Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,000 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: 3,900
- Columns: 18

Data Types:

- Numerical: Age, Income, Purchase Amount, Spending Score
- Categorical: Gender, Product Category, Payment Method
- Date/Time: Purchase Date

Important Columns Include:

- Customer ID
- Gender
- Age
- Annual Income
- Spending Score
- Product Category
- Purchase Amount
- Payment Method
- Shopping Location
- Date of Purchase

The dataset enables analysis of customer demographics, spending patterns, product popularity, and seasonal trends.

3. Exploratory Data Analysis using Python

3.1 Importing Libraries: Imported the dataset using pandas

3.2 Initial Exploration: Used df. Info () to check structure and .describe( ) for summary Statistics

```
[2]: import pandas as pd
     df= pd.read_csv('customer_shopping_behavior.csv')

[3]: df.head()
```

[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	Pay M
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	V
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	I
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	I

3.3 Checking Missing Values: Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

```
df.isnull().sum()
```

```
Customer ID                 0
Age                         0
Gender                      0
Item Purchased              0
Category                    0
Purchase Amount (USD)       0
Location                    0
Size                        0
Color                       0
Season                      0
Review Rating              37
Subscription Status         0
Shipping Type               0
Discount Applied            0
Promo Code Used             0
Previous Purchases          0
Payment Method              0
Frequency of Purchases      0
dtype: int64
```

3.4 Data Consistency Check: Verified if **discount_applied** and **promo_code_used** were redundant dropped promo_code_used.

3.5 Database Integration: Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

```
pip install psycopg2-binary sqlalchemy
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: psycopg2-binary in c:\users\hp\appdata\roaming\python\python313\site-packages (2.9.11)
Requirement already satisfied: sqlalchemy in c:\programdata\anaconda3\lib\site-packages (2.0.43)
Requirement already satisfied: greenlet>=1 in c:\programdata\anaconda3\lib\site-packages (from sqlalchemy) (3.2.4)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\programdata\anaconda3\lib\site-packages (from sqlalchemy) (4.15.0)
Note: you may need to restart the kernel to use updated packages.
```

```
from sqlalchemy import create_engine
username="postgres"
password="1234"
host="localhost"
port="5432"
database="customer_behavior"
engine= create_engine(f"postgresql+psycopg2://{username}:{password}@{host}:{port}/{database}")

table_name="customer"
df.to_sql(table_name,engine,if_exists="replace", index=False)
print(f"Data successfully loaded into table '{table_name}' in database '{database}',")
```

```
Data successfully loaded into table 'customer' in database 'customer_behavior',
```

4. Data Analysis using SQL

We perform structured analysis in PostgreSQL to answer key business questions:

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

	gender text	revenue numeric
1	Female	75191
2	Male	157890

2. **High – Spending Discount Users** – Identified customers who used discount 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

3. **Top 5 Product by Rating –** Found product with the highest average review rating.

	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 –** Compare average purchase amount between Standard and Express shipping.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

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

	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** – Identified 5 products with the highest percentage of discount 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	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**- Checked whether customers with >5 purchases are more likely to subscribe.

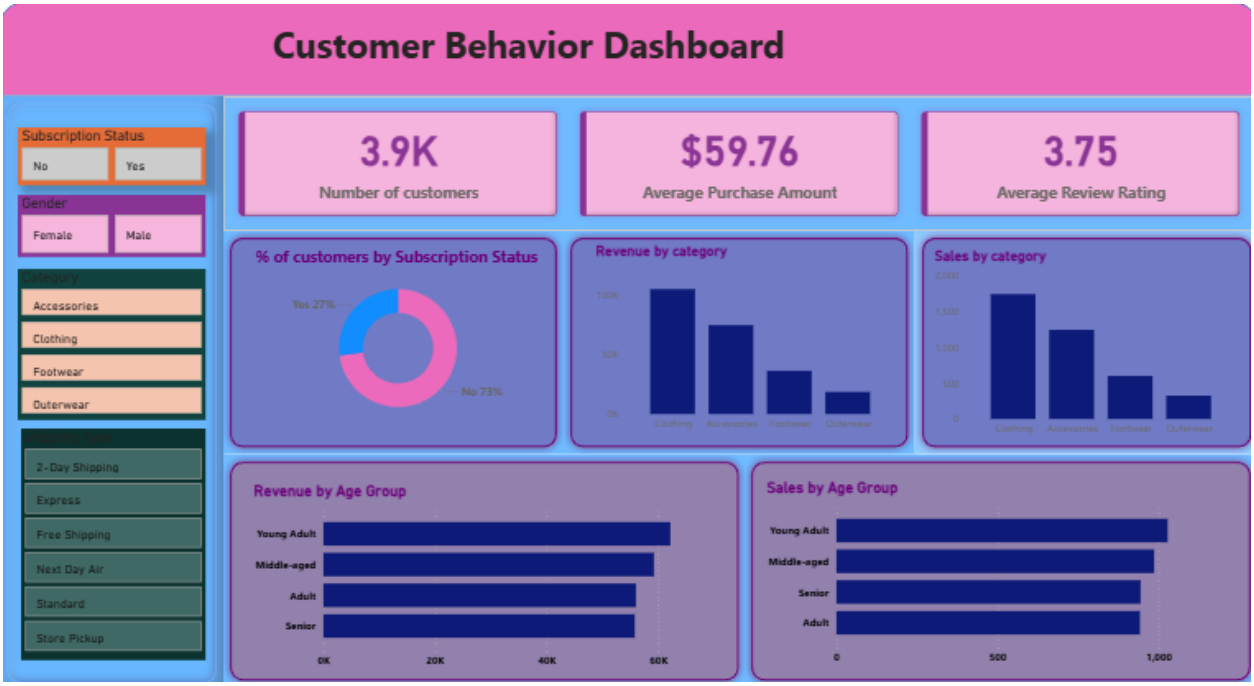
	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

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

Finally, we 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-rated and best-selling products in campaigns.
- **Targeted Marketing**- Focus efforts on high-revenue age groups and express-shipping users**.**