CUSTOMER BEHAVIOR ANALYSIS

---

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across multiple product categories. The primary goal is to uncover insights into spending patterns, customer segmentation, product preferences, and subscription trends that can guide strategic business decisions.

2. Dataset Summary

- **Total Records:** 3,900
- **Total Features:** 18
- **Key Attributes:**
    - Customer Demographics: Age, Gender, Location, Subscription Status
    - Purchase Details: Item Purchased, Category, Purchase Amount, Season, Size, Color
    - Shopping Behavior: Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type
- **Missing Data:** 37 missing values in the Review Rating column

3. Exploratory Data Analysis using Python

The dataset was cleaned and prepared in Python to ensure accuracy and consistency.

- **Data Loading:** Imported the dataset using pandas.
- **Initial Exploration:** Used df.info() to inspect data structure and .describe() for summary statistics.
-

	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	N
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	N
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	N
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	N
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	N
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	N
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	N

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

- **Handling Missing Values:** Identified missing values and imputed those in the Review Rating column using the median rating of each product category.
- **Column Standardization:** Renamed all columns using snake_case for improved readability.
- **Feature Engineering:**
    - Created a new column, age_group, by categorizing customer ages.
    - Derived purchase_frequency_days from purchase history data.
- **Data Consistency Check:** Evaluated whether discount_applied and promo_code_used were redundant and removed promo_code_used after confirming overlap.
- **Database Integration:** Connected the cleaned DataFrame to PostgreSQL for advanced SQL-based business analysis

.

4. Data Analysis using SQL (Business Transactions)

SQL queries were executed in PostgreSQL to answer key business questions:

1. **Revenue by Gender:** Compared total revenue between 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 spent above the average purchase amount.

	customer_id bigint	purchase_amount_(usd) bigint
8	16	81
9	20	90
10	22	62
11	24	88
12	29	94
13	32	79
14	33	67
15	35	91
16	37	69
17	40	60
18	41	76
19	43	100

Total rows: 839 Query complete 00:00:00.145

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	3.78

Total rows: 5 | Query complete 00:00:00.305

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**: Analyzed average spend and total revenue between subscribers and non-subscribers.

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

6. **Discount-Dependent Products**: Determined five 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 as New, Returning, or Loyal 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 frequently 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. **Repeat Buyers and Subscriptions:** Examined whether customers with more than five purchases were more likely to have a subscription.

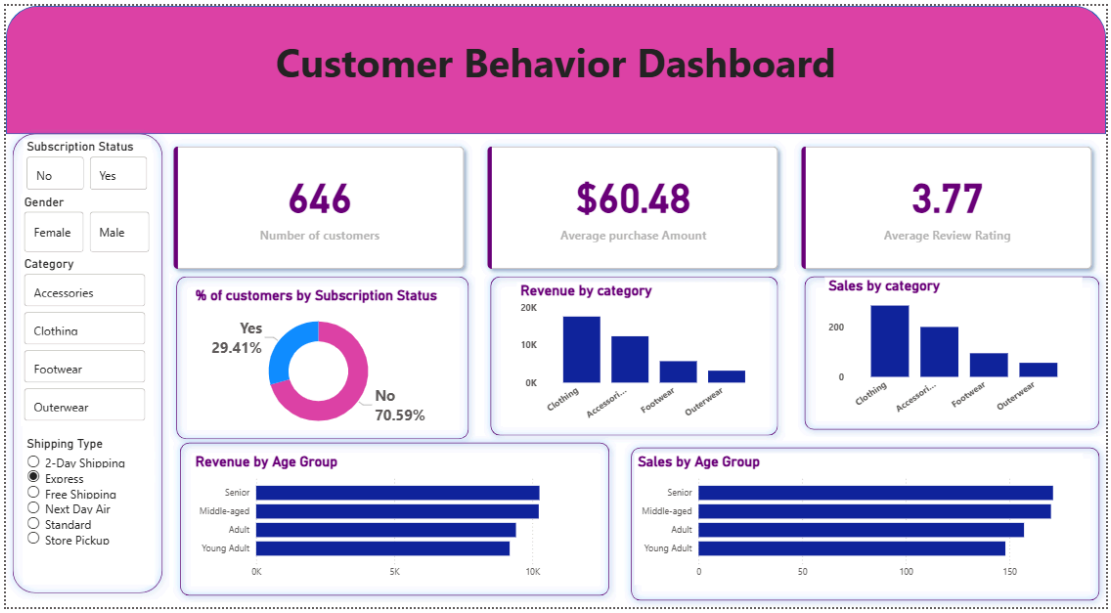
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. **Revenue by Age Group:** Calculated total revenue contributed by 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

An interactive dashboard was developed in Power BI to visually present the analytical findings and provide actionable insights for decision-makers.



6. Business Recommendations

- **Boost Subscriptions:** Offer exclusive perks and targeted campaigns to increase subscriber count.

- **Implement Loyalty Programs:** Reward returning customers to transition them into the Loyal segment.
- **Review Discount Policies:** Balance promotional discounts with profitability to maintain healthy margins.
- **Enhance Product Positioning:** Promote top-rated and high-selling products in marketing campaigns.
- **Focus Marketing Efforts:** Target high-revenue age groups and frequent express-shipping users for better ROI.