CUSTOMER SHOPPING BEHAVIORS ANALYSIS

1. Executive Summary

This project analyzed customer behaviors using transactional data from 3900 records of transactions across multiple product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences and subscription behaviors to guide strategic business decisions.

2. Data Understanding

- Source: Dataset was sourced from a publicly available Github repository
- Rows: 3900
- Columns: 18
- Key Features:
    - Customer Demographics (Age, Gender, Location, Subscription Status)
    - Purchase Details (Item Purchase, Category, Purchase, Size, Color, Season)
    - Shopping Behaviors (Review Rating, Shipping Type, Discount Applied, Promo Code Used, Previous Purchases, Payment Method, Frequency of Purchase)

3. Data Cleaning and Preparation

- **Handling Missing Values**: Missing values in the *review_rating* column were imputed using the median rating within each product category. The median was chosen instead of the mean to minimise the influence of outliers and maintain realistic customer sentiment patterns within each category.

- **Column Name Standalization**: Renamed column names to snake case for better readability and documentation

- **Featured Engineering**:
    - A new Columns *age_group* was created using quantile-based binning (qcut). Customers were categorized into 4 main groups (Young Adult, Adult, Middle-Aged, Senior). This segmentation enables clearer visualization and more meaningful comparison in revenue contribution

    - To facilitate quantitative analysis and comparisons, a new column, *purchase_frequency_days*, was created by converting the categorical values in *frequency_of_purchases* into numerical values representing the approximate number of days between purchases. This enables segmentation of customers (e.g., new, returning, loyal) and analysis of their contribution to revenue.

- **Data Consistency:** For consistency and clarity in the dataset, the *promo_code_used* column was dropped, as its values were redundant with the *discount_applied* column. Removing this duplicate column reduces potential confusion and streamlines analysis of promotional impact on customer spending.

4. Exploratory Data Analysis (EDA) using MySQL

1. **Revenue by Gender**: Male customers generated roughly twice the total revenue as females made, highlighting the disproportionate contribution of female customers to the total revenue.

GENDER	total_revenue
Male	157890
Female	75191

2. **Revenue by Age Group**: Customers in the Young Adult and Middle-Aged segments account for the largest share of total revenue, highlighting them as key target groups for marketing and promotional strategies. Efforts to engage these segments through personalized offers or loyalty programs are likely to have the greatest impact on revenue growth.

REV	AGE_GROUP
55763	Senior
55978	Adult
59197	Middle-aged
62143	Young Adult

3. **High-spending discount User**: Customers who applied promo codes but still spent above the average purchase amount represent high-value buyers who are motivated by product preference rather than discounts.

	CUSTOMER_ID	DISCOUNT_APPLIED	PURCHASE_AMOUNT
▶	2	Yes	64
	3	Yes	73
	4	Yes	90
	7	Yes	85
	9	Yes	97
	12	Yes	68
	13	Yes	72
	16	Yes	81
	20	Yes	90
	22	Yes	62
	24	Yes	88
	29	Yes	94
	32	Yes	79
	33	Yes	67
	35	Yes	91

4. **Subscriber vs Non-Subscriber:** Although subscribed customers make up less than one-third of the total customer base, they spend nearly as much per transaction as non-subscribers, indicating that subscribers are high-value customers. Meanwhile, non-subscribers, who represent the majority of customers, contribute nearly three-quarters of total revenue.

	SUBSCRIPTION_STATUS	TOTAL_CUSTOMERS	AVG_SPEND	TOTAL_REVENUE
▶	Yes	1053	59.4919	62645
	No	2847	59.8651	170436

5. Customer segmentation: Loyal customers make up nearly 80% of the total base, underscoring the importance of retention strategies. Returning customers represent a key growth opportunity, while the low proportion of new customers highlights the need for targeted acquisition campaigns.

	CUSTOMER_SEGMENT	CUSTOMER_COUNT
▶	LOYAL	3116
	RETURNING	701
	NEW	83

6. **Repeat buyers:** Most customers with more than five prior purchases are not subscribed.

	SUBSCRIPTION_STATUS	COUNT(*)
	Yes	958
▶	No	2518

7. **Top 5 Products with highest review rating:** Products such as Gloves, Sandals, Boots, Hat, and Handbag received the highest average review ratings, reflecting strong customer satisfaction.

	TOP_5_PRODUCTS	AVG_REVIEW_RATING
	Gloves	3.86
	Sandals	3.84
	Boots	3.82
	Hat	3.8
▶	Handbag	3.78

8. **Top 3 most-purchased Products within each category**:

	RNK	CATEGORY	ITEM_PURCHASED	TOTAL_CUSTOMERS
▶	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

5. Insights/ Recommendations

1. Target high-value male customers with personalized promotions to further increase revenue.

2. Focus marketing campaigns on Young Adult and Middle-Aged segments for maximum impact.

3. Upsell or offer VIP programs to customers who spend above average despite using promo codes.

4. Run campaigns to convert non-subscribers into subscribers to boost revenue and loyalty.

5. Strengthen retention for loyal customers and create acquisition campaigns for new and returning customers.

6. Offer subscription incentives to repeat buyers to convert them into subscribers.

7. Promote highly-rated products through marketing or bundle offers to drive sales.

8. Prioritize inventory, promotions, and cross-sell opportunities for top-selling items in each category.