Customer Shopping Behavior Analysis

1. Project Overview

This project focuses on deriving meaningful business insights from an e-commerce customer behavior dataset using **MySQL**. The objective is to analyze customer activity, purchasing patterns, revenue trends, and product performance to support data-driven business decisions.

---

2. Dataset Description

The dataset represents customer shopping behavior and includes the following key attributes: - **Customer ID** – Unique identifier for each customer - **Age / Gender** – Demographic details - **Category** – Product category (Accessories, Clothing, Footwear, etc.) - **Item Purchased** – Product name - **Purchase Amount** – Transaction value - **Purchase Frequency** – Number of purchases - **Subscription Status** – Yes / No - **Review Rating** – Customer feedback score - **Payment Method & Location** – Transaction details

The dataset is assumed to be stored in a MySQL database table.

---

3. Tools & Technologies

- **Database:** MySQL
- **Language:** SQL,Python
- **Visualization:** Power BI
- **Version Control:** GitHub

4. 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

[3]:

	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	P
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	

Previous Purchases	Payment Method	Frequency of Purchases
14	Venmo	Fortnightly
2	Cash	Fortnightly
23	Credit Card	Weekly
49	PayPal	Weekly
31	PayPal	Annually

● 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.

5. Business Questions Addressed

The analysis answers the following business-oriented questions:

1. How many total customers are present in the dataset?

gender	revenue
Male	157890
Female	75191

2. What is the average purchase amount per customer?

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97

3. Which product categories generate the see revenue?

item_purchased	avg_review_rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Handbag	3.78

4. What are the top-selling categories by sales volume?

shipping_type	avg(purchase_amount)
Express	60.4752
Standard	58.4602

5. How does subscription status impact customer behavior?

subscription_status	total_customer	avg_spend	total_revenue
Yes	1053	59.49	62645
No	2847	59.87	170436

6. What is the average review rating across categories?

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

7. Which products or categories contribute the least revenue?

customer_segment	Number of customer
Loyal	3116
Returning	701
New	83

8. How does gender-based purchasing behavior differ?

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

9. Which customers show high purchase frequency?

	subscription_status	repeat_buyers
▶	Yes	958
	No	2518

10. What insights can help improve business performance?

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

---

6. Key SQL Concepts Used

- `SELECT`, `WHERE`, `ORDER BY`
- Aggregate functions (`SUM`, `AVG`, `COUNT`)
- `GROUP BY` and `HAVING`
- Subqueries
- Common Table Expressions (CTEs)
- Conditional logic using `CASE WHEN`
- `Window Function ROW_NUMBER() OVER() PARTITION()`

---

7. Sample Insights

- **Customer Base:** The business serves several thousand customers, indicating a healthy user base.
- **Revenue Concentration:** A small number of product categories contribute the majority of total revenue.
- **Subscription Impact:** Subscribed customers tend to have higher purchase frequency and value.
- **Category Performance:** Clothing and Accessories outperform other categories in both revenue and sales count.
- **Customer Satisfaction:** Average review ratings remain above average, indicating good customer satisfaction.

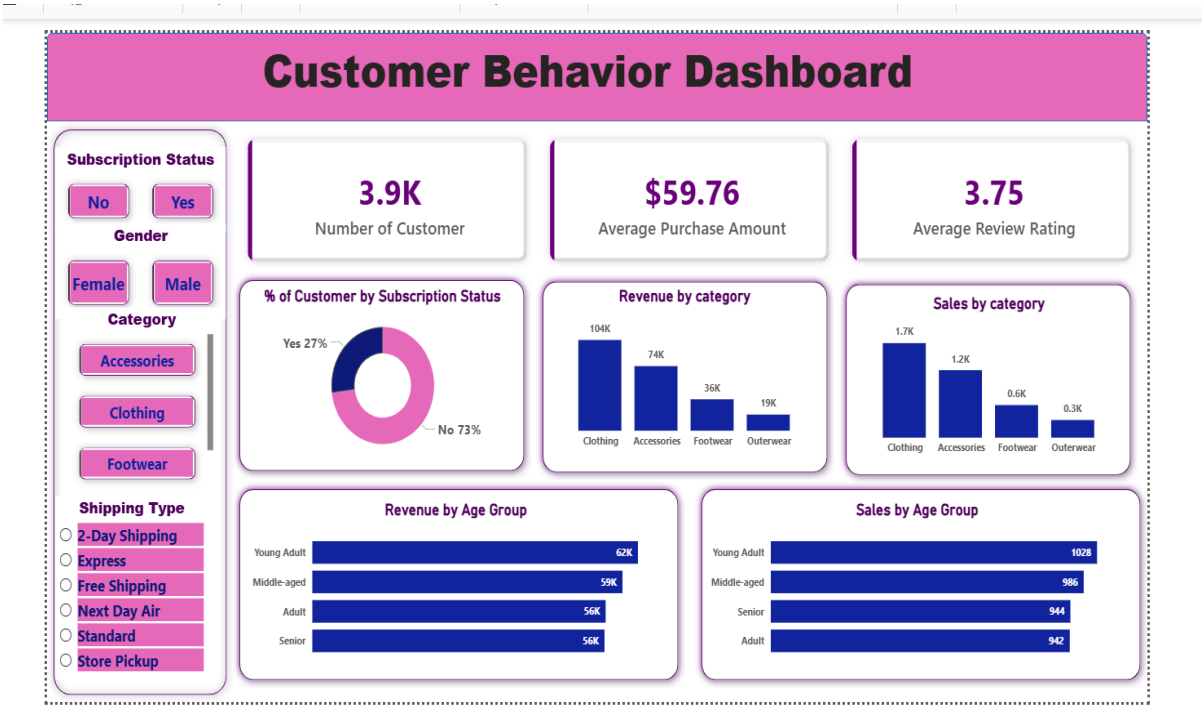
---

8. Business Recommendations

- Focus marketing efforts on **high-revenue categories**.
- Introduce offers to convert non-subscribers into subscribers.
- Improve underperforming categories through pricing or promotions.
- Use customer ratings to identify products needing quality improvement.

---

9. Dashboard in Power BI



10. Conclusion

This SQL-based analysis demonstrates how structured queries in MySQL can uncover valuable business insights from raw transactional data. The project highlights strong analytical skills, database proficiency, and the ability to translate data into actionable business recommendations.

---

11. Author

Satyam Pandey
B.Tech – Electronics & Communication Engineering
SQL & Data Analytics Enthusiast