Customer Shopping Behavior Analysis

1. Project Overview

This project examines customer shopping behavior using transactional data from 3,900 purchases across multiple product categories. The objective is to identify spending habits, customer segments, product performances, and subscription trends, providing insights that support informed business decisions.

2. Data Summary

- Rows: 3,900
- Columns: 18
- 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 Types.
- Missing Data: 37 Values in Review Rating Column.

3. Exploratory Data Analysis using Python

We initiated the data preparation and cleaning process using Python:

- Data Loading: The dataset was imported using the **Pandas** library
- Initial Exploration: Used `df.info()` to review the dataset structure and `df.describe()` to generate summary statistics.
- Handling Missing Data: Checked the null values and handled missing data in the Review Rating column by imputing the median rating within each product category.
- Column Standardization: Renamed all column names to snake case to improve readability and ensure consistent documentation.
- Feature Engineering:
  - Created an age_group column by categorizing customer ages into defined ranges using binning.
  - Created purchase_frequency_days column by converting textual purchase intervals (eg: Weekly, Annually, Quarterly) into their corresponding numeric values in days.
- Data Consistency Check: Evaluated the relationship between discount_applied and promo_code_used, identified redundancy, and removed promo_code_used column.
- Database Integration: Connected the Python workflow to PostgreSQL and loaded the cleaned data frame into a database table for further SQL based analysis.

4. Data Analysis using SQL (Business Transactions)

Conducted structured analysis in PostgreSQL to address key business questions and extract and extract actionable insights.

1. Revenue by Gender: Analyzed and compared total revenue generated from male and female customers.

	gender text	Total Revenue Generated numeric
1	Female	75191
2	Male	157890

2. High-Spending Discount Customers: Identified customers who applied discounts yet recorded purchase amounts above the overall average spending.

	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

3. Top 5 Products by Rating: Identified the products with the highest average review rating.

	item_purchased text	Average Rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

4. Shipping Type Comparison: Evaluated and compared the average purchase amounts for Standard vs Express shipping.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

5. Subscribers vs Non-Subscribers: Compared both the average spending and total revenue generated by subscribed and non-subscribed customers.

	subscription_status text	Total Customers bigint	Average Spend numeric	Total Revenue numeric
1	No	2847	59.87	170436
2	Yes	1053	59.49	62645

6. Discount-Dependent Products: Identified five products with the highest proportion of purchases made using discounts.

	item_purchased text	percentage_of_purchases_with_discounts numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. Customer Segmentation: Categorized customers as New, Returning, or Loyal based on their purchase history and frequency.

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

8. Top 3 Products per Category: Identified the three most frequently purchased products 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. Repeat Buyers and Subscriptions: Assessed whether customers with more than five purchases were more likely to hold a subscription.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. Revenue by Age Group: Calculated the total revenue generated by each customer age group.

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

5. Power BI Dashboard

As the final step, an interactive Power BI dashboard was developed to visually present and summarize key insights.

6. Business Recommendations:

- Product Positioning: Highlight top-rated and best-selling products in marketing campaigns to drive higher conversions.
- Targeted Marketing: Concentrate marketing efforts on high-revenue age groups and customers who prefer express shipping.
- Boost Subscription Rates: Promote exclusive benefits to encourage more customers to subscribe.
- Customer Loyalty Programs: Introduce targeted rewards for repeat buyers to encourage their progression into the “Loyal” segment.
- Review Discount Policy: Balance sales growth from discounts with margin control to ensure profitability.