Customer Shopping Dataset – Detailed Analysis Report

1. Introduction

This report presents an in-depth analysis of the **Customer Shopping Dataset**, focusing on customer demographics, product preferences, purchase behavior, discount impact, and location-based trends.
The goal is to extract actionable insights that can support decision-making in marketing, sales strategies, and retail optimization.

---

2. Dataset Overview

The dataset contains information related to customer purchases, including:

- **Customer ID**
- **Gender**
- **Location**
- **Item Purchased**
- **Category**
- **Purchase Amount (USD)**
- **Review Rating**
- **Discount Applied**
- **Promo Code Used**
- **Purchase Date**
- **Last Purchase Date**

After cleaning, the dataset contained no missing values in key columns, and all date fields were converted into proper datetime format.

---

3. Data Cleaning Summary

The following preprocessing steps were performed:

✔ Handled Missing Values

- Found missing dates → cleaned and converted to `NaT`.
- No missing values found in essential fields afterward.

✔ Added Two New Columns

1. **`days_since_last_purchase`** – difference between purchase date and previous purchase.
2. **`purchase_month`** – extracted month from the purchase date for time-based analysis.

✔ Standardized Data Types

- Converted numeric fields to integers/floats.
- Converted date fields to datetime.

✔ Filtered Invalid Entries

- Removed rows with negative purchase amounts.
- Standardized category names.

---

4. Exploratory Data Analysis (EDA)

4.1 Customer Demographics

- The dataset contains a balanced mix of genders.
- Locations vary widely, suggesting a diverse customer base.
- Some locations appear more frequently, indicating stronger market penetration.

4.2 Product Analysis

- **Top Selling Products**
  Calculated using product frequency.
- **Highest Rated Products**
  Using average review rating per item.

Key Insight:

Products with higher review ratings correlate with higher purchase frequency.

---

5. Discount & Promotion Analysis

5.1 Customers Using Discounts

Discounts were analyzed through:

- % of orders with **Discount Applied = Yes**
- % of orders using **Promo Code**

Findings:

- Some products have disproportionately high discount usage.
- Discounts significantly increase purchase frequency for selected categories.

5.2 Products With Highest Discount Usage

Calculated using:

```
(group of Discount Applied = "Yes") / (total purchases per product)
```

These items show the strongest dependency on discounts and may require pricing strategy review.

---

6. Location-Based Insights

6.1 Distinct Locations

Extracted using:

- `nunique()` in Pandas
- `DISTINCT Location` in SQL or Power BI

6.2 Top 5 Performing Locations (Power BI)

To get top 5 locations by **Purchase Amount**, use:

DAX:

```
TopLocations =
TOPN(
    5,
    SUMMARIZE(Customer, Customer[Location], "TotalSales",
SUM(Customer[Purchase Amount (USD)])),
    [TotalSales],
    DESC
)
```

Key Insight:

A small number of locations contribute to the majority of revenue.

---

7. Revenue Analysis

7.1 Total Revenue

Calculated by summing all purchase amounts.

7.2 Average Purchase Value

Helps understand customer spending behavior.

7.3 Category-Wise Revenue

Some categories consistently outperform others, indicating strong demand.

---

8. Review Rating Analysis

8.1 Highest Average Rating – Top 5 Products

Calculated using:

```
groupby('Item Purchased')['Review Rating'].mean().nlargest(5)
```

8.2 Rating Distribution

Review ratings are mostly positive, but a few products show low ratings, indicating potential quality issues.

---

9. Customer Behavior Insights

9.1 Repeat Purchase Behavior

Using the newly created `days_since_last_purchase`, we observe:

- Shorter intervals represent loyal, frequent buyers.
- Longer intervals may require retargeting campaigns.

9.2 Purchase Timing Patterns

Analysis shows customer activity peaks in certain months/seasons.

---

10. Key Insights & Recommendations

📌 Insights

- Top 5 locations generate a major share of revenue.
- Discounts significantly boost purchases for select items.
- Highly rated products are also best-selling.
- Customers with shorter repeat intervals are valuable for loyalty programs.

📌 Recommendations

1. **Increase inventory** for best-selling, high-rated products.
2. **Optimize discount strategy** by reducing unnecessary promotions on products that sell without discounts.
3. **Run targeted campaigns** in high-performing locations.
4. **Improve product quality** or customer experience for low-rated items.
5. **Launch loyalty programs** for repeat customers.

---

11. Conclusion

This analysis highlights important trends in customer shopping behavior and provides insights to improve marketing strategy, product management, and business growth. With these findings, the organization can make more data-driven decisions and enhance overall customer satisfaction.