

Customer Purchase Insights Report

1. Project Overview

In this project, I analyzed customer shopping behavior using transactional data from 3,900 purchases across multiple product categories. The primary goal was to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to inform data-driven business decisions.

2. Dataset Summary

- 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
- Rows: 3,900
- Columns: 18

3. Exploratory Data Analysis (Python)

I started by preparing and cleaning the dataset using Python:

- **Data Loading:** Imported the dataset using pandas
- **Initial Exploration:** Used `.info()` to check structure and `.describe()` for summary statistics

I visualized the initial summary statistics of the dataset, including count, mean, standard deviation, minimum, and maximum values, to understand the distribution and scale of each feature. The table below shows these 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	Frequency of Purchases
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	3900	3900	3900.000000	3900	3900
unique	NaN	NaN	2	25	4	NaN	50	4	25	4	NaN	2	6	2	2	NaN	6	7
top	NaN	NaN	Male	Blouse	Clothing	NaN	Montana	M	Olive	Spring	NaN	No	Free Shipping	No	No	NaN	PayPal	Every 3 Months
freq	NaN	NaN	2652	171	1737	NaN	96	1755	177	999	NaN	2847	675	2223	2223	NaN	677	584
mean	1950.500000	44.068462	NaN	NaN	NaN	59.764359	NaN	NaN	NaN	NaN	3.750065	NaN	NaN	NaN	NaN	25.351538	NaN	NaN
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	NaN	NaN	14.447125	NaN	NaN
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	NaN	NaN	1.000000	NaN	NaN
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	NaN	NaN	13.000000	NaN	NaN
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	NaN	NaN	25.000000	NaN	NaN
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	NaN	NaN	38.000000	NaN	NaN
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	NaN	NaN	50.000000	NaN	NaN

- **Handling Missing Data:** Imputed missing values in the *Review Rating* column using the median rating per product category
- **Column Standardization:** Renamed columns to snake case for better readability and documentation.
- **Feature Engineering:**
 - Created *age_group* column by binning customer ages
 - Created *purchase_frequency_days* column to measure customer purchase intervals
- **Data Consistency Check:** Evaluated if *discount_applied* and *promo_code_used* for redundancy; removed *promo_code_used* for clarity
- **Database Integration:** Connected the cleaned dataset to PostgreSQL for structured analysis

4. Data Analysis (SQL – PostgreSQL)

I conducted in-depth analysis using SQL to answer business-critical questions:

1. **Revenue by Gender** – Compared total revenue generated by male vs. female customers

	gender text	revenue numeric
1	Female	75191
2	Male	157890

2. **Revenue by Age Group** – Determined revenue contribution per age group

	age_group text	total_revenue numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

3. **Subscribers vs. Non-Subscribers** – Analyzed average spend and total revenue by subscription status

	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

4. **High-Spending Discount Users** – Identified customers who spent above average despite using discounts

	customer_id bigint	purchase_amount 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 00:00:00.133

5. **Discount-Dependent Products** – Found the 5 products with the highest percentage of discounted purchases

	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

6. Shipping Type Comparison – Compared purchase amounts for Standard vs. Express shipping

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

7. Top 5 Products by Rating – Highlighted products with the highest average review ratings

	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

8. Top 3 Products per Category – Identified the most purchased items within each 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
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. Customer Segmentation – Classified customers as New, Returning, or Loyal based on purchase history

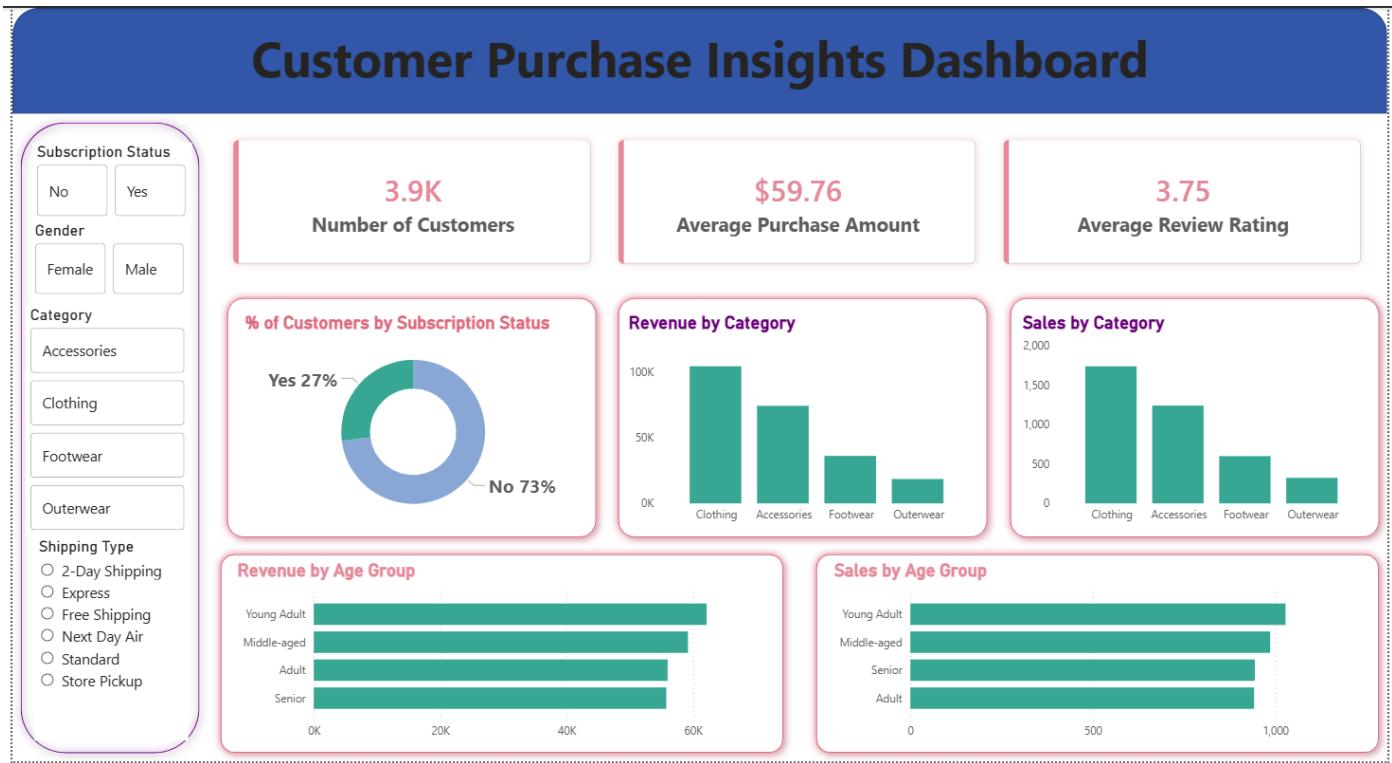
	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

10. Repeat Buyers & Subscriptions – Examined whether frequent buyers (>5 purchases) are more likely to subscribe

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

5. Dashboard (Power BI):

I created an interactive Power BI dashboard to visualize key insights, allowing stakeholders to explore revenue trends, product performance, and customer segments in a dynamic way.



6. Business Recommendations

Based on the analysis, I proposed the following actionable strategies:

- **Targeted Marketing:** Focus efforts on high-revenue age groups and express-shipping users
- **Review Discount Policy:** Balance discount strategies to improve margins
- **Boost Subscriptions:** Promote exclusive benefits to increase subscriber base
- **Product Positioning:** Highlight top-rated and best-selling products in marketing campaigns
- **Customer Loyalty Programs:** Reward repeat buyers to encourage loyalty