Customer Shopping Analysis

A Portfolio Project in Data Analytics & Business Intelligence

Project Overview

This project examines customer shopping behavior using transactional data from 3,900 purchases across multiple product categories. The aim was to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

Dataset Summary

- **Size:** 3,900 rows × 18 columns
- **Key Features:**
    - Customer demographics: Age, Gender, Location, Subscription Status
    - Purchase details: Item, Category, Amount, Season, Size, Color
    - Shopping behavior: Discounts, Frequency, Review Ratings, Shipping Type
- **Data Quality:**
    - 37 missing values in *Review Rating*
    - Inconsistent column naming conventions

Methodology

The analysis began with a structured data preparation workflow implemented in Python:

- Data Loading: Imported the dataset into a Pandas DataFrame for processing.
- Initial Exploration: Applied df.info() to examine the schema and df.describe() to generate descriptive statistics.
- Missing Data Handling: Identified null values and imputed missing entries in the *Review Rating* column using the median rating within each product category.
- Column Standardization: Renamed all variables to snake_case format to ensure consistency and improve readability.
- Feature Engineering:
    - Constructed an age_group variable by binning customer ages into defined ranges.

    - Derived a purchase_frequency_days feature from transaction timestamps to capture shopping frequency.

- Data Consistency Check: Evaluated redundancy between discount_applied and promo_code_used; removed the latter to streamline the dataset.
- Database Integration: Established a connection to PostgreSQL and loaded the cleaned DataFrame, enabling structured SQL analysis for business insights.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discou Appli
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	Na
std	1125.977353	15.207589	NaN	NaN	NaN	23.685392	NaN	NaN	NaN	NaN	0.716983	NaN	NaN	Na
min	1.000000	18.000000	NaN	NaN	NaN	20.000000	NaN	NaN	NaN	NaN	2.500000	NaN	NaN	Na
25%	975.750000	31.000000	NaN	NaN	NaN	39.000000	NaN	NaN	NaN	NaN	3.100000	NaN	NaN	Na
50%	1950.500000	44.000000	NaN	NaN	NaN	60.000000	NaN	NaN	NaN	NaN	3.800000	NaN	NaN	Na
75%	2925.250000	57.000000	NaN	NaN	NaN	81.000000	NaN	NaN	NaN	NaN	4.400000	NaN	NaN	Na
max	3900.000000	70.000000	NaN	NaN	NaN	100.000000	NaN	NaN	NaN	NaN	5.000000	NaN	NaN	Na

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

Business Analysis (SQL)

- **Revenue by Gender:** Male customers contributed nearly twice the revenue of female customers.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

- **High-Spending Discount Users:** 839 customers used discounts yet spent above average.

	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	99

Total rows: 839 Query complete 00:00:

- **Top Products by Rating:** Gloves, Sandals, and Boots ranked highest in satisfaction.

	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

- **Top 3 products per category**: listed the most purchased product 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

- **Shipping Type Comparison:** Express shipping correlated with slightly higher purchase amounts.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

- **Subscribers vs. Non-Subscribers:** Non-subscribers generated more revenue overall, but subscribers showed stronger engagement.

	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

- **Discount-Dependent Products:** Hats, Sneakers, and Coats were most reliant on discounts.

	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:** Majority classified as *Loyal* (3,116 customers).
- **Repeat Buyers & Subscriptions:** Repeat buyers were more likely to subscribe.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

- **Revenue by Age Group:** Young Adults contributed the highest revenue (~$62K).

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

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

Visualization (Power BI)

Built an interactive dashboard to present insights:

- Revenue and sales breakdown by category, age group, and subscription status.
- Customer distribution across loyalty segments.
- Comparative analysis of shipping types and discount usage.
- Highlighted KPIs:
    - Average Purchase Amount: $59.76

- Average Review Rating: 3.75
- Subscribers vs. Non-Subscribers: 27% vs. 73%

Key Insights

- Subscription growth opportunity exists, as subscribers spend consistently but remain a minority.
- Loyalty programs could strengthen retention among the majority of loyal customers.
- Heavy reliance on discounts for certain products risks margin erosion.
- Top-rated items should be prioritized in marketing campaigns.
- Young Adults and Express Shipping users represent high-value segments.

Business Recommendations

1. Promote exclusive benefits to increase subscriptions.

2. Incentivize repeat buyers to transition into loyal customers.
3. Balance promotional activity with profitability.
4. Highlight best-rated and best-selling products in advertising.
5. Focus marketing spend on high-revenue age groups and premium shipping users.

Tools & Technologies

- Python (Pandas, NumPy) – Data cleaning & feature engineering
- PostgreSQL – Structured business queries
- Power BI – Interactive dashboards & visualization
- Excel – Supporting analysis and validation

Portfolio Value

This project demonstrates end-to-end analytics capability:

- **Data Engineering:** Cleaning, transformation, and database integration
- **Business Intelligence:** SQL-driven insights aligned with strategic goals
- **Visualization:** Professional dashboards for stakeholder communication
- **Impact Orientation:** Translating analytics into actionable recommendations