Customer Behavior Analysis Report

1. Introduction

This report presents an end-to-end analysis of customer shopping behavior based on the provided dataset and SQL analyses. The objective is to understand purchasing patterns, customer segments, revenue drivers, and the impact of subscriptions, discounts, shipping types, and demographics on overall revenue. The insights are intended to support business decision-making, marketing strategies, and customer retention efforts.

2. Dataset Overview

- **Total records:** ~3,900 customers

- **Key attributes:** Age, Gender, Category, Item Purchased, Purchase Amount, Review Rating, Subscription Status, Shipping Type, Discount Applied, Previous Purchases, Frequency of Purchases, and derived features (Age Group, Purchase Frequency in Days).

- **Data quality:**

  o No missing values across most columns.

  o Review Rating had limited missing values, which were imputed using the **median rating within each product category**, preserving category-level trends.

3. Data Preparation & Feature Engineering

The following preprocessing steps were performed:

- Standardized column names (lowercase, underscores).

- Renamed purchase_amount_(usd) to purchase_amount for SQL compatibility.

- Created **Age Groups** using quartiles:

  o Young Adult

  o Adult

  o Middle-aged

  o Senior

- Converted **purchase frequency** categories (Weekly, Monthly, Quarterly, etc.) into numeric values representing days between purchases.

- Removed redundant column promo_code_used as it perfectly matched discount_applied.

These steps ensured clean, analysis-ready data for both Python-based exploration and SQL querying.

4. Key Dashboard Metrics (High-Level KPIs)

From the dashboard summary:

- **Average Purchase Amount:** $59.76

- **Average Review Rating:** 3.75

- **Total Customers:** ~4,000

- **Subscribers:** ~27%

- **Non-Subscribers:** ~73%

This indicates a large customer base with moderate spending and generally positive (but improvable) customer satisfaction.

5. Revenue Analysis

5.1 Revenue by Gender

- Total revenue was calculated separately for **male** and **female** customers.

- This comparison helps identify which gender segment contributes more to overall sales and can guide targeted marketing campaigns.

Insight: Both genders contribute significantly to revenue, indicating the product mix appeals broadly rather than being gender-skewed.

5.2 Revenue by Age Group

- Revenue contribution was analyzed across age groups.

- **Middle-aged** and **Young Adult** customers generated the highest revenue.

Insight: These age groups are the primary revenue drivers and should be the focus of loyalty programs and personalized offers.

5.3 Revenue by Category

- Clothing emerged as the **top revenue-generating category**, followed by Accessories and Footwear.

Insight: Inventory planning and promotional campaigns should prioritize Clothing, while Footwear shows potential for growth.

6. Subscription Analysis

6.1 Spending Behavior: Subscribers vs Non-Subscribers

- Metrics compared:
    - Number of customers
    - Average purchase amount
    - Total revenue

Key Finding:

- **Subscribed customers spend more on average** and contribute higher total revenue per customer compared to non-subscribers.

Business Implication:

- Increasing subscription adoption can directly boost revenue.
- Subscription-based incentives (exclusive discounts, faster shipping) are likely effective.

6.2 Repeat Buyers and Subscription Likelihood

- Customers with **more than 5 previous purchases** were analyzed.
- A higher proportion of repeat buyers were found among **subscribed customers**.

Insight: Subscription status and customer loyalty are strongly linked. Encouraging early subscription could increase long-term retention.

7. Discount & Promotion Analysis

7.1 High-Value Discount Users

- Customers who used discounts **and** spent more than the average purchase amount were identified.

Insight: Discounts are not only attracting low spenders; they also successfully incentivize high-value customers.

7.2 Products with Highest Discount Usage

- Top 5 products were ranked by **percentage of purchases with discounts applied**.

Insight:

- Certain products rely heavily on discounts to drive sales.

- These products may be price-sensitive or highly competitive.

Recommendation:

- Review pricing strategy for these items.

- Consider bundling instead of frequent discounting to protect margins.

8. Product & Category Insights

8.1 Top-Rated Products

- Products were ranked by **average review rating**.

Insight:

- Highly rated products represent strong quality perception and can be promoted as "best sellers" or "customer favorites."

8.2 Most Purchased Products by Category

- Top 3 products were identified within each category using window functions.

Insight:

- Clear category leaders exist.

- These items should be prioritized for stock availability and featured placements.

9. Shipping Type Analysis

- Average purchase amounts were compared between **Standard** and **Express** shipping.

Finding:

- Customers choosing **Express shipping** tend to have a slightly higher average purchase value.

Business Insight:

- Faster shipping appeals to high-intent buyers.

- Offering free or discounted express shipping for premium customers could increase basket size.

10. Customer Segmentation

Customers were segmented based on previous purchases:

- **New:** 1 purchase

- **Returning:** 2–10 purchases

- **Loyal:** More than 10 purchases

Observation:

- The majority fall into **Returning** customers.

- Loyal customers, though fewer, contribute disproportionately to revenue.

Recommendation:

- Introduce tiered loyalty rewards to convert returning customers into loyal ones.

11. Overall Business Insights & Recommendations

1. **Focus on Subscriptions:** Subscribers spend more and show higher loyalty.

2. **Target Key Age Groups:** Young Adults and Middle-aged customers are the main revenue drivers.

3. **Optimize Discounts:** Use discounts strategically on high-impact products without eroding margins.

4. **Promote Top Products:** Highlight top-rated and most-purchased items to increase conversions.

5. **Leverage Fast Shipping:** Express shipping correlates with higher spending—use it as a value-added incentive.

12. Conclusion

This analysis demonstrates how combining Python-based data preprocessing with SQL-driven business queries provides deep insight into customer behavior. The findings clearly show that **subscriptions, customer loyalty, and targeted product strategies** play a critical role in maximizing revenue. With data-driven refinements in marketing, pricing, and customer engagement, the business can significantly improve both profitability and customer satisfaction.