# Customer Shopping Behavior Analysis

### 1. Project Overview

This project examines customer purchasing patterns using transactional data from 3,900 orders across multiple product categories. The objective is to reveal insights into spending habits, customer segmentation, product preferences, and subscription trends to support informed business decision-making.

---

### 2. Dataset Summary

- **Total Records:** 3,900
- **Total Features:** 18
- **Key Attributes Include:**
    - Customer demographics: age, gender, location, subscription status
    - Transaction details: product purchased, category, amount spent, season, size, color
    - Behavioral indicators: discount usage, promo code use, previous purchases, buying frequency, review rating, shipping type
- **Missing Values:** 37 entries in *review_rating*

---

### 3. Exploratory Data Analysis (Python)

The analysis began with data cleansing and preparation using Python:

- **Data Import:** Loaded the dataset using pandas.
- **Initial Assessment:** Utilized df.info() and df.describe() to understand structure and statistical properties.

| | Customer ID | Age | Gender | Item Purchased | Category | Purchase Amount (USD) | Location | Size | Color | Season | Review Rating | Subscription Status | Shipping Type |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 55 | Male | Blouse | Clothing | 53 | Kentucky | L | Gray | Winter | 3.1 | Yes | Express |
| **1** | 2 | 19 | Male | Sweater | Clothing | 64 | Maine | L | Maroon | Winter | 3.1 | Yes | Express |
| **2** | 3 | 50 | Male | Jeans | Clothing | 73 | Massachusetts | S | Maroon | Spring | 3.1 | Yes | Free Shipping |
| **3** | 4 | 21 | Male | Sandals | Footwear | 90 | Rhode Island | M | Maroon | Spring | 3.5 | Yes | Next Day Air |
| **4** | 5 | 45 | Male | Blouse | Clothing | 49 | Oregon | M | Turquoise | Spring | 2.7 | Yes | Free Shipping |

| Discount Applied | Promo Code Used | Previous Purchases | Payment Method | Frequency of Purchases |
|---|---|---|---|---|
| Yes | Yes | 14 | Venmo | Fortnightly |
| Yes | Yes | 2 | Cash | Fortnightly |
| Yes | Yes | 23 | Credit Card | Weekly |
| Yes | Yes | 49 | PayPal | Weekly |
| Yes | Yes | 31 | PayPal | Annually |

- **Handling Missing Data:** Filled missing review ratings using category-wise median ratings.
- **Standardization:** Converted column names into snake_case for improved consistency.
- **Feature Engineering:**
    - Added an *age_group* variable by categorizing age ranges.
    - Derived *purchase_frequency_days* from purchase timestamps.
- **Quality Checks:** Assessed redundancy between *discount_applied* and *promo_code_used*; removed *promo_code_used*.
- **Database Integration:** Uploaded the cleaned dataset into PostgreSQL for SQL-driven analytics.

---

**4. SQL-Based Business Analysis**

Using MySQL, the following business questions were explored:

1. **Revenue by Gender:** Compared total sales generated by male and female customers.

| | gender | revenue |
|---|---|---|
| ▶ | Male | $ 157.89k |
| | Female | $ 75.19k |

2. **High-Spending Discount Users:** Identified customers who availed discounts yet exceeded the average spending threshold.

| customer_id | purchase_amount |
|---|---|
| 2 | 64 |
| 3 | 73 |
| 4 | 90 |
| 7 | 85 |
| 9 | 97 |
| 12 | 68 |
| 13 | 72 |
| 16 | 81 |
| 20 | 90 |
| 22 | 62 |
| 24 | 88 |

3. **Top 5 Products by Rating:** Ranked products based on average customer reviews.

| item_purchased | Average_product_rating |
|---|---|
| Gloves | 3.86 |
| Sandals | 3.84 |
| Boots | 3.82 |
| Hat | 3.8 |
| Handbag | 3.78 |

4. **Shipping Type Comparison:** Analyzed differences in average purchase amounts between various shipping types.

| shipping_type | amount |
|---|---|
| Express | $ 60.48 |
| Free Shipping | $ 60.41 |
| Next Day Air | $ 58.63 |
| Standard | $ 58.46 |
| 2-Day Shipping | $ 60.73 |
| Store Pickup | $ 59.89 |

5. **Subscriber vs. Non-Subscriber Behavior:** Compared spending patterns and revenue contribution.

| subscription_status | Total_customers | average_spending | total_revenue |
|---|---|---|---|
| Yes | 1053 | 59.49 | 62645 |
| No | 2847 | 59.87 | 170436 |

6. **Discount-Heavy Products:** Identified products most frequently bought with discounts.

| item_purchased | discount_rate |
|---|---|
| Hat | 50.00 % |
| Sneakers | 49.66 % |
| Coat | 49.07 % |
| Sweater | 48.17 % |
| Pants | 47.37 % |

7. **Customer Segmentation:** Classified customers as New, Returning, or Loyal based on purchase history.

| customer_segment | Number_of_customers |
|---|---|
| Loyal | 3116 |
| Returning | 701 |
| New | 83 |

8. **Top 3 Products per Category:** Determined best-selling items 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 |

9. **Repeat Purchase & Subscription Trends:** Checked if customers with more than five purchases are more likely to subscribe.

| subscription_status | repeat_buyers |
|---|---|
| Yes | 958 |
| No | 2518 |

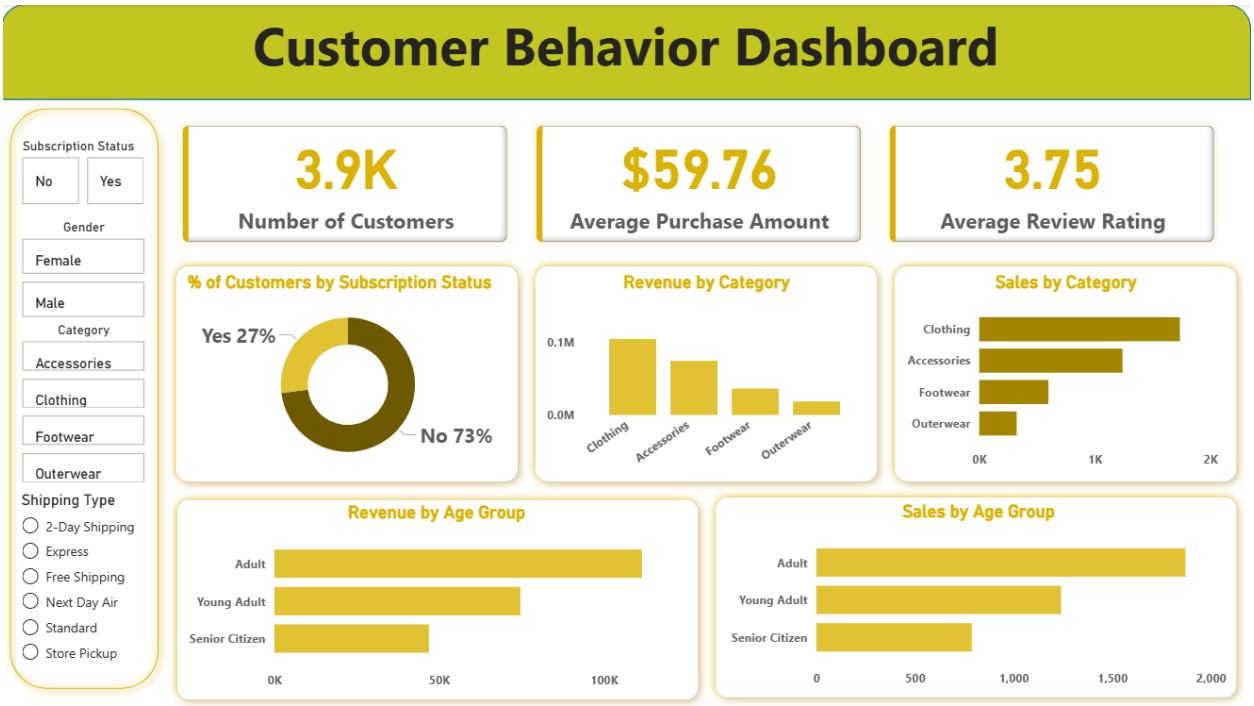
10. **Revenue by Age Group:** Calculated the contribution of each age segment to total revenue.

| age_group | revenue |
|---|---|
| Adult | $ 111.53k |
| Young Adult | $ 74.66k |
| Senior Citizen | $ 46.89k |

---

**5. Power BI Dashboard**

An interactive Power BI dashboard was developed to visually showcase trends, comparisons, and customer behavior insights, enabling faster decision-making.




---

**6. Business Recommendations**

- **Increase Subscription Adoption:** Highlight exclusive subscriber benefits to drive conversions.
- **Strengthen Loyalty Initiatives:** Incentivize frequent buyers to transition into the Loyal segment.
- **Optimize Discount Strategy:** Reassess discount campaigns to balance sales growth with profitability.
- **Enhance Product Marketing:** Promote high-rated and best-selling items more prominently.
- **Targeted Campaigns:** Focus marketing efforts on age groups and customer sets that deliver higher revenue, especially express-shipping customers.