Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 50000 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: 50000
- Columns: 23

Key Features: - 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)

Data Drill | Data Science | Jobs | Career

3. Exploratory Data Analysis using Python

We began with data preparation and Cleaning in Python
Data Loading : Import the dataset through pandas
Initial Exploration : used df.info() to check the structure and .describe() for summary stats.

	Customer_ID	Age	Gender	Location	Item_Purchased	Category	Size	Color	Season	Price	...	Subscription	Promo_Code_Used	Previous_Purchases	Payment_
0	1	56	Male	Ohio	Sandals	Outerwear	L	Purple	Fall	52.85	...	No	No	53	Cre
1	2	69	Female	California	Shoes	Clothing	L	Gray	Winter	240.54	...	Yes	Yes	29	De
2	3	46	Female	Florida	Jeans	Accessories	XL	White	Spring	207.77	...	No	No	12	Bank
3	4	32	Male	New York	Coat	Outerwear	M	Blue	Fall	174.06	...	No	No	49	Bank
4	5	60	Male	Virginia	Shorts	Footwear	M	Pink	Fall	43.50	...	No	No	27	Cre
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
95	96	42	Male	Florida	Sweater	Footwear	S	Blue	Fall	284.84	...	No	No	37	
96	97	62	Female	Virginia	Sandals	Footwear	S	Gray	Fall	239.12	...	No	No	8	
97	98	58	Female	Virginia	Blouse	Footwear	XS	Purple	Winter	181.00	...	No	Yes	5	Cre
98	99	46	Female	North Carolina	Pants	Accessories	S	Pink	Fall	64.58	...	No	No	58	De
99	100	32	Male	Georgia	Handbag	Footwear	S	Green	Winter	90.93	...	Yes	Yes	18	

100 rows × 23 columns

Promo_Code_Used	Previous_Purchases	Payment_Method	Frequency_of_Purchase	Return_Status	Delivery_Time (Days)	Customer_Satisfaction	Purchase_Date	Total_Spent
No	53	Credit Card	Annually	Yes	13	Happy	2022-03-28	2079.92
Yes	29	Debit Card	Quarterly	No	1	Happy	2023-10-05	2429.96
No	12	Bank Transfer	Bi-Weekly	No	5	Neutral	2023-06-27	4585.96
No	49	Bank Transfer	Bi-Weekly	Yes	4	Happy	2021-01-24	3496.37
No	27	Credit Card	Annually	No	11	Happy	2023-05-07	1688.55
...	...	...	...	...	...	...	...	...
No	37	PayPal	Weekly	No	6	Happy	2021-11-12	3962.03
No	8	PayPal	Quarterly	No	12	Unhappy	2021-12-16	1097.57
Yes	5	Credit Card	Weekly	No	8	Happy	2023-09-06	704.11
No	58	Debit Card	Bi-Weekly	No	3	Neutral	2021-05-18	629.26
Yes	18	PayPal	Quarterly	No	5	Happy	2022-03-03	2745.04

Missing Data Handling: Checked for null values and imputed missing values in the **Review Rating** column 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_applied** and **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.

SQL Querries:

Revenue by male vs female customers

Gender	Revenue
Male	60673021.45000004
Female	60852846.38000017
Other	4679577.2700000005

-- Which Customer used a discount but still spent more than the average purchase amout??

Customer_ID	Total_Spent
3	4585.96
4	3496.37
6	3921.33
9	3215.06
11	4357.12
14	4266.62
16	2592.21
17	4738.95
20	4939.03
21	2673.63
22	3299.23
23	3016.19
25	2635.28
27	4002.36

-- which are the top 5 products with the highest average reveiw rating ??

Item_Purchased	Average product rating
Sweater	3.04
Shirt	3.04
Sneakers	3.02
Blouse	3.01
Handbag	3.01

3.01

-- compare the average purchase between standard and express shipping

Result Grid | Filter Rows:

Shipping_Type	Avg_Total_Spent
Standard	2530.14
Express	2528.45

-- Do subscibed customers spend more ? Compare average spend and total revenue between sub and non sub ..??

Subscription	Total_Customers	Average_Spent	Total_Revenue
No	35040	2523.44	88421180.17
Yes	14960	2525.69	37784264.93

-- Which 5 products have the highest percentage of purchase with discount applied ??

Item_Purchased	Discount_Rate
Sandals	89.90
Shorts	89.66
Shoes	89.32
Handbag	89.10
Sunglasses	89.06

-- What are the 3 most purchased product within each category

Category	Item_Purchased	Purchase_Count
Accessories	Sneakers	955
Accessories	Blouse	919
Accessories	Pants	917
Clothing	Sweater	933
Clothing	Handbag	927
Clothing	Jeans	924
Footwear	Pants	939
Footwear	Shirt	934
Footwear	Sandals	923
Outerwear	Jeans	922
Outerwear	Shirt	917
Outerwear	Sneakers	912

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

Buyer_Type	Subscription	Total_Customers	Percentage
Non-Repeat Buyer	No	3584	70.07
Non-Repeat Buyer	Yes	1531	29.93
Repeat Buyer	No	31456	70.08
Repeat Buyer	Yes	13429	29.92

-- what is the revenue contribution of each age group ??

Age_Group	Total_Revenue	Revenue_Contribution_Percentage
56+	33789792.38	26.77
26-35	24633289.03	19.52
36-45	24630486.96	19.52
46-55	23969913.62	18.99
18-25	19181963.11	15.2

Customer Shopping Behavior Analysis

Total Revenue
$126.21M

Average Ratings
3.00

Total Customers
50K

6. Business Recommendations

- **Boost Subscriptions** – Promote exclusive benefits for subscribers.

- **Customer Loyalty Programs** – Reward repeat buyers to move them into the "Loyal" segment.

- **Review Discount Policy** – Balance sales boosts with margin control.

- **Product Positioning** – Highlight top-rated and best-selling products in campaigns.

- **Targeted Marketing** – Focus efforts on high-revenue age groups and express-shipping users.