

Customer Shopping Behavior Analysis

1. 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.

2. Dataset Summary

- **Rows** = 3,900
- **Columns** = 18
- **Key features** -
 - Customer demographics (Age, Gender, Location, etc.)
 - Purchase details (Items purchased, Category, etc.)
 - Shopping behavior (Discount applied, Promo code used, Frequency of purchase)
- **Missing data** = 37 values in review_rating column

3. Exploratory Data Analysis

- **Data loading** - Imported dataset using pandas.
- **Initial exploration** - Used `df.info()` to check structure and `.describe()` for summary statistics.

	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
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900	3900.000000	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.351538	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.447125	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN	1.000000	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN	13.000000	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN	25.000000	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN	38.000000	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN	50.000000	NaN

- **Missing data handling** - Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

- **Column standardization** - Renamed columns to snake case for better readability and documentation.
- **Feature engineering** -
 - Created the `age_group` column by binning customer ages.
 - Created the `purchase_frequency_days` column from the purchase data.
- **Data consistency check** - Verified if `discount_applied` and `promo_code_used` were redundant; dropped `promo_code_used`.

4. Data Analysis using SQL

- The following questions were answered using the SQL queries -
 - Q1. What is the total revenue generated by male vs. female customers?
 - Q2. Which customers used a discount but still spent more than the average purchase amount?
 - Q3. Which are the top 5 products with the highest average review rating?
 - Q4. Compare the average Purchase Amounts between Standard and Express Shipping.
 - Q5. Do subscribed customers spend more? Compare average spend and total revenue between subscribers and non-subscribers.
 - Q6. Which 5 products have the highest percentage of purchases with discounts applied?
 - Q7. Segment customers into New, Returning, and Loyal based on their total number of previous purchases, and show the count of each segment.
 - Q8. What are the top 3 most purchased products within each category?
 - Q9. Are customers who are repeat buyers (more than 5 previous purchases) also likely to subscribe?
 - Q10. What is the revenue contribution of each age group?

5. Dashboard in Power BI

- Here is the Power BI dashboard -



6. Business Recommendations

- **Boost subscriptions** - Promote exclusive benefits for subscribers.
- **Customer loyalty programs** - Reward repeat buyers to move them into the "Loyal" segment.
- **Review discount policy** - Balance sales boosts with margin control.
- **Product positioning** - Highlight top-rated and best-selling products in campaigns.
- **Targeted marketing** - Focus efforts on high-revenue age groups and express-shipping users.