Customer Shopping Behaviour Analysis

1. Project Overview

This project analyses customer shopping behaviour using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behaviour to guide strategic business decisions.

2. Dataset Summary

- Rows : 3,900
- Columns : 18
- Key Features :
  - Customer Demographics : Age, Gender, Location, and Subscription Status.
  - Purchase Details : Item Purchased, Category, Purchase Amount, Season, Size, and Colour.
  - Shopping Behaviour : Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, and Shipping Type.
- Missing Data : 37 values in Review Rating column

3. Data Cleaning and Feature Engineering using Python

- **Data Loading** : We began by importing the dataset using **pandas**.
- **Initial Exploration** : Then we used **df.info()** to check structure of the dataset.
- **Removing unnecessary whitespaces** : We separated string columns and used **strip** to remove any unnecessary leading, trailing or double whitespaces.
- **Check Duplicates** : Then we checked for duplicate records.
- **Check Missing Values** : Checked null values and **imputed** null values in Review Rating column using **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.
  - Create **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 database for SQL analysis.

4. Data Analysis using Python and PostgreSQL

Performed structured analysis in PostgreSQL and Python to answer key business questions:

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

	gender text	sum numeric
1	Female	75191
2	Male	157890

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

	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

Total rows: 839 Query complete 00:00:00.089

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

	item_purchased text	average_review_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** – Compared average purchase amounts between Standard and Express shipping.

	shipping_type text	avg_purchase_amount numeric
1	Express	60.48
2	Standard	58.46

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

	subscription_status text	avg_amount_spent numeric	total_revenue numeric
1	No	59.8651211801896733	170436
2	Yes	59.4919278252611586	62645

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	49.00
3	Coat	49.00
4	Sweater	48.00
5	Pants	47.00

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

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

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

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessori...	Jewelry	171
2	2	Accessori...	Sunglasses	161
3	3	Accessori...	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** – Checked whether customers with >5 purchases are more likely to subscribe.

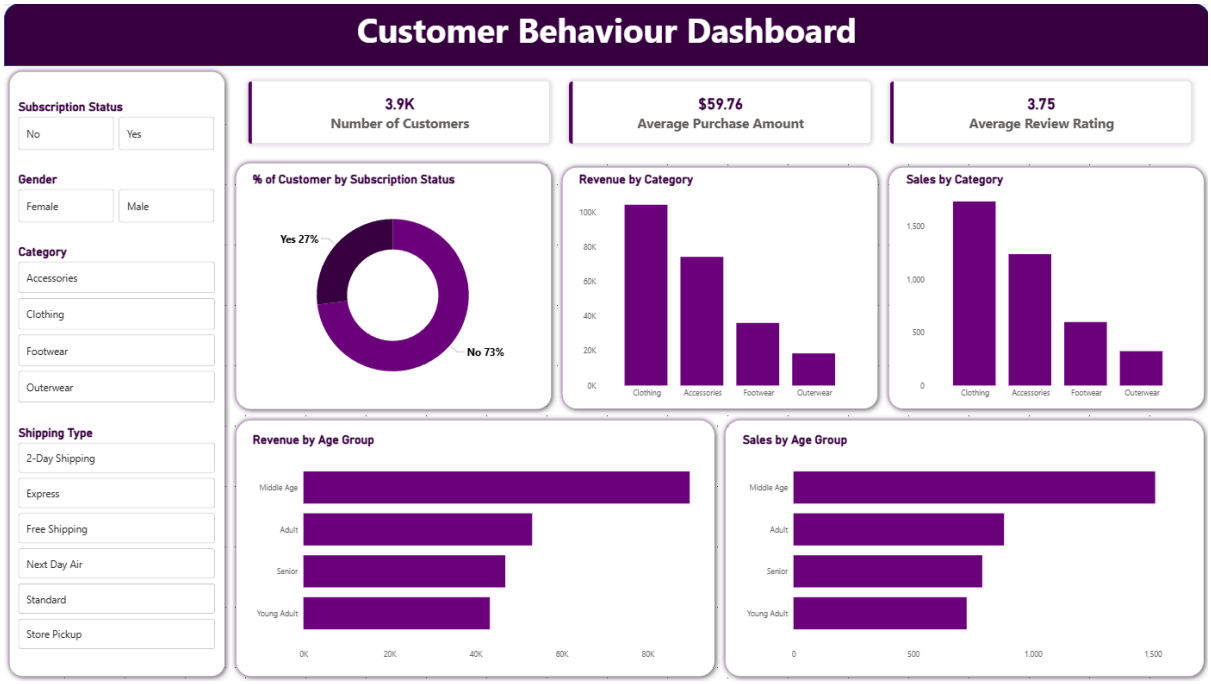
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

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

	age_group text	total_revenue numeric
1	Middle Age	89741
2	Adult	53144
3	Senior	46894
4	Young Adult	43302

5. Dashboard in Power BI

Finally, built an interactive dashboard in Power BI to present insights visually.



6. Business Recommendations

- **Boost Subscriptions :** Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs :** Reward repeat buyers to move them into the “Loyal” segment.
- **Review Discount Policy :** Balance sales boosts with margin control.
- **Product Positioning :** Highlight top-rate and best-selling products in campaigns.
- **Targeted marketing :** Focus efforts on high-revenue age groups and express-shipping users.