

# Customer Shopping Behavior Analysis

## 1. Project Overview

This project analyze customer shopping behavior using transactional data from 3900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segment, product preference, and subscription behavior to guide strategic business decisions.

---

## 2. Data Summary

- Row : 3900
  - Columns : 18
  - Key Feature:
    - Customer Demographic ~ Age, Gender, Location, Subscription Status
    - Purchase Detail ~ Item Purchased, Category, Purchase Amount, Season, Size, Color
    - Shopping Behavior ~ Discount Applied, Promo Code Used, Previous Purchase Frequency of Purchase, Review Rating, Shipping Type
    - Missing Data ~ 37 values in Review Rating column
-

### 3. Exploratory Data Analysis using Python

We began with the data preparation and data cleaning using pandas:

- Data Loading: Imported data using pandas
- Initial Exploration : Used [`df.info\(\)`](#) to check the structure and [`df.describe\(\)`](#)

	Customer ID	Age	Purchase Amount (USD)	Review Rating \
count	3900.000000	3900.000000	3900.000000	3863.000000
mean	1950.500000	44.068462	59.764359	3.750065
std	1125.977353	15.207589	23.685392	0.716983
min	1.000000	18.000000	20.000000	2.500000
25%	975.750000	31.000000	39.000000	3.100000
50%	1950.500000	44.000000	60.000000	3.800000
75%	2925.250000	57.000000	81.000000	4.400000
max	3900.000000	70.000000	100.000000	5.000000
Previous Purchases				
count	3900.000000			
mean	25.351538			
std	14.447125			
min	1.000000			
25%	13.000000			
50%	25.000000			
75%	38.000000			
max	50.000000			

- **Missing Data Handling:** Checked for the null values in our data and imputed the missing values in the Review Rating column using median rating of each product category.
- **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 clean DataFrame into database for SQL Analysis.

## 4.Data Analysis Using MySQL

We performed structured analysis in MySQL to answer the key Business question:

1. **Revenue By Gender:** Compared total revenue generated by male vs female customer

	gender	revenue
▶	Male	157890
	Female	75191

2. **High-Spending Discount User:** Identified customers who discounts but still spent above the average Purchase Amount.

	customer id	purchase amount (USD)
▶	43	100
	96	100
	194	100
	205	100
	244	100
	249	100
	456	100
	519	100
	582	100
	616	100
	770	100
	862	100
	1209	100
	1301	100
	1406	100
	1413	100
	1422	100
	1457	100
	1480	100
	1592	100

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

	Item Purchased	Average Product Rating
▶	Gloves	3.8627737226277383
	Sandals	3.8446540880503144
	Boots	3.818881118881119
	Hat	3.801307189542483
	Skirt	3.7853503184713366

4. **Shipping Type Comparison:** Compared average purchase amount between Standard and Express Shipping.
5. **Subscribers Vs Non-Subscribers:** Compared the spend and total revenue across the all subscription Status.

	Shipping Type	avg(`purchase amount (USD)`)
▶	Express	60.4752
	Standard	58.4602

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 customer into New, Returning and Loyal segment based on purchase history.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

8. **Top 3 Categories per Category:** Listed the most purchased products within each category

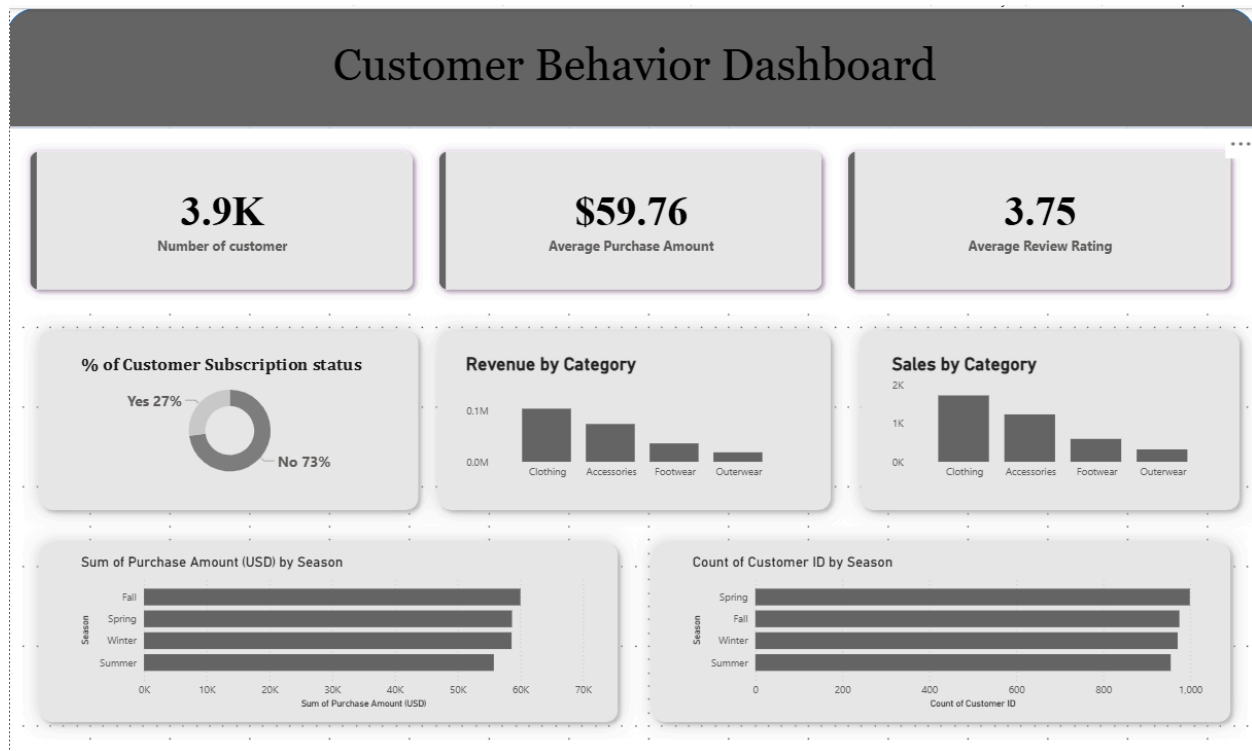
	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. **Repeat Buyers & Subscription:** Checked whether customer with > 5 purchases are more likely to subscribe.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

## 5. Dashboard in PowerBI

Finally, We built an interactive dashboard in **PowerBI** to present insights visually.



## 6. Business Recommendations

- **Boost Subscription:** Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs:** Balance sales boosts with margin control.
- **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.