

CUSTOMER SHOPPING BEHAVIOR ANALYSIS REPORT

• PROJECT OVERVIEW

This project analyzes customer shopping behavior 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 behavior to guide strategic business decisions.

• 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, Color)
 - Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

• DATA ANALYSIS USING PYTHON

Data Loading & Inspection

- The dataset was imported using pandas from a CSV file (customer_shopping_behavior.csv).
- Initial inspection with .head(), .info(), and .describe() helped understand data types, structure, and summary statistics.

Data Quality Check

- Missing and null values were identified using df.isnull().sum().
- Data consistency across columns like *Age*, *Purchase Amount (USD)*, and *Review Rating* was verified.

Exploratory Data Analysis (EDA)

- Descriptive statistics were computed to examine key measures such as mean, median, and standard deviation of numerical fields.
- The distribution of categorical variables like *Gender*, *Category*, and *Shipping Type* was analyzed.

Customer Demographics

- Insights into customer segments by gender and age group were derived to understand the buyer composition.
- Age distribution analysis revealed which groups contribute most to revenue.

Revenue & Spending Patterns

- Average purchase amount and total revenue were computed per category and customer type.
- Purchase frequency and previous purchase data were used to identify loyal vs. new customers.

Discount & Subscription Analysis

- Compared spending patterns between customers who received discounts and those who didn't.
- Assessed how subscription status correlates with average purchase amount and revenue contribution.

Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping

```
<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
```

	age	age_group
0	55	Middle-aged
1	19	Young Adult
2	50	Middle-aged
3	21	Young Adult
4	45	Middle-aged
5	46	Middle-aged
6	63	Senior
7	27	Young Adult
8	26	Young Adult
9	57	Middle-aged

```
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
```

- ANALYSIS USING SQL

Revenue Comparison by Gender

	Gender text 	revenue numeric 
1	Female	75191
2	Male	157890

High-Spending Customers Using Discounts

	Customer ID bigint 	Purchase Amount (USD) bigint 
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72

Top 5 Products by Average Review Rating

	Item Purchased text 	Average Product Rating numeric 
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.79

Average Purchase Amount Comparison by Shipping Type

	Shipping Type text 	Average Purchase Amount numeric 
1	Standard	58.46
2	Express	60.48

Impact of Subscription on Customer Spending

	Subscription Status text	Total Customers bigint	Average Spend numeric	Total Revenue numeric
1	No	2847	59.87	170436.00
2	Yes	1053	59.49	62645.00

Top 5 Products with Highest Discount Usage

	Item Purchased text	Discount Rate (%) numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

Customer Segmentation by Loyalty Level

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

Top 3 Most Purchased Products per Category

	item_rank bigint	Category text	Item Purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160

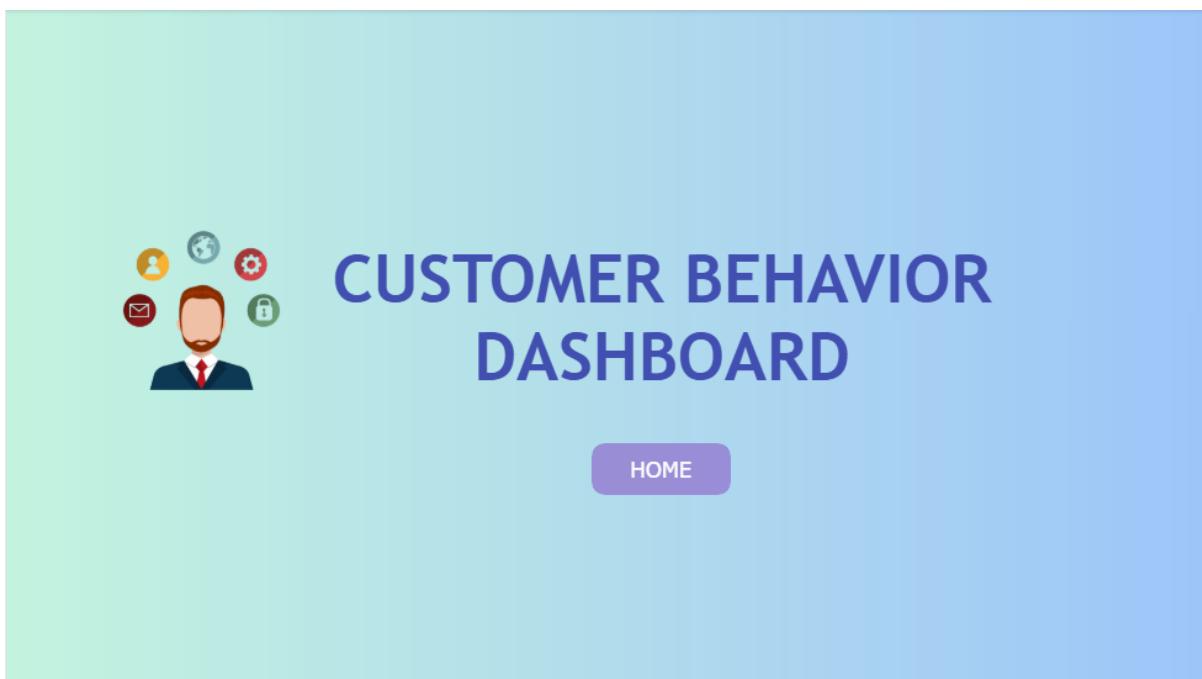
Subscription Trends Among Repeat Buyers

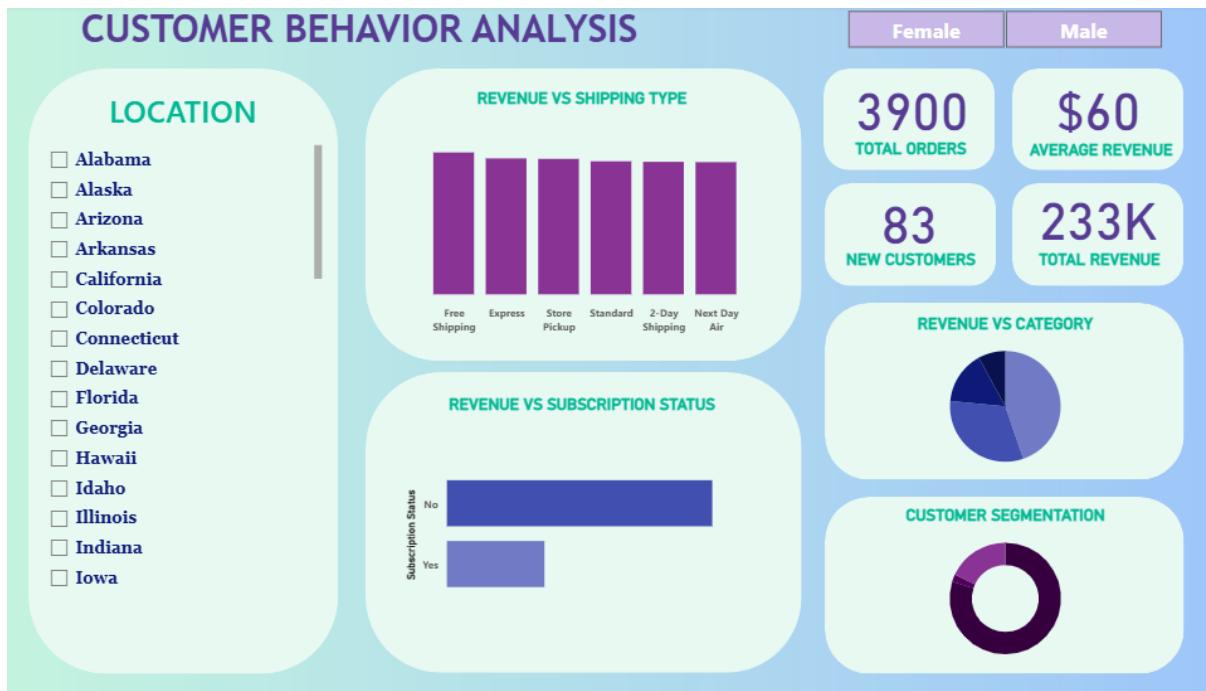
	Subscription Status text	Repeat Buyers bigint
1	No	2518
2	Yes	958

Revenue Contribution by Age Group

	Age Group text	Total Revenue numeric
1	41-60	89445
2	25-40	71214
3	60+	43164
4	Under 25	29258

- DATA VISUALISATION USING POWER BI





• RECOMMENDATIONS

Promote Subscriptions:

Subscribed customers spend significantly more than non-subscribers — offer exclusive deals, early access, or free shipping to encourage more sign-ups.

Reward Loyal Customers:

Loyal and repeat buyers contribute major revenue; strengthen retention through loyalty points, personalized offers, and referral bonuses.

Optimize Discount Strategy:

Discounts should target new or low-spending customers to boost conversion, while minimizing overuse on already high-performing products.

Enhance Express Shipping Offers:

Customers choosing express shipping show higher spending patterns — promote this option with small incentives or free upgrades to increase order value.

Focus on Top Products and Age Groups:

Invest in marketing top-rated, best-selling products, especially to the 25–40 age segment which generates the highest revenue.

Use Data Insights for Decisions:

Continuously monitor key metrics (revenue, discount usage, loyalty trends) in Power BI to guide pricing, marketing, and inventory decisions.