RETAIL CUSTOMER BEHAVIOR ANALYSIS

1. Project Overview

This project looks into retail customer shopping behavior using data from 3,900 purchases across different product categories. The aim is to uncover insights about spending habits, customer groups, product preferences, and subscription behavior to help guide business decisions.

2. Dataset Summary

- Rows: 3900
- Key Features:
  - Customer information (Age, Gender, Location, Subscription Status)
  - Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Colour)
  - Shopping behaviour (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 missing values in the Review Rating column

3. Exploratory Data Analysis using Python

We started by preparing and cleaning the data in Python:

- Data Loading: We imported the dataset using pandas.

- Initial Exploration: Used **df.info()** to check the structure and **describe()** for summary statistics.

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN

Previous Purchases	Payment Method	Frequency of Purchases
3900.000000	3900	3900
NaN	6	7
NaN	PayPal	Every 3 Months
NaN	677	584
25.351538	NaN	NaN
14.447125	NaN	NaN
1.000000	NaN	NaN
13.000000	NaN	NaN
25.000000	NaN	NaN
38.000000	NaN	NaN
50.000000	NaN	NaN

- **Handling Missing Data:** Checked for missing values and filled in the missing Review Rating data with the median rating of each product category.
- **Column Standardisation:** Renamed columns to snake case for clearer reading and better documentation.
- **Feature Engineering:**
    - Created an **age_group** column by grouping customer ages.
    - Created a **purchase_frequency_days** column from purchase data.
- **Data Consistency Check:** Verified whether the columns **discount_applied** and **promo_code_used** were redundant and removed **promo_code_used**.
- **Database Integration:** Connected the Python script to PostgreSQL and loaded the cleaned data into the database for SQL analysis.

4. Exploratory Data Analysis using Python

a. **Revenue by Gender** – Compared total revenue from male and female customers.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

b. **High-Spending Discount Users** – Identified customers who used discounts but still spent more than the average 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

Total rows: 839 Query complete 00:00:00.090

c. **Top 5 Products by Rating** – Found the top-rated products based on average review scores.

	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

d. **Shipping Type Comparison** – Compared average purchase amounts for Standard and Express shipping.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

e. **Subscribers vs Non-Subscribers** – Compared the average spend and total revenue between subscribed and non-subscribed customers.

	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

f. **Discount-Dependent Products** – Identified the five 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

g. **Customer Segmentation** – Grouped customers into New, Returning, and Loyal segments based on their purchase history.

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

h. **Top 3 Products per Category** – Listed the top three 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

i. **Repeat Buyers & Subscriptions** – Checked if customers with more than five purchases are more likely to subscribe.

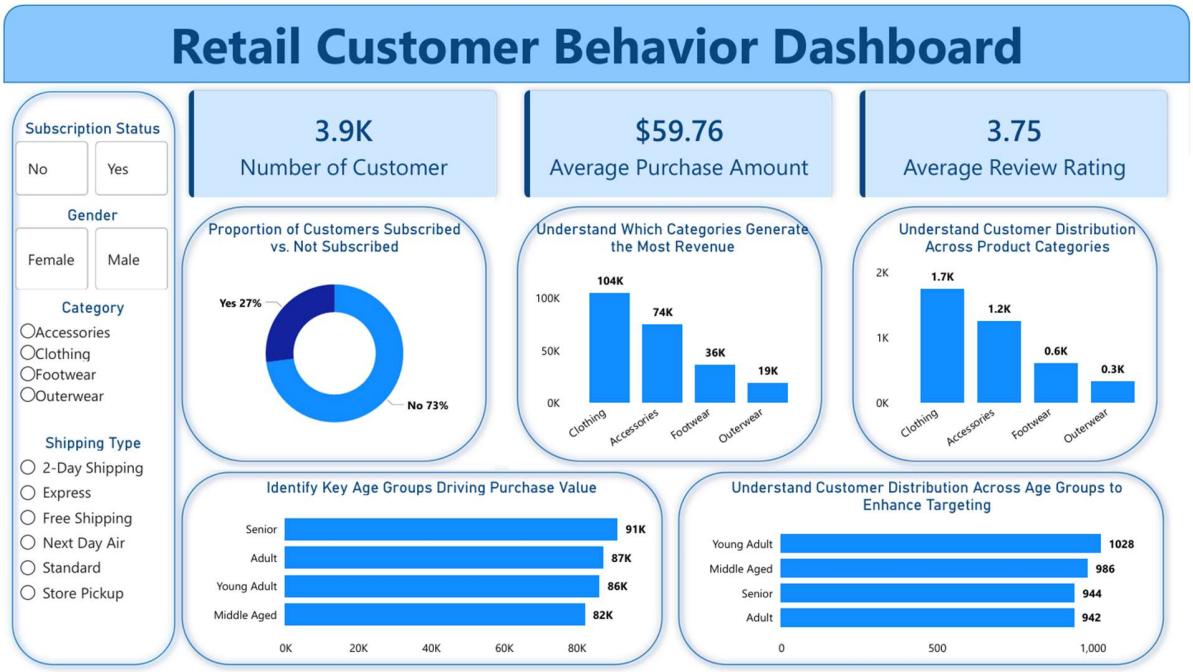
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

j. **Revenue by Age Group** – Calculated the total revenue contribution of each age group.

	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle Aged	59197
3	Adult	55978
4	Senior	55763

5. Dashboard in Power BI

Created an interactive dashboard in Power BI to display the insights visually



6. Business Recommendations

- **Boost Subscriptions**: Offer exclusive benefits for subscribers to encourage more sign-ups.
- **Customer Loyalty Programs**: Reward repeat customers to help move them into the "Loyal" segment.
- **Review Discount Policy**: Find a balance between increasing sales through discounts and controlling profit margins.
- **Product Positioning**: Focus on highlighting the best-rated and best-selling products in marketing campaigns.
- **Targeted Marketing**: Focus marketing efforts on high-revenue age groups and customers who use express shipping.