Customer Behavior Analysis Report

▢ Project Overview:

The Customer Behavior Analysis project focuses on analyzing transactional data to identify patterns in customer purchasing behavior and preferences. The analysis includes customer segmentation, subscription trends, and buying habits. The insights derived from this project support strategic business decisions related to marketing optimization, revenue growth, and improved customer engagement.

▢ Data Description :

The initial data contains :

- **3900** customer transaction records
- **18** features

In which **Review Rating** columns contain **37** missing value. It includes customer demographic information, product details, and sales-related attributes.

The dataset covers the following categories of information:

- **Customer Demographics:** Age, Gender, Location
- **Product Details:** Item Purchased, Category, Size, Color, Season
- **Transaction & Sales Data:** Purchase Amount (USD), Payment Method, Shipping Type, Discount Applied, Promo Code Used
- **Customer Behavior:** Previous Purchases, Frequency of Purchases, Subscription Status, Review Rating.

	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	Payment Method	Frequency of Purchases
1155	1156	67	Male	T-shirt	Clothing	53	Tennessee	S	Red	Fall	2.6	No	Express	Yes	Yes	47	PayPal	Quarterly
3213	3214	27	Female	Sunglasses	Accessories	56	Texas	M	Brown	Winter	3.3	No	Standard	No	No	33	Credit Card	Monthly
2328	2329	45	Male	Pants	Clothing	36	North Dakota	M	Maroon	Fall	2.6	No	Express	No	No	29	Venmo	Bi-Weekly
2194	2195	61	Male	Socks	Clothing	29	Kentucky	M	Silver	Summer	3.5	No	Standard	No	No	48	Bank Transfer	Quarterly
3279	3280	59	Female	Hat	Accessories	50	Wyoming	S	Red	Fall	3.0	No	Express	No	No	1	Debit Card	Fortnightly

▢ Exploratory Data Analysis using Python :

Began with data Cleaning and preprocessing using python.

- **Initial Exploration:** Gathered details about the dataset using data.info() and data.describe().

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Customer ID           3900 non-null   int64
 1   Age                   3900 non-null   int64
 2   Gender                3900 non-null   object
 3   Item Purchased        3900 non-null   object
 4   Category              3900 non-null   object
 5   Purchase Amount (USD) 3900 non-null   int64
 6   Location              3900 non-null   object
 7   Size                  3900 non-null   object
 8   Color                 3900 non-null   object
 9   Season                3900 non-null   object
 10  Review Rating         3863 non-null   float64
 11  Subscription Status   3900 non-null   object
 12  Shipping Type         3900 non-null   object
 13  Discount Applied      3900 non-null   object
 14  Promo Code Used       3900 non-null   object
 15  Previous Purchases    3900 non-null   int64
 16  Payment Method        3900 non-null   object
 17  Frequency of Purchases 3900 non-null  object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
Shape: (3900, 18)
Size: 70200
```

```
Statistical Description:
        Customer ID          Age  Purchase Amount (USD)  Review Rating  \
count  3900.000000  3900.000000            3900.000000    3863.000000
mean   1950.500000    44.068462              59.764359       3.750065
std    1125.977353    15.207589              23.685392       0.716983
min       1.000000    18.000000              20.000000       2.500000
25%     975.750000    31.000000              39.000000       3.100000
50%    1950.500000    44.000000              60.000000       3.800000
75%    2925.250000    57.000000              81.000000       4.400000
max    3900.000000    70.000000             100.000000       5.000000

       Previous Purchases
count         3900.000000
mean            25.351538
std             14.447125
min              1.000000
25%             13.000000
50%             25.000000
75%             38.000000
max             50.000000
```

- **Handling Missing Value:** Found missing entries in the Review Rating data and filled them using the median for each individual category.

- **Feature Engineering:**

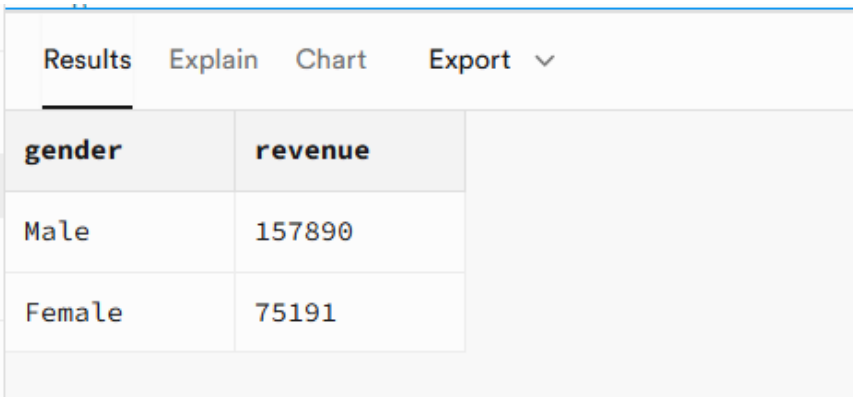
1. Categorized ages into young-adult, adult, middle-aged, and senior.
2. Created purchase_frequency_days from number of frequent purchase .

- Eliminated the promo_code_used column because it provided the same information as discount_applied.

▢ Data Analysis using SQL :

Performed structured analysis using SQL to answer key business question :

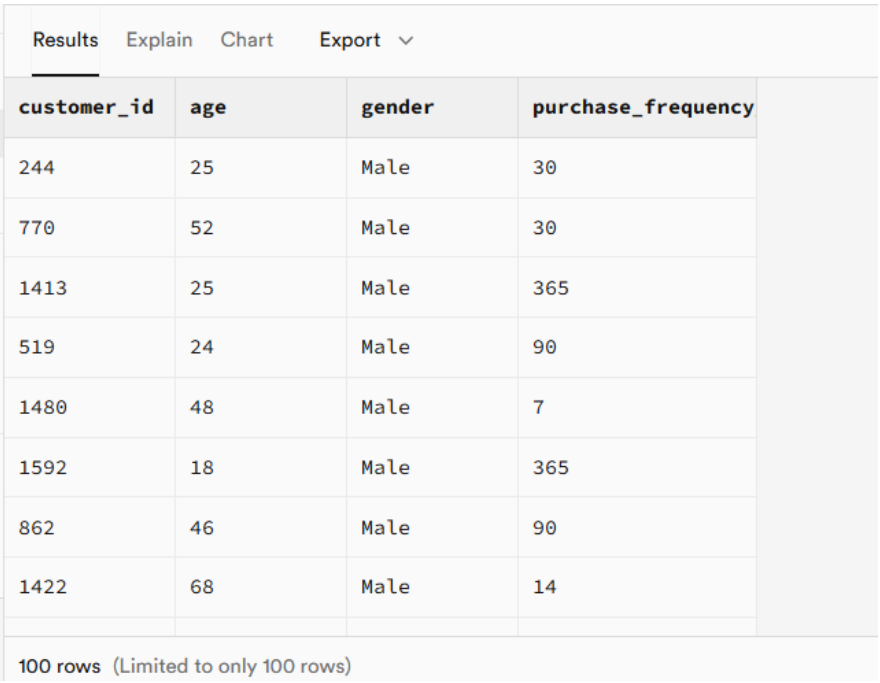
- **Revenue by gender :** Comparison of revenue generated by male vs female.



Results Explain Chart Export

gender	revenue
Male	157890
Female	75191

- **High-spending discount customers:** These are customers who used a discount but still spent more than the average purchase amount.



Results Explain Chart Export

customer_id	age	gender	purchase_frequency
244	25	Male	30
770	52	Male	30
1413	25	Male	365
519	24	Male	90
1480	48	Male	7
1592	18	Male	365
862	46	Male	90
1422	68	Male	14

100 rows (Limited to only 100 rows)

- **Top 5 product by review rating :** The top 5 product with the highest average review

rating.

Results Explain Chart Export

item_purchased	avg_rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.80
Skirt	3.78

- **Shipping type comparison :** Comparing the average purchase amount between Standard shipping and Express shipping.

Results Explain Chart Export

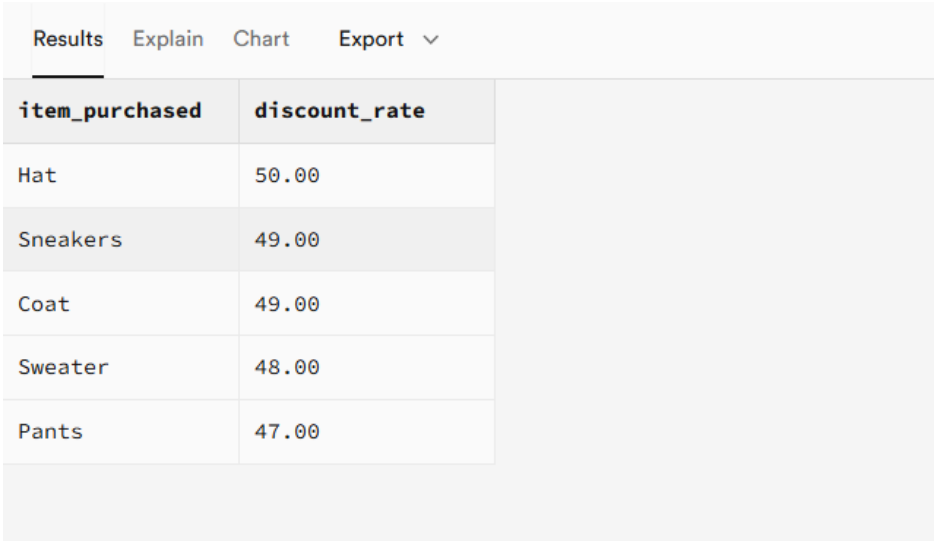
shipping_type	average_purchase_amount
Express	60.4752321981424149
Standard	58.4602446483180428

- **Subscribed Vs Non subscribed Customers :** Compare average spend and total revenue between subscribed and non-subscribed customer.

Results Explain Chart Export

subscription_status	average_spend	total_revenue	total_customers
Yes	59.49	62645	1053
No	59.87	170436	2847

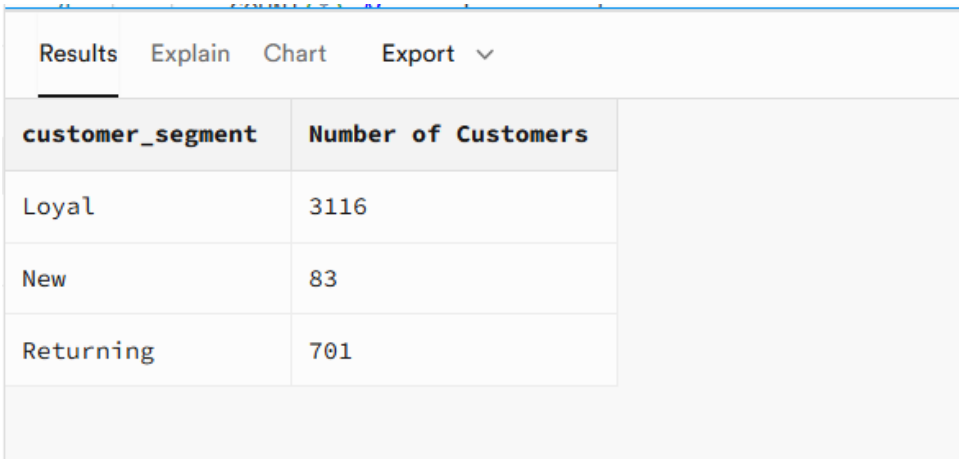
- **Discount dependent product :** 5 products have the highest percentage of purchases with discounts applied.



Results | Explain | Chart | Export

item_purchased	discount_rate
Hat	50.00
Sneakers	49.00
Coat	49.00
Sweater	48.00
Pants	47.00

- **Segmenting customer by previous purchases :** Segment and count customers into New, Returning , and Loyal based on their total number of previous purchases.



Results | Explain | Chart | Export

customer_segment	Number of Customers
Loyal	3116
New	83
Returning	701

- **Top 3 products per category :** The top3 most purchased products within each category

Results Explain Chart Export

item_rank	category	item_purchased	total_orders
1	Accessories	Jewelry	171
2	Accessories	Sunglasses	161
3	Accessories	Belt	161
1	Clothing	Blouse	171
2	Clothing	Pants	171
3	Clothing	Shirt	169
1	Footwear	Sandals	160
2	Footwear	Shoes	150

11 rows

- **Relation between Repeat buyers and Subscription:** Check whether Customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe.

Results Explain Chart Export

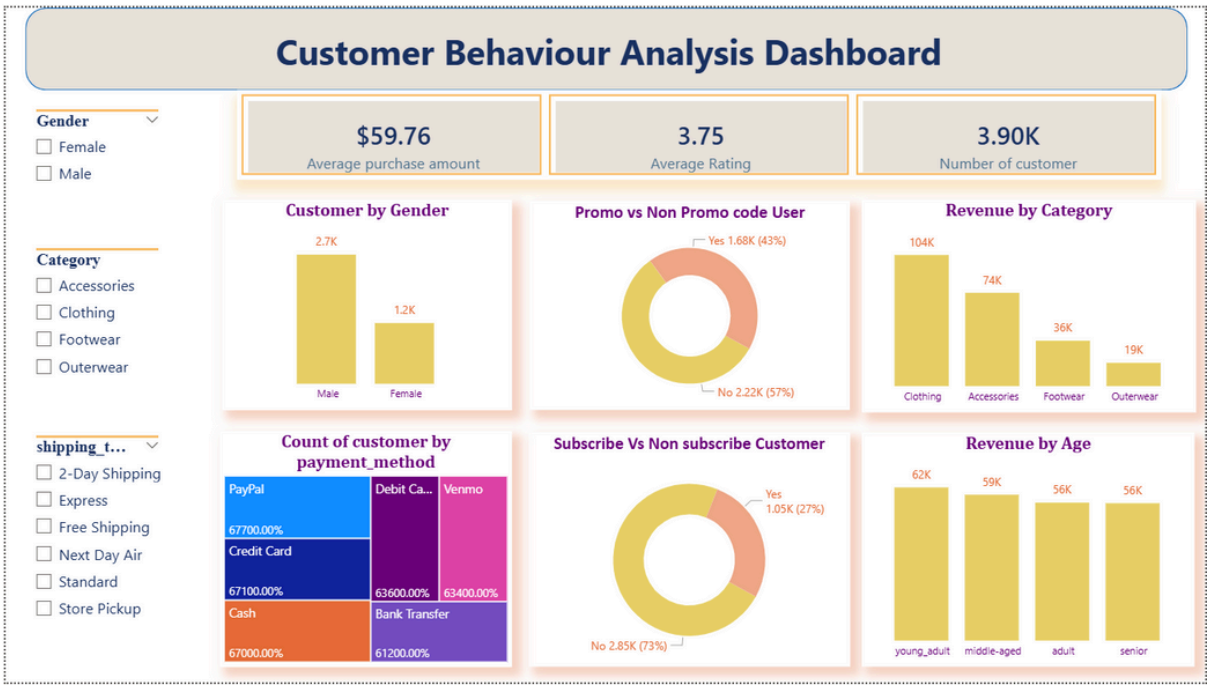
subscription_status	repeat_buyers
No	2583
Yes	980

- **Revenue by age_group :** Calculate Contribution of revenue by each age_group.

Results Explain Chart Export

age_group	revenue
young_adult	62143
middle-aged	59197
adult	55978
senior	55763

▢ Power BI Dashboard



▢ Business Recommendations

1. **Subscription Strategy Optimization:**

Insight

- Non-subscribed customers generate **much higher total revenue (₹170K vs ₹62K)** simply due to volume.
- Average spend is **almost the same** for subscribers and non-subscribers (~₹59).

Recommendation

- Improve the **value proposition of subscriptions** (exclusive discounts, early access, free express shipping).
- Target **high-spending non-subscribers** with personalized subscription offers.
- Introduce **tiered subscription plans** to increase adoption without reducing revenue.

2. **Focus on Loyal Customers**

Insight

- Loyal customers dominate the base (**3116 users**).
- New customers are extremely low (**only 83**).

Recommendation

- Launch **loyalty rewards** (points, cashback, referrals).
- Run **win-back campaigns** for returning customers to convert them into loyal ones

3. **Discount Effectiveness Review**

Insight

- Heavy discounts (47–50%) are applied on items like **Hats, Sneakers, Coats**.
- Some discounted items still don't appear in top-purchased products.

Recommendation

- Reduce discounts on **low-impact items** and reallocate offers to high-demand products.

4. **Gender-Based Revenue Strategy**

Insight

- Male customers generate **~68% of total revenue**.
- Female segment is under-monetized.

Recommendation

- Design **female-focused campaigns** .
- Personalize recommendations based on **gender preferences**.