Customer Shopping Behavior Analysis

Data Analyst Project

1. Project Overview

This project analyzes customer shopping behavior using transactional data from 3,900 purchases across diverse product categories. The primary objective is to derive actionable insights related to spending patterns, customer segmentation, product performance, and subscription behavior to support data-driven business decisions.

2. Dataset Summary

- Total Records: 3,900 rows
- Total Attributes: 18 columns
- Customer Demographics: Age, Gender, Location, Subscription Status
- Purchase Details: Item Purchased, Category, Purchase Amount, Season, Size, Color
- Behavioral Features: Discount Applied, Previous Purchases, Purchase Frequency, Review Rating, Shipping Type
- Data Quality: 37 missing values identified in the Review Rating column

3. Exploratory Data Analysis (Python)

- Data Loading and Inspection using Pandas
- Used df.info() and df.describe() for structural and statistical understanding
- Handled missing Review Rating values using median imputation by product category
- Standardized column names using snake_case convention
- Engineered new features including age_group and purchase_frequency_days
- Removed redundant promo_code_used column after consistency validation
- Exported cleaned dataset to MySQL for SQL-based analysis

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN

Previous Purchases	Payment Method	Frequency of Purchases
3900.000000	3900	3900
NaN	6	7
NaN	PayPal	Every 3 Months
NaN	677	584
25.351538	NaN	NaN
14.447125	NaN	NaN
1.000000	NaN	NaN
13.000000	NaN	NaN
25.000000	NaN	NaN
38.000000	NaN	NaN
50.000000	NaN	NaN

```
<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
```

4. Data Analysis (SQL)

- Revenue comparison by gender

gender	sum(purchase_amount)
Male	157890
Female	75191

- Identification of high-spending customers using discounts

customer_id	purchase_amount
2	64
3	73
4	90
7	85
9	97
12	68
13	72

- Top 5 products ranked by average review rating

item_purchased	avg(review_rating)
Gloves	3.8614285714285725
Sandals	3.8443750000000003
Boots	3.8187500000000005
Hat	3.8012987012987005
Skirt	3.784810126582278

- Purchase amount comparison across shipping types

shipping_type	round(avg(purchase_amount),2)
Express	60.48
Standard	58.46

- Spending analysis for subscribers vs. non-subscribers

subscription_status	count(customer_id)	total_revenue	average_spend
Yes	1053	62645	59.4919
No	2847	170436	59.8651

2847

- Identification of products highly dependent on discounts

item_purchased	discount_rate
Hat	50.00000
Sneakers	49.65517
Coat	49.06832
Sweater	48.17073
Pants	47.36842

- Customer segmentation into New, Returning, and Loyal groups

customer_segment	Number of customer
Loyal	3116
returning	701
New	83

- Top 3 most purchased products per 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

Result 8

- Correlation between repeat purchases and subscription status

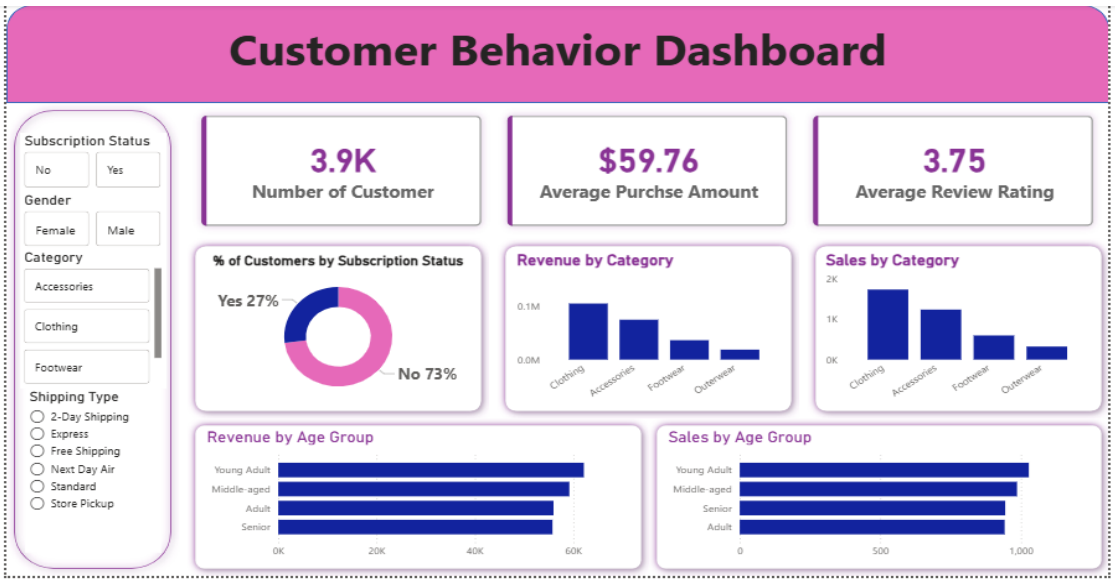
subscription_status	count(customer_id)
Yes	958
No	2518

- Revenue contribution analysis by age group

age_group	total_revenue
Senior	55763
Adult	55978
Middle-aged	59197
Young Adult	62143

5. Dashboard Development (Power BI)

An interactive Power BI dashboard was developed to visualize key insights. The dashboard enables effective comparison across customer segments, products, and revenue drivers, making the findings accessible to both technical and non-technical stakeholders.



6. Business Recommendations

- Enhance subscription models with exclusive incentives
- Implement loyalty programs to convert returning customers into loyal customers
- Optimize discount strategies to maintain profitability
- Promote top-performing and highly-rated products
- Focus marketing efforts on high-revenue age groups and express-shipping customers