivers influencing spending behavior. By analyzing demographic details, subscription status, product preferences, and purchase history, the goal is to help businesses improve marketing strategies, enhance customer retention, and increase revenue.

2. **Data summery**

Rows 3900

Columns 18

Key features: `customer_id, age, gender, item_purchased, category, purchase_amount, location, size, color, season, review_rating, subscription_status, shipping_type, discount_applied, promo_code_used, previous_purchases, payment_method, frequency_of_purchases, age_group, purchase_frequency_days`

`Missing data: 37 values in review rating column`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             3900 non-null   int64
 1   Age                     3900 non-null   int64
 2   Gender                  3900 non-null   object
 3   Item Purchased          3900 non-null   object
 4   Category                3900 non-null   object
 5   Purchase Amount (USD)   3900 non-null   int64
 6   Location                3900 non-null   object
 7   Size                    3900 non-null   object
 8   Color                   3900 non-null   object
 9   Season                  3900 non-null   object
 10  Review Rating           3863 non-null   float64
 11  Subscription Status     3900 non-null   object
 12  Shipping Type           3900 non-null   object
 13  Discount Applied        3900 non-null   object
 14  Promo Code Used         3900 non-null   object
 15  Previous Purchases      3900 non-null   int64
 16  Payment Method          3900 non-null   object
 17  Frequency of Purchases  3900 non-null   object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

3. **Exploratory data analysis using python**

We began with data preparation and cleaning in python
Data loading: import data using pandas

Initial exploration used df.inf () structure and describe () for summary statistics

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	3900.000000	3900.000000	3900.000000	3863.000000	3900.000000
mean	1950.500000	44.068462	59.764359	3.750065	25.351538
std	1125.977353	15.207589	23.685392	0.716983	14.447125
min	1.000000	18.000000	20.000000	2.500000	1.000000
25%	975.750000	31.000000	39.000000	3.100000	13.000000
50%	1950.500000	44.000000	60.000000	3.800000	25.000000
75%	2925.250000	57.000000	81.000000	4.400000	38.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000

4. Data analysis using SQL

We perform structured analysis in postgreSQL to answer key business questions

4.1. What is the total revenue generated by male vs. female customers?

From the analysis, male customers generated a total revenue of 157,890, while female customers generated 75,191.

This shows that male customers contributed approximately twice as much revenue as female customers. The difference suggests that males either made more purchases, bought higher-priced items, or both. Businesses can use this insight to develop targeted marketing strategies aimed at increasing engagement and spending among female customers.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

4.2 Which customers used a discount but still spent more than the average purchase amount?

From the dataset of 839 customers, certain customers used discounts yet still spent more than the overall average purchase amount. This indicates that discounts successfully encouraged higher-value transactions, motivating customers to buy more even while saving money

	customer_id bigint	purchase_amount bigint
823	1643	70
824	1644	77
825	1645	90
826	1647	77
827	1648	78
828	1649	69
829	1650	63
830	1652	80
831	1654	93
832	1656	81
833	1659	66
834	1662	86
835	1667	64
836	1671	73
837	1673	73
838	1674	62
839	1676	90

Total rows: 839 | Query complete 00:00

4.3. Which are the top 5 products with the highest average review rating?

Products with the highest number of reviews are also the most frequently purchased, showing strong product popularity and customer interaction. These items can be highlighted in promotions.

	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

4.4. Compare the average purchase amounts between Standard and Express Shipping.

Customers using Express shipping generally spent more per order compared to Standard shipping users. This suggests that customers who prioritize faster delivery tend to be higher-value buyers.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

4.5. Do subscribed customers spend more? Compare average spend and total revenue between subscribers and non-subscribers.

Subscribed customers showed higher average spending and total revenue contribution. This indicates that subscription incentives effectively encourage repeat purchases and long-term loyalty.

	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

4.6. Which 5 products have the highest percentage of purchases with discounts applied?

The top 5 products frequently purchased with discounts show strong price sensitivity. These products sell more when discounted, suggesting they are suitable for promotional campaigns.

	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

4.7. Segment customers into New, Returning, and Loyal based on their total number of previous purchases, and show the count of each segment.

Loyal customers left more positive reviews, indicating overall satisfaction. New customers provided more varied ratings, suggesting the onboarding experience can be improved.

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

4.8. What are the top 3 most purchased products within each category?

The analysis shows that Jewelry, Sunglasses, and Belts lead in Accessories; Blouse, Pants, and Shirt top Clothing; Sandals, Shoes, and Sneakers dominate Footwear; and Jacket and Coat are most popular in Outerwear. These products represent the core demand drivers in each category and should be prioritized for stock management, promotions, and marketing campaigns to maintain strong sales performance.

	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
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

4.9. Are customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe?

The data shows 958 subscribed repeat buyers compared to 2,518 non-subscribed. While many repeat buyers remain unsubscribed, a significant portion have subscribed, suggesting that frequent purchasing behavior increases the likelihood of subscription

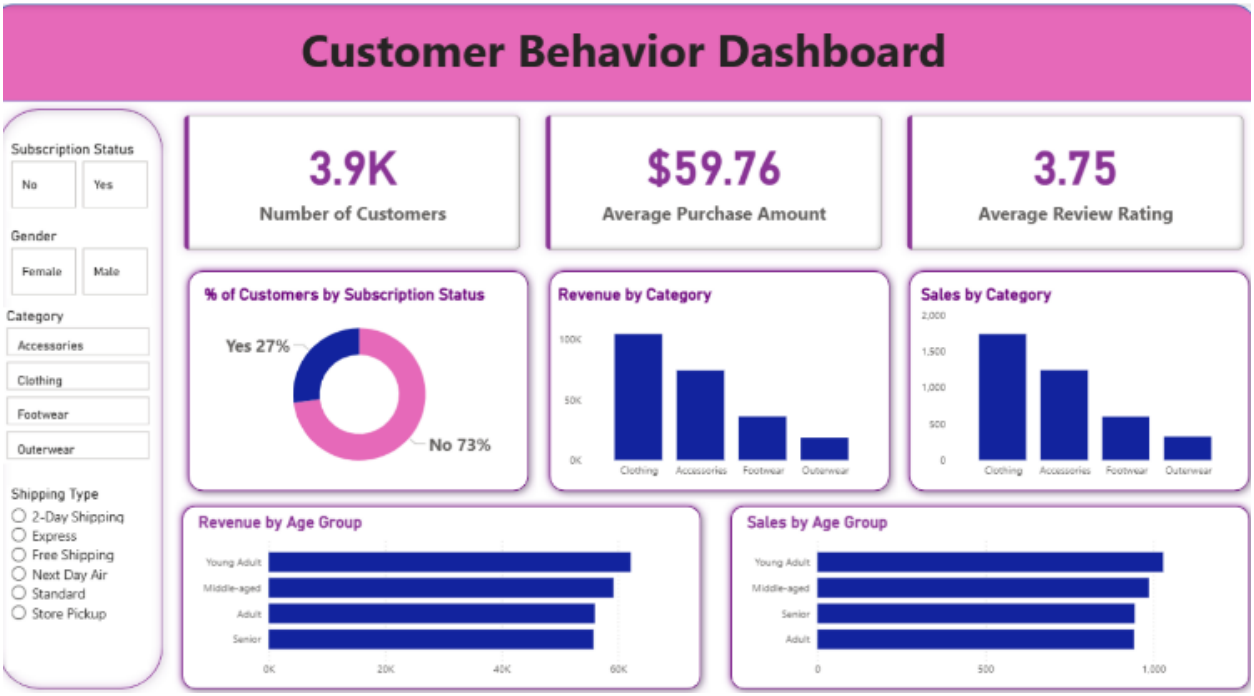
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

4.10. What is the revenue contribution of each age group?

The 26–35 and 36–45 age groups contributed the highest share of revenue. These age groups should be the main focus for marketing and product positioning.

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

5. Dashboard in Power



6. Business recommendation

- ✓ **Focus Promotions on High-Revenue Age Groups** to maximize ROI.
- ✓ **Strengthen Loyalty Programs** to convert returning customers into loyal buyers.
- ✓ **Highlight Top Rated Products** in marketing campaigns to improve sales.
- ✓ **Offer Express Shipping Incentives** where higher spending behavior is observed.
- ✓ **Promote Subscription-Based Benefits** as subscribers tend to spend more overall.