Customer Shopping Behavior Analysis

1. Project Overview

In this project, I analyzed customer shopping behavior using transactional data of **3,900 purchases** across multiple product categories. The main objective was to understand **spending patterns, customer segments, product preferences, and subscription behavior** to support better business decisions.

---

2. Dataset Summary

- **Total Records:** 3,900
- **Total Features:** 18
- **Key Data Included:**
  - Customer details: Age, Gender, Location, Subscription Status
  - Purchase details: Item Purchased, Category, Purchase Amount, Season, Size, Color
  - Behavior metrics: Discounts, Promo Codes, Previous Purchases, Purchase Frequency, Review Ratings, Shipping Type
- **Missing Values:** 37 missing entries in the *Review Rating* column

---

3. Exploratory Data Analysis (Python)
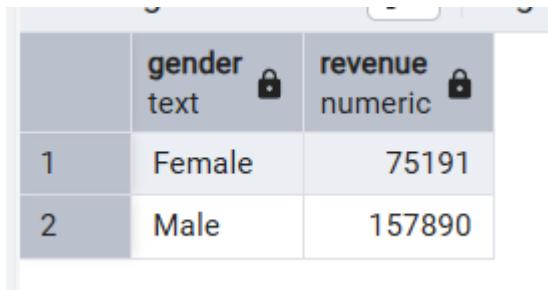
The analysis started with data cleaning and preparation using Python:

- Loaded data using **pandas** and performed initial inspection using `info()` and `describe()`.
- Handled missing values by imputing the **median review rating per product category**.
- Standardized column names into **snake_case** for consistency.
- Performed feature engineering:
  - Created **age_group** buckets
  - Derived **purchase_frequency_days**
- Checked redundancy between *discount_applied* and *promo_code_used* and removed the promo code column.
- Stored the cleaned dataset into **PostgreSQL** for SQL-based analysis.

---

4. Business Analysis using SQL

Using PostgreSQL, several business-focused queries were performed:

1. **Revenue by Gender** – Male customers generated higher total revenue compared to 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 value.

	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
12	29	94
13	32	79
14	33	67
15	35	91
16	37	69

3. **Top 5 Products by Rating** – Gloves, Sandals, and Boots ranked highest based on average customer 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** – Express shipping users had a slightly higher average purchase amount than standard shipping users.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

5. **Subscribers vs Non-Subscribers** – Non-subscribers contributed higher total revenue, while average spend remained similar.

	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** – Products like Hats and Sneakers had 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** – Customers were categorized as New, Returning, and Loyal, with the majority falling into the Loyal segment.

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

8. **Top Products per Category** – Identified top 3 most purchased products in 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** – Customers with more than 5 purchases showed higher subscription adoption.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. **Revenue by Age Group** – Young Adults contributed the highest revenue overall.
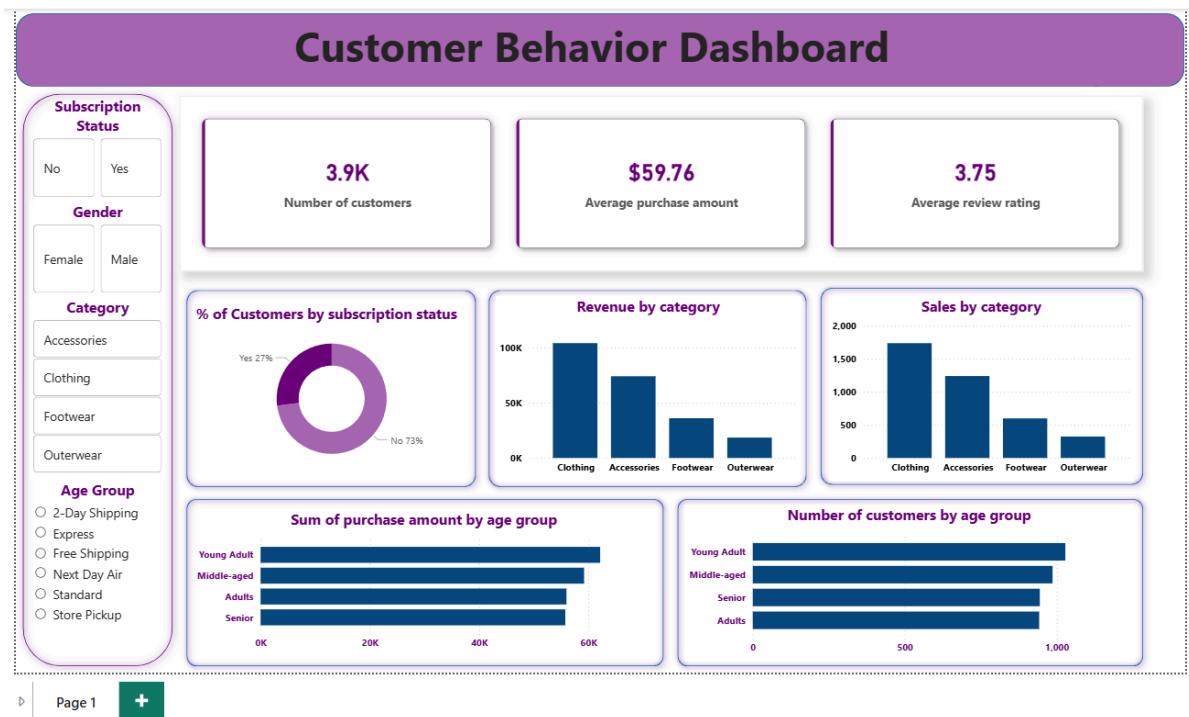
|	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adults	55978
4	Senior	55763

5. Power BI Dashboard

An interactive **Power BI dashboard** was created to visualize:

- Total customers and average purchase value
- Revenue and sales by category
- Subscription split
- Revenue and sales by age group

This dashboard helps stakeholders quickly understand customer behavior and performance trends.

6. Business Recommendations

- **Increase Subscriptions:** Offer exclusive deals and benefits to convert repeat buyers.
- **Customer Loyalty Programs:** Reward loyal customers to improve retention.
- **Optimize Discount Strategy:** Control discounts on highly dependent products to protect margins.
- **Product Promotion:** Highlight top-rated and best-selling products in marketing campaigns.
- **Targeted Marketing:** Focus on high-revenue age groups and express-shipping users for better ROI.

Tools Used: Python (Pandas), PostgreSQL, Power BI