DATA ANALYSIS REPORT

Customer Shopping Behavior Analysis for Growth and Retention

📌 Business Problem

Retail companies collect large volumes of customer shopping data, but without proper analysis, this data is underutilized.
The business currently lacks clear visibility into customer purchasing behavior, high-value customer segments, category performance, and factors influencing revenue and retention.

As a result:

- Marketing campaigns are not fully data-driven
- High-value customers are not clearly identified
- Product and category-level decisions are inefficient

🎯 Business Objective

The objective of this project is to analyze customer shopping behavior data and generate actionable insights that help the business:

- Increase overall revenue
- Improve customer retention
- Identify high-performing product categories
- Understand demographic and seasonal purchasing patterns

❓ Key Business Questions

1. Which product categories generate the highest revenue?
2. Which customer segments contribute most to total sales?
3. Do subscription customers spend more than non-subscribers?
4. How does purchase behavior vary by age and gender?
5. Which seasons show peak shopping activity?
6. What factors influence repeat purchases?

DATA OVERVIEW

Key Features:

• Customer demographics (Age, Gender, Location, Subscription Status)

• Purchase details (Item, Category, Amount, Season, Size, Color)

• Shopping behavior (Discount, Previous Purchases, Reviews, Shipping)

• Missing Data: 37 values in Review Rating

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.

```
df.describe()
```

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	3900.000000	3900.000000	3900.000000	3863.000000	3900.000000
mean	1950.500000	44.068462	59.764359	3.750065	25.351538
std	1125.977353	15.207589	23.685392	0.716983	14.447125
min	1.000000	18.000000	20.000000	2.500000	1.000000
25%	975.750000	31.000000	39.000000	3.100000	13.000000
50%	1950.500000	44.000000	60.000000	3.800000	25.000000
75%	2925.250000	57.000000	81.000000	4.400000	38.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000

DATA CLEANING AND PREPARATION

● **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**: Loaded cleaned data into SQL database for analysis.

.

```
df[['purchase_frequency_days','frequency_of_purchases']].head(10)
```

	purchase_frequency_days	frequency_of_purchases
0	14	Fortnightly
1	14	Fortnightly
2	7	Weekly
3	7	Weekly
4	365	Annually
5	7	Weekly
6	90	Quarterly
7	7	Weekly
8	365	Annually
9	90	Quarterly
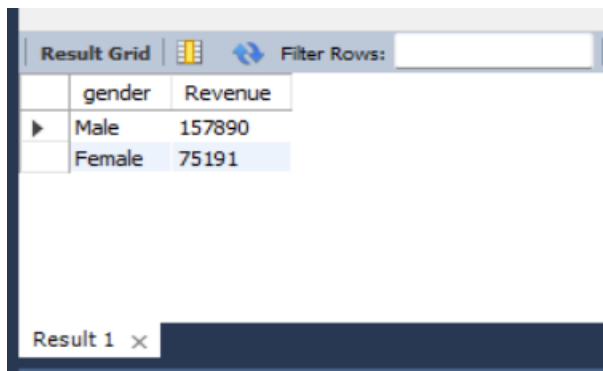
```
df[['discount_applied','promo_code_used']].head(10)
```

	discount_applied	promo_code_used
0	Yes	Yes
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	Yes
5	Yes	Yes
6	Yes	Yes
7	Yes	Yes
8	Yes	Yes
9	Yes	Yes

Data Analysis using SQL (Business Transactions)

```sql
-- Q1. What is the total revenue generated by male vs.
female customers?

select gender , sum(purchase_amount_amount)  as Revenue
 from customer
 group by gender;
```

gender	Revenue
Male	157890
Female	75191

```sql
 -- Q2. Which customers used a discount but still spent
more than the average purchase amount?
select * from customer
 where  discount_applied = 'Yes'
 and  purchase_amount_amount >=(select
avg(purchase_amount_amount) from customer);
```
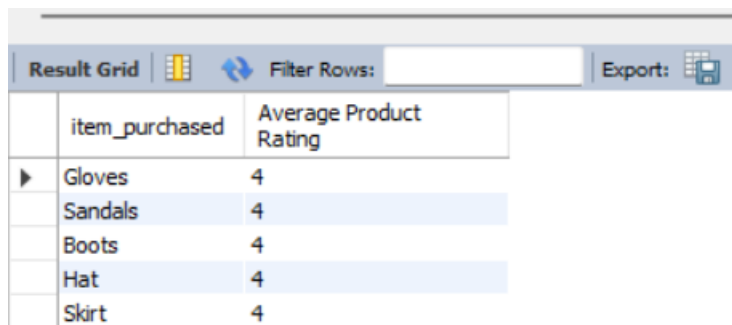
customer_id	age	gender	item_purchased	category	purchase_amount_amount	location	size	color	season	review_rating	subscription_status	shipping_type	discount_applied	previous_purchases	payment_method
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	2	Cash
3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	23	Credit Card
4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	49	PayPal
7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Free Shipping	Yes	49	Cash
9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6	Yes	Express	Yes	8	Venmo
12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9	Yes	Store Pickup	Yes	10	Bank Transfer
13	61	Male	Coat	Outerwear	72	Delaware	M	Gold	Winter	4.5	Yes	Express	Yes	37	Venmo
16	64	Male	Skirt	Clothing	81	Rhode Island	M	Teal	Winter	2.8	Yes	Store Pickup	Yes	8	PayPal
20	66	Male	Pants	Clothing	90	Rhode Island	M	Green	Summer	3.3	Yes	Standard	Yes	46	Debit Card
22	31	Male	Pants	Clothing	62	North Carolina	M	Charcoal	Winter	4.1	Yes	Store Pickup	Yes	22	Debit Card
24	31	Male	Pants	Clothing	88	Oklahoma	XL	White	Winter	4.4	Yes	Express	Yes	40	Credit Card
29	54	Male	Handbag	Accessories	94	North Carolina	M	Gray	Fall	4.4	Yes	Free Shipping	Yes	41	PayPal
32	33	Male	Dress	Clothing	79	West Virginia	L	Brown	Winter	4.7	Yes	Store Pickup	Yes	45	Venmo
33	36	Male	Jacket	Outerwear	67	Kansas	M	Silver	Summer	4.9	Yes	Free Shipping	Yes	37	Venmo
35	36	Male	T-shirt	Clothing	91	North Dakota	L	Violet	Spring	4.6	Yes	2-Day Shipping	Yes	38	PayPal
37	35	Male	T-shirt	Clothing	69	Illinois	M	Maroon	Winter	4.6	Yes	Free Shipping	Yes	44	PayPal
40	70	Male	Pants	Clothing	60	Arizona	S	Turquoise	Summer	4.2	Yes	Express	Yes	18	Credit Card
41	69	Male	Handbag	Accessories	76	Louisiana	L	Beige	Winter	4.6	Yes	Next Day Air	Yes	31	Debit Card
43	20	Male	Coat	Outerwear	100	Tennessee	M	Beige	Spring	4.1	Yes	Free Shipping	Yes	15	PayPal
44	25	Male	Scarf	Accessories	69	Ohio	L	Lavender	Fall	3.7	Yes	Store Pickup	Yes	19	PayPal
55	47	Male	Coat	Outerwear	94	New Mexico	M	Brown	Summer	4.2	Yes	Next Day Air	Yes	35	Debit Card

-- Q3. Which are the top 5 products with the highest average review rating?

```sql
select item_purchased, round(avg(review_rating)) as
"Average Product Rating"
from customer
group by item_purchased
order by avg(review_rating) desc
limit 5;
```
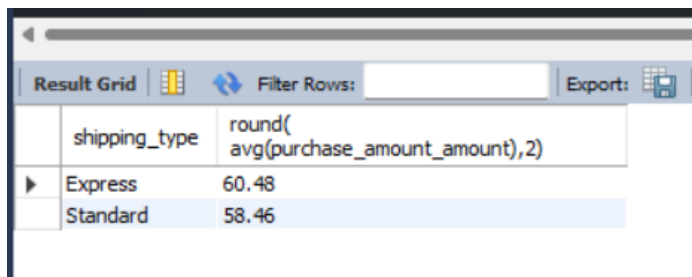
item_purchased	Average Product Rating
Gloves	4
Sandals	4
Boots	4
Hat	4
Skirt	4

-- Q4. Compare the average Purchase Amounts between Standard and Express Shipping.

```sql
select shipping_type , round(
avg(purchase_amount_amount),2)
 from customer group by shipping_type
 having shipping_type in ('Standard','Express');
```
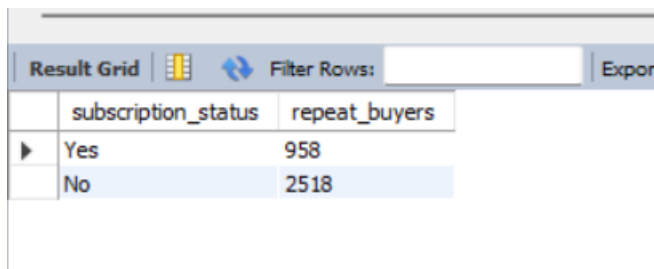
shipping_type	round(avg(purchase_amount_amount),2)
Express	60.48
Standard	58.46

```sql
-- Q9. Are customers who are repeat buyers (more than 5
previous purchases) also likely to subscribe?

SELECT subscription_status,
       COUNT(customer_id) AS repeat_buyers
FROM customer
WHERE previous_purchases > 5
GROUP BY subscription_status;
```
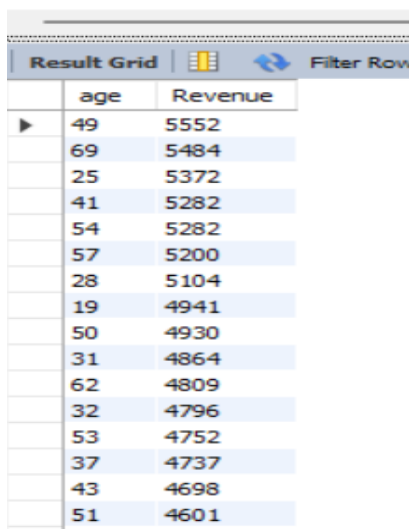
subscription_status	repeat_buyers
Yes	958
No	2518

```sql
-- Q10. What is the revenue contribution of each age
group?
select age , sum(purchase_amount_amount)  as Revenue
from customer group by age order by Revenue desc ;
```

age	Revenue
49	5552
69	5484
25	5372
41	5282
54	5282
57	5200
28	5104
19	4941
50	4930
31	4864
62	4809
32	4796
53	4752
37	4737
43	4698
51	4601

Dashboard in Power BI:

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

◆ **Dashboard Objectives**

- Provide a **single view of business performance**
- Enable management to **identify revenue drivers and customer trends**
- Support **data-driven decision-making** through interactivity

Key Visualizations

- **Category-wise Revenue Analysis** – Identifies top-performing product categories
- **Customer Demographic Analysis** – Revenue distribution by age group and gender
- **Subscription vs Non-Subscription Analysis** – Comparison of spending behavior
- **Seasonal Sales Trends** – Purchase patterns across seasons
- **Shipping Type Analysis** – Impact of shipping method on purchase amount.

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.**