Customer 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, 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 values in Review Rating column
3. **Exploratory Data Analysis by Python**
   We majorly performed data preparation and cleaning portion in python as:
   - **Data Loading:** We loaded data in Python using python library
   - **Explored data** using .info() and .describe()

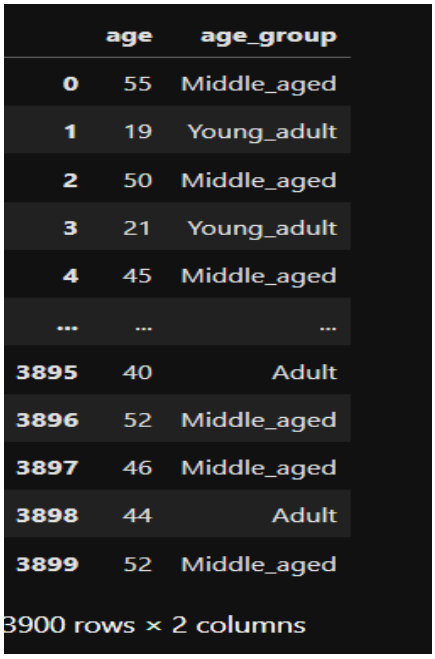
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             3900 non-null   int64
 1   Age                     3900 non-null   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
```

	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

- **Handling Missing Data:** Looked for empty values and filled missing review ratings using the median rating of each product category.

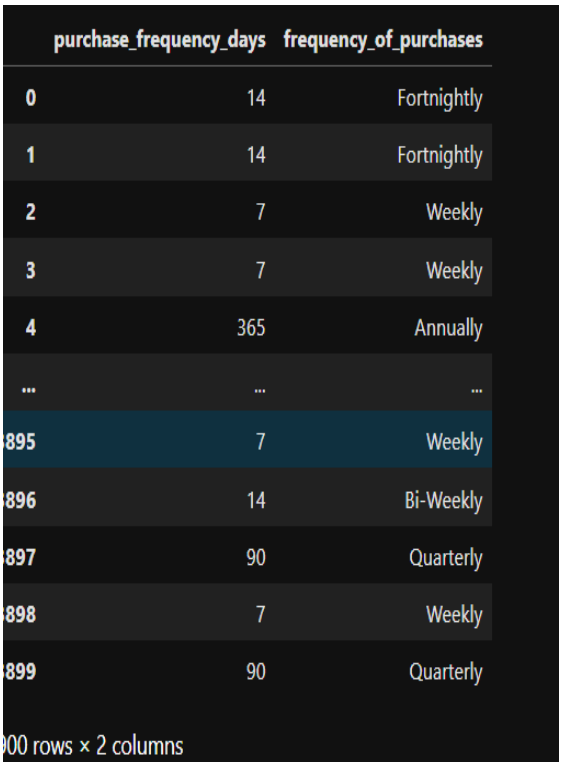
```
Customer ID                 0
Age                         0
Gender                      0
Item Purchased              0
Category                    0
Purchase Amount (USD)       0
Location                    0
Size                        0
Color                       0
Season                      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
```

- **Standardizing Columns:** Changed column names into lower case to make them easier to read and document.
- **Feature Engineering:**
    - Made an **age_group** column by grouping customer ages.
    - Created **purchase_frequency_days** using purchase-related information.

- 

	age	age_group
0	55	Middle_aged
1	19	Young_adult
2	50	Middle_aged
3	21	Young_adult
4	45	Middle_aged
...	...	...
3895	40	Adult
3896	52	Middle_aged
3897	46	Middle_aged
3898	44	Adult
3899	52	Middle_aged

3900 rows × 2 columns



	purchase_frequency_days	frequency_of_purchases
0	14	Fortnightly
1	14	Fortnightly
2	7	Weekly
3	7	Weekly
4	365	Annually
...	...	...
895	7	Weekly
896	14	Bi-Weekly
897	90	Quarterly
898	7	Weekly
899	90	Quarterly

00 rows × 2 columns

- **Checking Data Consistency:** Checked whether *discount_applied* and *promo_code_used* meant the same thing; removed *promo_code_used* after confirming it was redundant.
- **Database Integration:** Connected Python to MySQL and uploaded the cleaned Data Frame into the database for further SQL analysis.

4. **<u>Data Analysis by SQL:</u>**

We tried answering these business related questions using MySQL:

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

Result Grid | Filter Rows:

gender	Total_Revenue
Male	157890
Female	75191

- Which customers used a discount but still spent more than the average purchase amount?

Result Grid | Filter Rows:

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72
16	81
20	90
22	62

mytable 1 ×

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

Result Grid | Filter Ro

product	avg_rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Handbag	3.78

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

Result Grid | Filter Rows:

shipping_type	purchase_amount
Express	60.4752
Standard	58.4602

- Do subscribed customers spend more? Compare average spend and total revenue between subscribers and non-subscribers.

Result Grid | Filter Rows: | Export: | Wrap Cell Conte

subscription_status	COUNT(customer_id)	Average	avg_purchase_amount
Yes	1053	59.49	62645
No	2847	59.87	170436

- Which 5 products have the highest percentage of purchases with discounts applied?

item_purchased	discount_rate
Hat	50.00
Sneakers	49.66
Coat	49.07
Sweater	48.17
Pants	47.37

- Segment customers into New, Returning, and Loyal based on their total number of previous purchases, and show the count of each segment.

esult Grid | Filter Rows:

customer_segment	Number of Customers
Loyal	3116
Returning	701
New	83

- What are the top 3 most purchased products within each category?

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
3	Footwear	Sneakers	145
1	Outerwear	Jacket	163

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

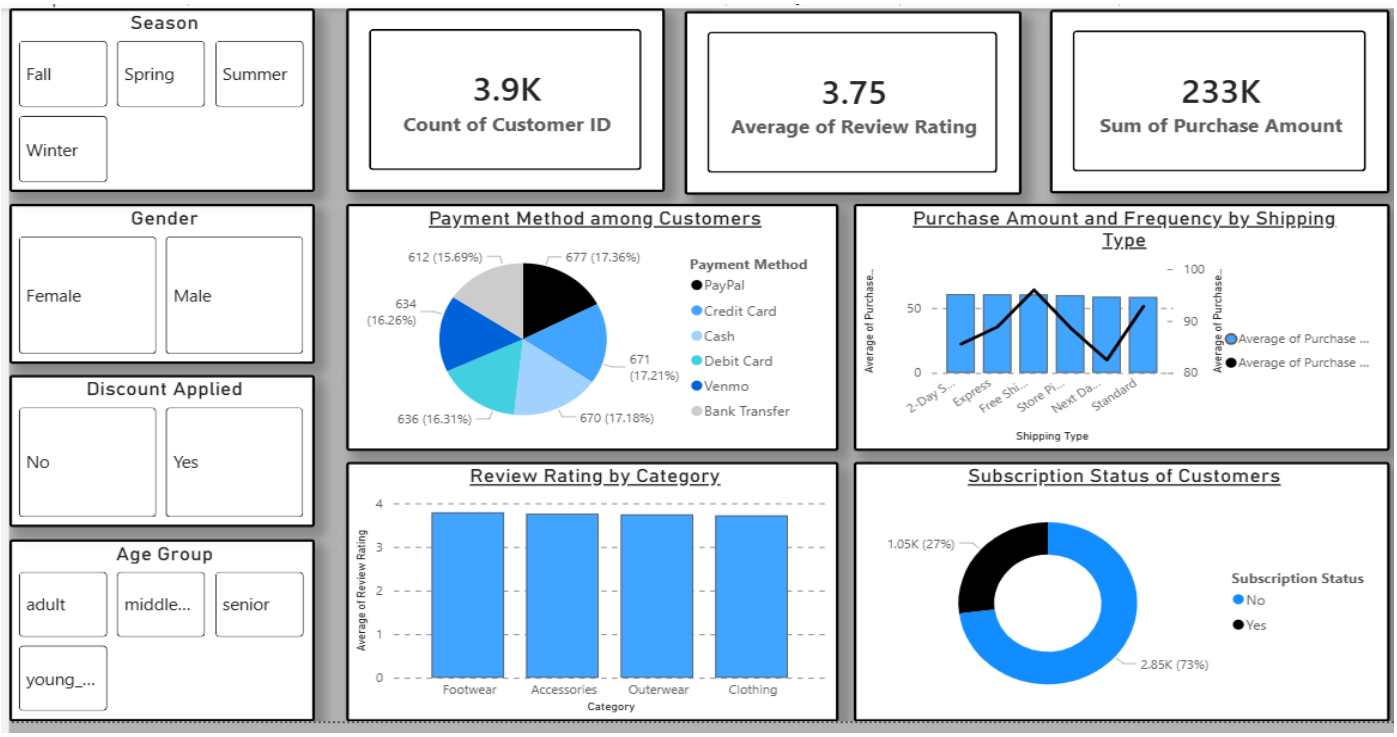
subscription_status	repeat_buyers
Yes	958
No	2518

- What is the revenue contribution of each age group?

sult Grid | Filter Rows:

age_group	total_revenue
Young_adult	62143
Middle_aged	59197
Adult	55978
Senior	55763

5. Dashboard in Power BI:



6. Business Recommendations

- **Increase Subscriptions:** Highlight special perks to encourage more customers to subscribe.
- **Strengthen Loyalty Programs:** Give rewards to frequent buyers to shift them into the “loyal customer” category.
- **Optimize Discount Strategy:** Make sure discounts attract customers without hurting profit margins.
- **Improve Product Promotion:** Feature the highest-rated and most popular products in marketing campaigns.
- **Use Targeted Marketing:** Direct marketing efforts toward age groups that generate higher revenue and customers who prefer express delivery.