Customer Behavior Analysis Report:

Project Overview:

The Project analyze Customer Shopping Behavior using transactional data from purchases across various product categories.

The goal is to uncover insights into spending patterns, subscription behavior, product preferences, customer segment behavior to guide strategic business decisions.

Dataset Info:

- Rows: 3900
- Columns: 18
- Customer Details: Age, gender, location, subscription status
- Purchase details: Item Purchased, Category, Amount, Season, size, color
- Shopping details: Discount, Previous purchase, Frequency purchase, Review rating, Shipping type.
  (Missing Data 37 in Review Rating)

Exploratory Data Analysis using Python:

Begin with data explorations checking columns Data structure

- Data loading: Import data using Pandas
- Exploration: used df.info() and describe for statistical summary

```
df.info()
```

```
<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
```

```
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

- Missing data handled: Check null values and imputed missing value in Review Rating column using median

```
df["Review Rating"] = df.groupby("Category")["Review Rating"].transform(lambda x:x.fillna(x.median()))
```

```
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             0
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
```

- Column standardization: Rename columns for better readability
- Featured Engineering: Created Age_Groups column from Ages,

Age_Groups	Purchase_Frequency_Days
Mid-Age	14
Young	14
Mid-Age	7
Young	7
Mid-Age	365
Mid-Age	7
Senior	90
Young	7
Young	365
Mid-Age	90

  Created Purchase_Frequency_days column from purchase data
- Data consistency check: Verified Discount and Promocode are redundant dropped Promocode column

Data Analysis using SQL:

- Perform structured analysis in MS SQL Server to solved

Business Questions and Solutions

-- Q1. What are the total revenue generated by Male vs Female Customers:

	Gender	Total_Revenue
1	Male	157890
2	Female	75191

-- Q2. Which Customer used a discount but still spend more than average purchase amount :

	Customer_ID	Purchase_Amount
1	43	100
2	96	100
3	194	100
4	205	100
5	244	100
6	249	100
7	456	100
8	519	100

-- Q3. Which are the Top 5 Products with the highest Average Review Rating:

	Item_Purchased	Avg_Review_Rating
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.8
5	Handbag	3.78

-- Q4. Compare the Average Purchase Amount between Standard and Express shipping:

	Shipping_Type	Avg_Purchase
1	Standard	58
2	Express	60

-- Q5. Do subscribed customers spend more ? Compare average spend and total revenue between Subscribers and Non-subscribers:

	Subscription_Status	Total_Customer	Avg_purchase	Total_Revenue
1	Yes	1053	59	62645
2	No	2847	59	170436

-- Q6. Which 5 Products have the highest percentage of purchase with Discount applied :

	Item_Purchased	Discount_Rate
1	Hat	50
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

--Q7. Segment the Customers into New, Returning, Loyal based on their total number of Previous Purchase & show count of each segment :

	Customer_Segment	Total_Customer
1	Loyal	3116
2	Returning	701
3	New	83

-- Q8. What are the Top 3 most Purchase Product within each Category :

	Customer_Segment	Total_Customer
1	Loyal	3116
2	Returning	701
3	New	83

-- Q9. Are Customers who are repeat buyers (More than 5 Previous Purchases) also likely to subscribe:

	Subscription_Status	Total_buyers
1	Yes	958
2	No	2518

-- Q10. What is the Revenue Contribution of each Age Group:

	Subscription_Status	Total_buyers
1	Yes	958
2	No	2518

- **Dashboard in Power BI:**

  Built an interactive dashboard in Power BI to present insights visually.