Customer Shopping Behavior Analysis

1.Project Overview

This project aims to analyze customer shopping behavior using transactional data collected from 3,900 purchases across various product categories. The primary objective is to identify customer spending patterns, segment customers based on their behavior, understand product preferences, and evaluate the impact of subscription status on purchasing decisions. The insights derived from this analysis can support data-driven business strategies and improve customer engagement.

2. Dataset Summary

- The dataset contains a total of 3,900 records and 18 features, covering customer demographics, purchase information, and shopping behavior.
- Customer demographic attributes include Age, Gender, Location, and Subscription Status. Purchase-related attributes consist of Item Purchased, Product Category, Purchase Amount, Season, Size, and Color. Shopping behavior is represented by Discount Applied, Promo Code Used, Previous Purchases, Purchase Frequency, Review Rating, and Shipping Type.
- Regarding data quality, the dataset has 37 missing values in the *Review Rating* column. All other features are complete and suitable for analysis.

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.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7
5	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9
6	7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2
7	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2

Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
Yes	Express	Yes	Yes	14	Venmo	Fortnightly
Yes	Express	Yes	Yes	2	Cash	Fortnightly
Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
Yes	Next Day Air	Yes	Yes	49	PayPal	Weekly
Yes	Free Shipping	Yes	Yes	31	PayPal	Annually
Yes	Standard	Yes	Yes	14	Venmo	Weekly
Yes	Free Shipping	Yes	Yes	49	Cash	Quarterly
Yes	Free Shipping	Yes	Yes	19	Credit Card	Weekly

```
a.describe()
```

	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

- **Missing Data Handling:** Checked for null values and imputed missing values in the Review Ratingcolumn 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_appliedand promo_code_used were redundant; dropped promo_code_used.
- **Database Integration:** Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

4.Data Analysis using SQL (Business Transactions)

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

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

	gender text	total_revenue numeric
1	Female	75191
2	Male	157890

2. Which customers used a discount but still spent more than 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

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

	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. Compare the average Purchase Amounts between Standard and Express Shipping.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

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

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

	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

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

	customer_segment text	number_of_customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

8. What are the top 3 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
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

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

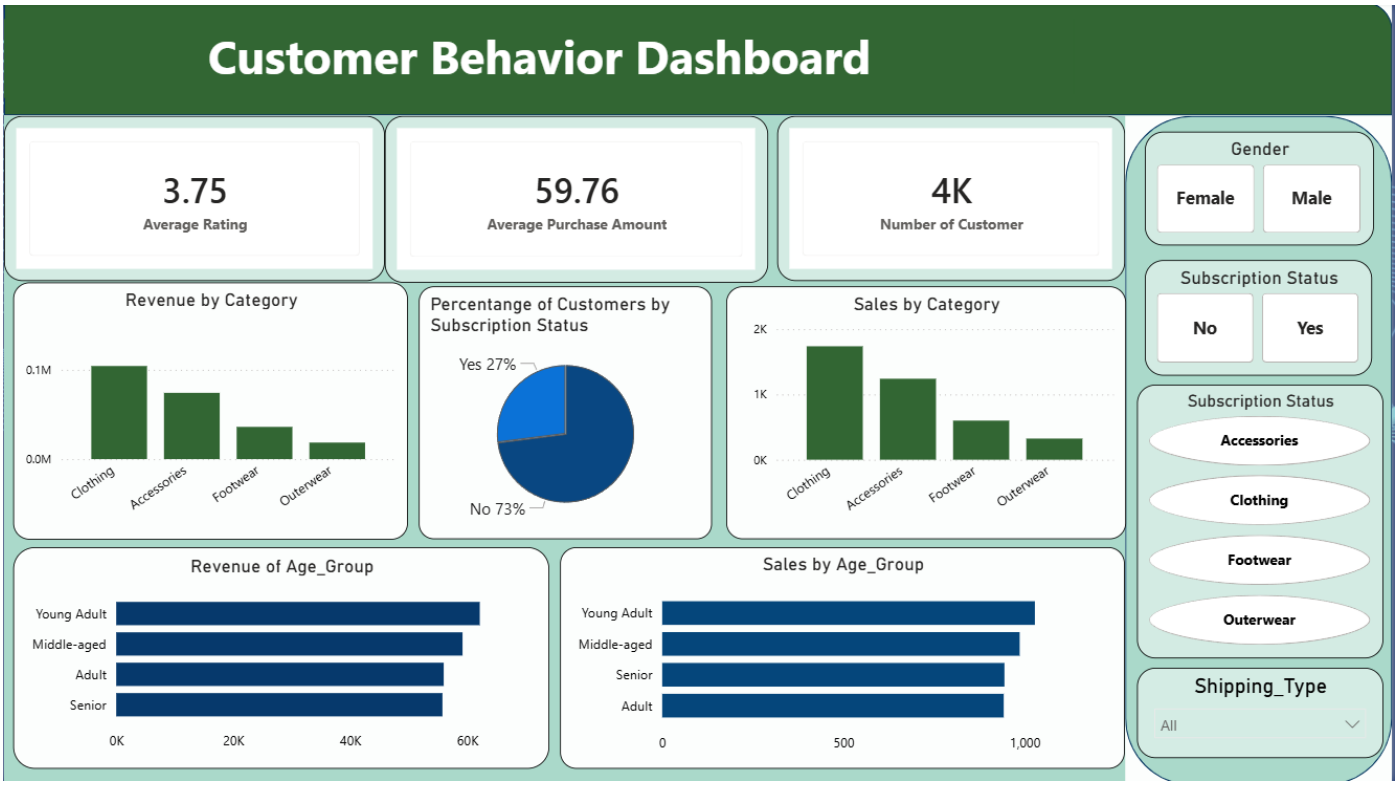
	subscription_status text	repeat_buyers bigint
1	No	2847
2	Yes	1053

10. What is the 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.

- **Increase Subscription Adoption.**

  Promote exclusive offers, early access to products, and special discounts to encourage more customers to opt for subscription plans.

- **Strengthen Customer Loyalty Programs**

  Introduce reward points, cashback, or tier-based benefits to incentivize repeat purchases and transition customers into the loyal segment.

- **Optimize Discount Strategy**

  Evaluate discount usage to ensure it effectively drives sales while maintaining healthy profit margins.

- **Enhance Product Positioning**

  Prioritize top-rated and best-selling products in marketing campaigns to improve visibility and conversion rates.

- **Implement Targeted Marketing Campaigns**

  Focus marketing efforts on high-revenue age groups and customers who prefer express shipping to maximize return on investment.