Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 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: 3,900
- Columns: 18
- 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)
- Missing Data: 37 values in Review Rating column

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	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnightly
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnightly
3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Weekly
4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	PayPal	Weekly
5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	PayPal	Annually
6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Standard	Yes	Yes	14	Venmo	Weekly
7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Free Shipping	Yes	Yes	49	Cash	Quarterly
8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2	Yes	Free Shipping	Yes	Yes	19	Credit Card	Weekly
9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes	Express	Yes	Yes	8	Venmo	Annually
10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8	Yes	2-Day Shipping	Yes	Yes	4	Cash	Quarterly
11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1	Yes	Store Pickup	Yes	Yes	26	Bank Transfer	Bi-Weekly
12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9	Yes	Store Pickup	Yes	Yes	10	Bank Transfer	Fortnightly
13	61	Male	Coat	Outerwear	72	Delaware	M	Gold	Winter	4.5	Yes	Express	Yes	Yes	37	Venmo	Fortnightly
14	65	Male	Dress	Clothing	51	New Hampshire	M	Violet	Spring	4.7	Yes	Express	Yes	Yes	31	PayPal	Weekly
15	64	Male	Coat	Outerwear	53	New York	L	Teal	Winter	4.7	Yes	Free Shipping	Yes	Yes	34	Debit Card	Weekly
16	64	Male	Skirt	Clothing	81	Rhode Island	M	Teal	Winter	2.8	Yes	Store Pickup	Yes	Yes	8	PayPal	Monthly
17	25	Male	Sunglasses	Accessories	36	Alabama	S	Gray	Spring	4.1	Yes	Next Day Air	Yes	Yes	44	Debit Card	Bi-Weekly
18	53	Male	Dress	Clothing	38	Mississippi	XL	Lavender	Winter	4.7	Yes	2-Day Shipping	Yes	Yes	36	Venmo	Quarterly
19	52	Male	Sweater	Clothing	48	Montana	S	Black	Summer	4.6	Yes	Free Shipping	Yes	Yes	17	Cash	Weekly
20	66	Male	Pants	Clothing	90	Rhode Island	M	Green	Summer	3.3	Yes	Standard	Yes	Yes	46	Debit Card	Bi-Weekly
21	21	Male	Pants	Clothing	51	Louisiana	M	Black	Winter	2.8	Yes	Express	Yes	Yes	50	Cash	Every 3 Months
22	31	Male	Pants	Clothing	62	North Carolina	M	Charcoal	Winter	4.1	Yes	Store Pickup	Yes	Yes	22	Debit Card	Quarterly
23	56	Male	Pants	Clothing	37	California	M	Peach	Summer	3.2	Yes	Store Pickup	Yes	Yes	32	Debit Card	Annually
24	31	Male	Pants	Clothing	88	Oklahoma	XL	White	Winter	4.4	Yes	Express	Yes	Yes	40	Credit Card	Weekly
25	18	Male	Jacket	Outerwear	22	Florida	M	Green	Fall	2.9	Yes	Store Pickup	Yes	Yes	16	Debit Card	Weekly
26	18	Male	Hoodie	Clothing	25	Texas	M	Silver	Summer	3.6	Yes	Express	Yes	Yes	14	PayPal	Annually
27	38	Male	Jewelry	Accessories	20	Nevada	M	Red	Spring	3.6	Yes	Next Day Air	Yes	Yes	13	Credit Card	Annually
28	56	Male	Shorts	Clothing	56	Kentucky	L	Cyan	Summer	5	Yes	Next Day Air	Yes	Yes	7	Bank Transfer	Every 3 Months
29	54	Male	Handbag	Accessories	94	North Carolina	M	Gray	Fall	4.4	Yes	Free Shipping	Yes	Yes	41	PayPal	Every 3 Months
30	31	Male	Dress	Clothing	48	Wyoming	S	Black	Fall	4.1	Yes	Store Pickup	Yes	Yes	14	Credit Card	Weekly
31	57	Male	Jewelry	Accessories	31	North Carolina	L	Black	Winter	4.7	Yes	Standard	Yes	Yes	16	Credit Card	Monthly
32	33	Male	Dress	Clothing	79	West Virginia	L	Brown	Winter	4.7	Yes	Store Pickup	Yes	Yes	45	Venmo	Monthly
33	36	Male	Jacket	Outerwear	67	Kansas	M	Silver	Summer	4.9	Yes	Free Shipping	Yes	Yes	37	Venmo	Annually
34	54	Male	Pants	Clothing	38	Colorado	L	Green	Summer	3.3	Yes	Store Pickup	Yes	Yes	45	Cash	Quarterly
35	36	Male	T-shirt	Clothing	91	North Dakota	L	Violet	Spring	4.6	Yes	2-Day Shipping	Yes	Yes	38	PayPal	Quarterly

- **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 MySQL and loaded the cleaned DataFrame into the database for SQL analysis.

4. Data Analysis using SQL (Business Transactions)

We performed structured analysis in MySQL 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.

	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 00:00:0

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

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. **ubscribers 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

6. **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

7. **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

8. **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

9. **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

10. **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

- **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.