

Customer Shopping Behavior Analysis Report

1. Project Overview and Data Summary

This analysis leverages transactional data from **3,900 customer purchases** to uncover key insights into spending habits, product popularity, and subscription trends. The main objective is to provide a data-driven foundation for strategic business decisions.

Dataset Snapshot

Feature	Details
Total Purchases (Rows)	3,900
Key Features	Demographics, Purchase Details (Amount, Category), Shopping Behavior (Discount, Reviews)
Missing Data	37 values in the Review Rating column
Average Purchase Amount	\$59.76
Average Review Rating	3.75

2. Data Preparation and Exploratory Data Analysis using Python

The initial data exploration in Python included loading the dataset and checking its structure.

Data Cleaning and Handling

- Missing Data Imputation:** The 37 missing values in the **Review Rating** column were handled by imputing them with the median rating specific to each product category.
- Column Standardization:** Column names were converted to snake case for improved documentation and readability.

- **Data Consistency:** The promo_code_used column was dropped after verifying its redundancy with discount_applied.

New Features Created

- **age_group:** Customers were segmented by age.
- **purchase_frequency_days:** Created from existing purchase data.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	39
unique	Nan	Nan	2	25	4	Nan	50	4	25	4	Nan	2	6	
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	M	Olive	Spring	Nan	No	Free Shipping	
freq	Nan	Nan	2652	171	1737	Nan	96	1755	177	999	Nan	2847	675	22
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan	Nan	Nan	3.750065	Nan	Nan	Nan
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan	Nan	Nan	0.716983	Nan	Nan	Nan
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan	Nan	Nan	2.500000	Nan	Nan	Nan
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan	Nan	Nan	3.100000	Nan	Nan	Nan
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan	Nan	Nan	3.800000	Nan	Nan	Nan
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan	Nan	Nan	4.400000	Nan	Nan	Nan
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan	Nan	Nan	5.000000	Nan	Nan	Nan

Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases
3900	3900	3900.000000	3900	3900
2	2	Nan	6	7
No	No	Nan	PayPal	Every 3 Months
2223	2223	Nan	677	584
Nan	Nan	25.351538	Nan	Nan
Nan	Nan	14.447125	Nan	Nan
Nan	Nan	1.000000	Nan	Nan
Nan	Nan	13.000000	Nan	Nan
Nan	Nan	25.000000	Nan	Nan
Nan	Nan	38.000000	Nan	Nan
Nan	Nan	50.000000	Nan	Nan

The cleaned data was then loaded into a MySQL database for structured SQL analysis.

3. Key Business Insights from SQL Analysis

Structured analysis was performed to address specific business questions:

⌚ Revenue Generation

- **Gender Comparison:** Male customers generated significantly more revenue (\$157,890) than Female customers (\$75,191).

	gender 	revenue 
	text	numeric
1	Female	75191
2	Male	157890

- **Age Group Revenue:** Young Adults contributed the highest total revenue (\$62,143), followed closely by Middle-aged customers (\$59,197).
- **Discount Users:** A total of **839** customers used a discount but still spent above the average purchase amount.

	customer_id 	purchase_amount 
	bigint	bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88

Total rows: 839 Query complete 00:00:00

📦 Shipping and Subscription

- **Shipping Preference:** The **Express** shipping type is associated with a slightly higher average purchase amount (\$60.48) compared to Standard shipping (\$58.46).

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

- **Subscription Status:** Only **27%** of customers have a subscription. Non-subscribers generate the majority of total revenue (\$170,436) and have a slightly higher average spend (\$59.87) than subscribers (\$59.49).

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

- **Repeat Buyers & Subscriptions:** Out of the repeat buyers (customers with \$>5\$ purchases), **2,518** do not have a subscription, while **958** do.

🛒 Product Performance

- **Top 5 Highest-Rated Products:** **Gloves** have the highest average rating (3.86), followed by Sandals (3.84) and Boots (3.82).

	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	3.78

- **Discount-Dependent Products:** Hat has the highest percentage of discounted purchases (50.00%), followed by Sneakers (49.66%) and Coat (49.07%).

	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:** The majority of the customer base (3,116) falls into the **Loyal** segment, with **701** returning and only **83** new customers.

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

- **Top Products by Category:**

- **Accessories:** Jewelry (171 orders)
- **Clothing:** Blouse (171 orders)
- **Footwear:** Sandals (160 orders)
- **Outerwear:** Jacket (163 orders)

	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
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

4. Dashboard Creation in Power BI

Finally, we built an interactive dashboard in **Power BI** to present insights visually.



5. Strategic Business Recommendations

Based on the analysis, the following recommendations are suggested to drive growth:

- Boost Subscriptions:** Given that 73% of customers are non-subscribers, promote **exclusive benefits** to increase the subscription rate.
- Enhance Customer Loyalty:** Implement programs to reward repeat buyers and incentivize them to transition into the "Loyal" segment.
- Review Discount Policy:** Re-evaluate the discount strategy to ensure a balance between boosting sales and maintaining profit margins.
- Targeted Marketing:** Direct marketing efforts toward the **Young Adult** age group and users who choose **Express** shipping, as they contribute significantly to revenue and average spend, respectively.
- Strategic Product Positioning:** Feature the top-rated (e.g., Gloves, Sandals) and best-selling products (e.g., Jewelry, Blouse) prominently in marketing campaigns.

