

Consumer Behavior Analysis Project

Python • SQL • Power BI

Records: 3,900 • Features: 18

1. Executive Summary

This project analyzes 3,900 transactional records to identify patterns in customer shopping behavior across demographics, product categories, and sales channels (online vs offline). The main objective was to answer the business question:

"How can the company leverage consumer shopping data to identify trends, improve customer engagement, and optimize marketing and product strategies?"

2. Dataset & Preparation

Rows: 3,900

Columns: 18

Key Data Fields

- **Customer Demographics:** age, gender, location, subscription status
 - **Purchase Details:** item purchased, category, purchase amount, size, color, season
 - **Shopping Behavior:** discount applied, promo code used, shipping type, frequency of purchases, previous purchases
 - **Customer Feedback:** review rating
 - **Data Quality:** 37 missing values in *review_rating*,
-

3. Exploratory Data Analysis Using Python

- Loaded dataset using pandas
- Verified structure using df.info() and explored descriptive statistics using df.describe()

	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	

- Review_rating had 37 missing values — imputed using median rating per category.
- Converted column names into snake_case for readability and consistent documentation
- New features created:
 - age_group (e.g., Young Adult, Adult, Middle-aged, Senior)
 - purchase_frequency_days (days between purchases)
- promo_code_used dropped because it was redundant with discount_applied.
- Cleaned dataset exported from Google Colab as cleaned_customer.csv.

4. Data Analysis using SQL (Business Transactions)

Total revenue generated by male vs. female customers

	gender	revenue
	Male	157890
▶	Female	75191

Customers used a discount but still spent more than the average purchase amount

	customer_id	purchase_amount
▶	2	64
	3	73
	4	90
	7	85
	9	97
	12	68
	13	72
	16	81
	20	90

Total Rows: 839

Top 5 products with the highest average review rating

	item_purchased	avg_rating
▶	Gloves	3.86
	Sandals	3.84
	Boots	3.82
	Hat	3.8
	Handbag	3.78

Average Purchase Amounts between Standard and Express Shipping.

	shipping_type	avg_purchase_amount
▶	Express	60.48
	Standard	58.46

Average spend and total revenue between subscribers and non-subscribers.

	subscription_status	total_customers	avg_spend	total_revenue
▶	Yes	1053	59.49	62645
	No	2847	59.87	170436

Top 5 products with the highest percentage of purchases with discounts applied

	item_purchased	discounted_purchases	discount_rate
▶	Hat	77	50.00
	Sneakers	72	49.66
	Coat	79	49.07
	Sweater	79	48.17
	Pants	81	47.37

Segment customers into New, Returning, and Loyal based on their total number of previous purchases.

	customer_segment	Number of Customers
▶	Loyal	3116
	Returning	701
	New	83

Top 3 most purchased products within each 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
	2	Footwear	Shoes	150
	3	Footwear	Sneakers	145
	1	Outerwear	Jacket	163
	2	Outerwear	Coat	161

Checked whether customers with more than 5 purchases are more likely to subscribe or not

	subscription_status	repeat_buyers
▶	Yes	958
	No	2518

Revenue contribution of each age group

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

5. Dashboard in Power BI



5.1. KPI Cards: Number of Customers, Avg Purchase Amount, Avg Review Rating

These top-level indicators summarize the overall business performance.

- Number of Customers (3,900) reflects the transaction volume used in the analysis.
- Avg Purchase Amount (\$59.8) provides insight into typical spending behavior.
- Avg Review Rating (3.8) highlights general customer satisfaction across products.

5.2. Customer Subscription Status

This donut chart shows that 27% of customers are subscribed, while 73% are non-subscribed.

This segmentation helps identify the size of the high-value subscriber base and evaluate opportunities to increase subscription adoption.

5.3. Revenue by Category

Clothing and Accessories generate the largest portions of revenue, followed by Footwear and Outerwear.

The visual helps determine which product lines drive financial performance and where to prioritize inventory and marketing efforts.

5.4. Sales by Category (Count of Purchases)

Clothing leads with 1,737 sales, followed by Accessories, Footwear, and Outerwear. Comparing sales volume to revenue reveals whether high sales come from high-value items or frequent low-cost purchases.

5.5. Revenue by Age Group

The horizontal bar chart segments revenue contribution by demographic groups. Young Adults generate the highest revenue, followed closely by Middle-aged and Adult groups, with Seniors contributing slightly less.

5.6. Sales by Age Group

Young Adults again lead with the highest transaction count, confirming that they are both frequent shoppers and strong revenue contributors.

5.7. Slicers: Discount Status, Gender, Payment Mode, Shipping Type

These interactive filters allow users to dynamically analyze specific segments—for example:

- How discount vs. non-discount purchases differ
- Gender-based spending and product preferences
- Preferred payment methods (e.g., credit card, cash)
- How shipping types influence customer choices