Customer Shopping Behaviour Analysis

1. Project Overview

During my Project I Analyzed customer shopping behaviour using transactional data to identify purchasing patterns, sales trends, and high-value customers. Used SQL and Python for data cleaning and analysis, and built interactive Power BI dashboards to visualize KPIs and customer segmentation. The insights support data-driven marketing and inventory decision-making.

2. Dataset Summary

- Dataset Size: 3,900 rows with 18 Columns
- Data Coverage:
  - Customer profile information: (Age, Gender, Location, and Subscription Status)
  - Transaction information: (Purchased Item, Product Category, Order Value, Season, Size, and Colour)
  - Behavioural indicators: (Discount Usage, Promo Code Usage, Purchase Frequency, Past Purchases, Review Ratings, and Shipping Preferences)
- Data Quality Notes: Minor missing values observed in the Review Rating field (37rows)

3. Exploratory Data Analysis using Python

In this project I Began data preparation and cleaning in python:

- **Data Loading:** Imported the dataset using Pandas
- **Initial Exploration:** Use df.info( ) to check structure and describe( ) for summary statistics

	customer_id	age	gender	item_purchased	category	purchase_amount	location	size	color	season	review_rating	subscription_status	shipping_type	discount_applied
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN

previous_purchases	payment_method	frequency_of_purchases	age_group	purchase_frequency_days
3900.000000	3900	3900	3900	3900.000000
NaN	6	7	4	NaN
NaN	PayPal	Every 3 Months	Young Adult	NaN
NaN	677	584	1028	NaN
25.351538	NaN	NaN	NaN	89.133077
14.447125	NaN	NaN	NaN	119.037566
1.000000	NaN	NaN	NaN	7.000000
13.000000	NaN	NaN	NaN	14.000000
25.000000	NaN	NaN	NaN	30.000000
38.000000	NaN	NaN	NaN	90.000000
50.000000	NaN	NaN	NaN	365.000000

- **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)

Performed Structures analysis in MySQL to answer key business questions:

1) **Revenue by Gender –** Compared total revenue generated by male vs. female customers.

	gender	total_revenue
▶	Male	157890
	Female	75191

2) **High-Spending Discount Users** – Identified customers who used discounts but still spent above the average purchase amount.

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90
22	62
24	88
29	94
32	79
33	67
35	91
37	69
40	60
41	76

3) **Top 5 Products by Rating** – Found products with the highest average review ratings.

item_purchased	Avergae_product_rating
Sandals	3.82
Boots	3.79
Gloves	3.78
Hat	3.78
Handbag	3.78

4) **Shipping Type Comparison** – Compared average purchase amounts between Standard and Express shipping.

shipping_type	Avg_Purchase_Amount
Express	60.48
Standard	58.46

5) **Subscribers vs. Non-Subscribers** – Compared average spend and total revenue across subscription status.

subscription_status	Total_customer	Avg_spend	Total_revenue
Yes	1053	59.49	62645
No	2847	59.87	170436

6) **Discount-Dependent Products** – Identified 5 products with the highest percentage of discounted purchases.

item_purchased	discount_rate
Hat	50.00
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37

7) **Customer Segmentation** – Classified customers into New, Returning, and Loyal segments based on purchase history.

customer_segment	Number of customers
Loyal	3116
Returning	701
New	83

8) **Top 3 Products per Category** – Listed the most purchased products within each category.

item_rank	category	item_purchased	total_orders
1	Accessories	Jewelry	171
2	Accessories	Sunglasses	161
3	Accessories	Belt	161
1	Clothing	Blouse	171
2	Clothing	Pants	171
3	Clothing	Shirt	169
1	Footwear	Sandals	160
2	Footwear	Shoes	150
3	Footwear	Sneakers	145
1	Outerwear	Jacket	163
2	Outerwear	Coat	161

9) **Repeat Buyers & Subscriptions** – Checked whether customers with >5 purchases are more likely to subscribe.

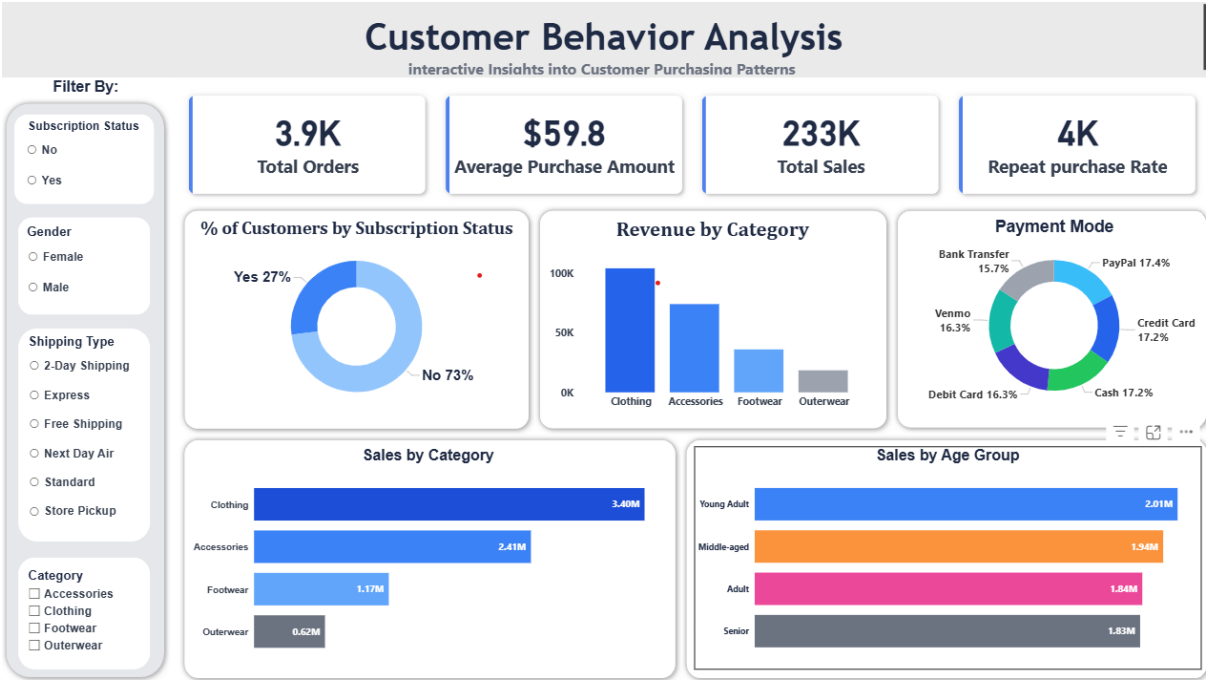
subscription_status	repeat_buyers
Yes	958
No	2518

10) **Revenue by Age Group** – Calculated total revenue contribution of each age group.

age_group	revenue_contribution
Young Adult	62143
Middle-aged	59197
Adult	55978
Senior	55763

5. Dashboard in Power BI

➢ Based on the results Built an interactive dashboard in Power BI to present insights visually.



6. Business Recommendations

I. Increase Focus on High-Revenue Categories

- **Recommendation:**
  Allocate more marketing and inventory resources to Clothing and Accessories, as they generate the highest sales and revenue.
- **Business Impact:**
  Improves revenue growth and reduces stock-out risk for top-performing products.

II. Target Young Adult Customers with Personalized Campaigns

- **Recommendation:**
  Design targeted promotions and loyalty programs for the Young Adult age group, which contributes the highest sales.
- **Business Impact:**
  Increases customer engagement, conversion rates, and lifetime value.

III. Improve Subscription Conversion Strategy

- **Recommendation:**
  Encourage non-subscribed customers to join subscription plans through exclusive discounts, early access, or free shipping benefits.
- **Business Impact:**
  Boosts recurring revenue and improves customer retention.

IV. Optimize Underperforming Categories

- **Recommendation:**
  Re-evaluate pricing, promotions, and product assortment for Footwear and Outerwear, which show lower sales contribution.
- **Business Impact:**
  Helps identify growth opportunities and prevents revenue stagnation.

V. Maintain Multiple Payment Options

- **Recommendation:**
  Continue supporting multiple payment modes, as customers show balanced usage across cards, digital wallets, and cash.
- **Business Impact:**
  Enhances customer convenience and reduces cart abandonment.