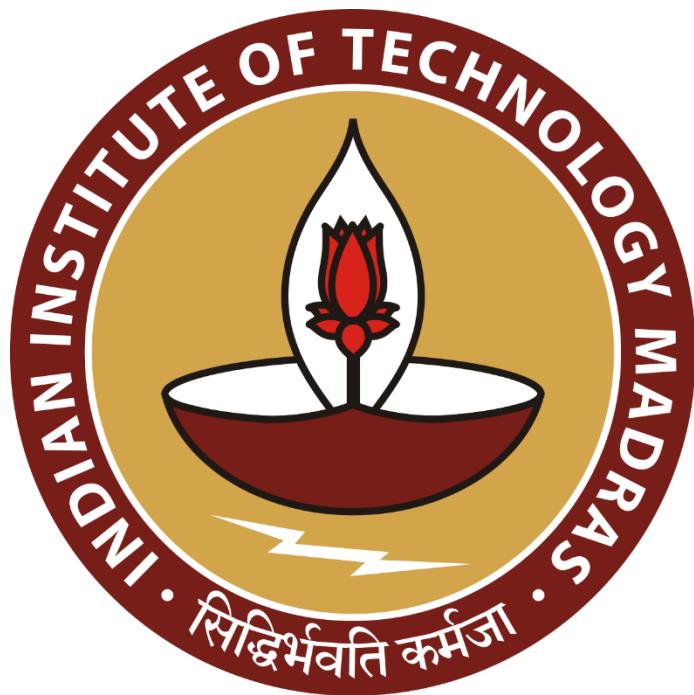


## **Business Data Analysis of E-commerce Sales**

Submitted by

Name: HARMEET KAUR

Roll number: 23F2001610



IITM Online BS Degree Program,  
Indian Institute of Technology, Madras, Chennai  
Tamil Nadu, India, 600036

## **Table of Contents**

<b>1. Executive Summary .....</b>	<b>3</b>
<b>2. Proof of Originality .....</b>	<b>4</b>
2.1. Declaration of Original Work	
2.2. Data Source and Acknowledgment	
<b>3. Meta Data and Descriptive Statistics.....</b>	<b>5</b>
<b>4. Detailed Explanation of Analysis Process / Method .....</b>	<b>7</b>
4.1. Method for Financial Overview	
4.2. Method for Financial Overview	
4.3. Profit Margin Analysis	
<b>5. Results and Findings .....</b>	<b>15</b>
1. Category-Wise Distribution Analysis	
2. Payment Method Distribution	
3. Discount and Pricing Relationship	
4. Discount and Profit Relationship	
5. Final Price Distribution	
6. Time-Based Sales Insights	
<b>6. Interpretation of Results and Recommendations .....</b>	<b>23</b>
6.1. Interpretation of Results	
6.2. Business Recommendations	
<b>7. Presentation and legibility of report.....</b>	<b>29</b>

## **1. Executive Summary**

This project presents a comprehensive Business Data Analysis of an E-commerce transactional dataset with the objective of evaluating sales performance, pricing effectiveness, customer behavior, and overall profitability across multiple product categories. The business currently faces key challenges related to aggressive discounting, negative profit margins, and misalignment between pricing strategy and cost structure.

Before analysis, the dataset was cleaned by checking for missing values, validating date formats, correcting derived temporal variables (Month, Weekday, Year), and ensuring numerical consistency across Price, Discount, and Final Price columns. Derived variables such as Profit and Profit Margin were computed programmatically to avoid manual inconsistencies.

Exploratory and financial analyses were conducted using Python (Pandas and Matplotlib) and validated with Excel to ensure consistency between code outputs and reported results. Key analyses included category-wise sales distribution, profit margin evaluation, payment method preference, discount–price relationships, discount–profit relationships, final price distribution, and time-based sales trends.

The results reveal that although the business generates substantial revenue, it operates at an overall loss with an average profit margin of  $-18.83\%$ , primarily due to discounts being applied directly on cost price without a base markup. Clothing and Electronics emerge as the strongest revenue-contributing categories, while Credit Cards dominate payment preferences. Moderate discounts increase sales volume but consistently erode profitability.

Key recommendations include introducing a base markup before discounting, implementing category-specific discount limits, optimizing inventory toward high-performing categories, leveraging digital payment incentives, and adopting data-driven pricing and forecasting strategies to ensure long-term financial sustainability.

## 2. Proof of originality

### Declaration of Original Work

I, HARMEET KAUR, hereby certify that the analysis, interpretation, model development, and all narrative content presented in this Business Data Management project report are my original work. All methodologies, visualizations, and conclusions derived from the dataset are the exclusive product of my own research and effort, unless explicitly cited otherwise.

This report relies heavily on the provided dataset, which serves as the foundational data input. The preparation, cleaning, transformation, and management of this raw data, as detailed in the methodology section, constitute the core original work of this project.

### Data Source and Acknowledgment

The transactional data used for this project, titled "E-Commerce Dataset," was sourced from a publicly accessible third-party repository. The raw data structure has been preserved for the sake of integrity, and all analysis is based exclusively on the provided CSV file.

#### Dataset Source Details:

Attribute	Detail
<b>Dataset Name</b>	E-Commerce Dataset(Transactional Data)
<b>Data Creator</b>	Steve Rogg
<b>Platform</b>	Kaggle
<b>URL</b>	<a href="https://www.kaggle.com/datasets/steve1215rogg/e-commerce-dataset">https://www.kaggle.com/datasets/steve1215rogg/e-commerce-dataset</a>
<b>Licence</b>	Shared under the CCO: Public Domain license
<b>Dataset Link:</b>	<a href="https://www.kaggle.com/datasets/steve1215rogg/e-commerce-dataset">https://www.kaggle.com/datasets/steve1215rogg/e-commerce-dataset</a>
<b>Excel Working Sheet:</b>	 BDM
<b>Google Colab (Analysis Code):</b>	 BDM PLOTS  BDM PLOT2

The data is used solely for educational and non-commercial project purposes. By using this publicly available dataset, I acknowledge the original creator and the platform hosting the data.

### 3. Meta Data Table

Attribute Name	Data Type	Description
User_ID	Categorical	Unique identifier assigned to each customer
Product_ID	Categorical	Unique identifier for each product
Category	Categorical	Product category such as Clothing, Electronics,etc.
Price(Rs.)	Numerical	Original price of the product, treated as cost price
Discount(%)	Numerical	Percentage discount applied to the product
Final Price(Rs.)	Numerical	Selling price after discount is applied
Payment Method	Categorical	Mode of Payment used (UPI, Credit Card etc.)
Purchase_Date	Date	Date on which transaction is made
Month	Categorical(Derived)	Month extracted from the purchase date
Weekday	Categorical(Derived)	Day of the week extracted from the purchase date
Year	Categorical(Derived)	Year extracted from the purchase date

### Descriptive Statistics

Descriptive statistics were used to summarize the key numerical variables in the dataset and understand their central tendency and variability. Metrics such as mean, minimum, maximum, and standard deviation were computed for Price (Rs.), Discount (%), and Final Price (Rs.). This analysis provides insights into the overall pricing structure, discount behavior, and transaction value distribution before performing profitability and financial analysis.

```
df[["Price (Rs.)", "Discount (%)", "Final_Price(Rs.)"]].describe()
```

Fig(1) Descriptive Statistics using python library

The `describe()` function was used to compute descriptive statistics for key numerical variables. It provides measures of central tendency, dispersion, and range, enabling a preliminary understanding of pricing, discount behavior, and final transaction values.

	<b>Price(Rs.)</b>	<b>Discount(%)</b>	<b>Final_Price(Rs.)</b>
mean	254.80	18.82	206.90
std	141.68	14.73	122.68
min	10.09	0.00	5.89
25%	134.01	5.00	104.51
50%	253.84	15.00	199.18
75%	377.59	25.00	304.11
max	499.96	50.00	496.82

Matplotlib library was used for plotting graphs for easier data visualization.

Frequency vs Variables

<b>Metric</b>	<b>Freq vs Price(Rs.)</b>	<b>Freq vs. Discount (%)</b>	<b>Freq vs. Final Price (Rs.)</b>
Mean (Average)	254.80	18.83	206.91
Median (Middle)	253.85	15.00	199.19

Std. Deviation	141.68	14.73	122.69
Minimum	10.09	0.00	5.89
Maximum	499.96	50.00	496.82
Skewness	-0.005 (Highly Symmetric)	0.77(Right-Skewed)	0.24 (Slightly Symmetric)
Distribution Shape	<b>Uniform/Flat:</b> Items are evenly distributed across the price range.	<b>Clustered:</b> Most discounts are at lower tiers (0-15).	<b>Normal-like:</b> The spread peaks slightly toward the lower-middle range.

### Relationship Statistics

Plot Name	Correlation (r)	Type of Relationship
Discount vs. Price	0	None / Independent
Final Price vs. Discount	-0.3115	Weak Negative

## 4. Detailed Explanation of Analysis Process/Method

### 4.1 Method for Overall Financial Summary

The company's consolidated financials reveal the following key statistics:

- Total Revenue Estimation

Total Revenue is calculated by summing the Final Price(Rs.) of all transactions. It represents the actual sales value realized after discounts are applied.

$$\text{Total Revenue} = \Sigma \text{Final Price (Rs.)}$$

- Total Cost Estimation

Total cost is estimated by aggregating the original Price (Rs.) of all products, considered as the cost price. This reflects the total expenditure incurred on goods sold.

$$\text{Total Cost} = \Sigma \text{Price (Rs.)}$$

- Total Profit Calculation

Total profit is obtained by subtracting total cost from total revenue. This indicates the overall financial gain or loss across all transactions.

$$\text{Total Profit} = \text{Total Revenue} - \text{Total Cost}$$

Category	Price (Rs.)	Discount (%)	Final_Price(Rs.)	Profit	Profit Margin	MONTH
Sports	36.53	15	31.0	=D2-B2	-15.00136874	December
Clothing	232.79	20	186.23	-46.56	-20.00085914	November
Sports	317.02	25	237.76	-79.26	-25.00157719	October
Toys	173.19	25	129.89	-43.3	-25.0014435	November
Beauty	244.8	20	195.84	-48.96	-20	November

Fig(2) Formula applied in excel sheet to calculate profit

- Average Profit per Sale

Average profit per sale is calculated by dividing total profit by the total number of transactions. It shows the average profitability of each sale.

$$\text{Average Profit per Sale} = \text{Total Profit} \div \text{Number of Transactions}$$

Calculations done in Analysis:

$$\text{Total Profit(Rs.)} = \text{Total Revenue(Rs.)} - \text{Total Cost(Rs.)} = \text{Total Profit(Rs.)}$$

$$= 757,278.08 - 932,570.47$$

$$= -175,292.39$$

$$\text{Number of Transactions} = 3660$$

$$\text{So, Average Profit per Sale(Rs.)} = -175,292.39 / 3660$$

$$= -47.89$$

- Average Profit Margin

Average profit margin measures profitability relative to cost and is expressed as a percentage. It evaluates how efficiently costs are converted into profit.

$$\text{Profit Margin (\%)} = (\text{Total Profit} \div \text{Total Cost}) \times 100$$

Calculations done in Analysis (For Beauty Category):

Total Profit(Sum of Profit)= -23417.94

Total Cost(SUM of Price (Rs.))= 127633.04

Profit Margin (%) =  $(-23417.94 \div 127633.04) * 100$

$$= -18.47544187$$

AVERAGE of Profit Margin= (Sum of Profit Margin across all categories) / Number of categories

$$= -131.7583814 / 7$$

$$= -18.83 \text{ (approx)}$$

Metric	Values
Total Revenue (Rs.)	757,278.08
Total Cost (Rs.)	932,570.47
Total Profit (Rs.)	-175,292.39
Average Profit Per Sale (Rs.)	-47.89
Average Profit Margin (%)	-18.83

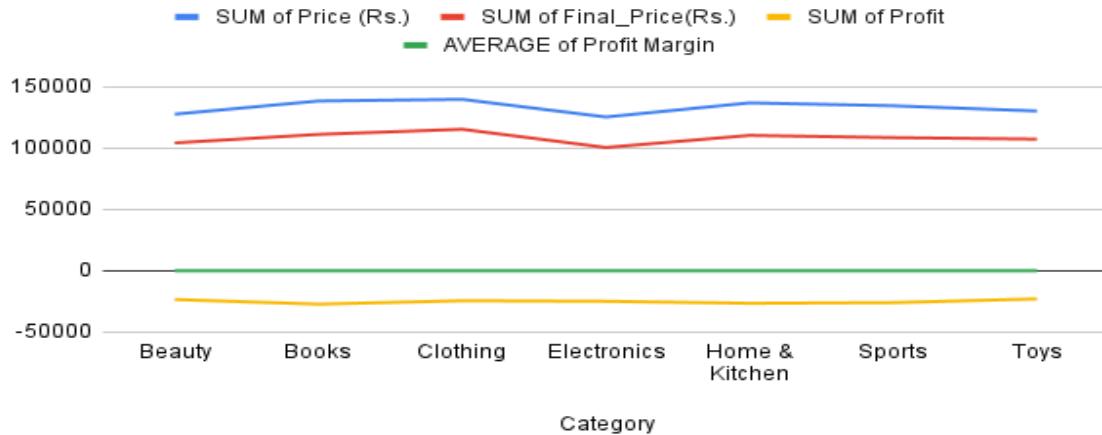
The figures clearly indicate that the business is operating at a **loss**, with an average negative profit margin of approximately **18.83 %**. This suggests that products are being sold below cost, most likely due to aggressive discounting or misaligned pricing strategies.

Category	SUM of Price (Rs.)	SUM of Final_Price(Rs.)	SUM of Profit	AVERAGE of Profit Margin
Beauty	127633.04	104215.1	-23417.94	-18.47544187
Books	138371.71	111149.35	-27222.36	-19.03530586
Clothing	139714.55	115314.84	-24399.71	-17.91905313
Electronics	125320.14	100462.23	-24857.91	-19.26674076
Home & Kitchen	136841.51	110328.08	-26513.43	-19.60847392
Sports	134481.99	108518.79	-25963.2	-19.32670685
Toys	130207.53	107289.69	-22917.84	-18.12665898
<b>Grand Total</b>	<b>932570.47</b>	<b>757278.08</b>	<b>-175292.39</b>	<b>-18.82513333</b>

Fig(3) Calculations of sum of price, final price, profit and average of profit margin in excel

A pivot table was constructed to calculate the category wise financial summary

### SUM of Price (Rs.), SUM of Final\_Price(Rs.), SUM of Profit and AVERAGE of Profit Margin



Fig(4) Data visualization of above calculation using line chart

X axis: Category    Y axis: Financial Summary values

```
[15] ❶ import pandas as pd
      df = pd.read_csv("bdmcsv.csv")
      total_revenue = df["Final_Price(Rs.)"].sum()
      total_cost = df["Price (Rs.)"].sum()
      total_profit = total_revenue - total_cost
      average_profit_per_sale = total_profit / len(df)
      average_profit_margin = (total_profit / total_cost) * 100

[16] ❷ print(f"Total Revenue: {total_revenue:.2f}")
      print(f"Total Cost: {total_cost:.2f}")
      print(f"Total Profit: {total_profit:.2f}")
      print(f"Average Profit per Sale: {average_profit_per_sale:.2f}")
      print(f"Average Profit Margin: {average_profit_margin:.2f}%")

... Total Revenue: ₹757278.08
Total Cost: ₹932570.47
Total Profit: ₹-175292.39
Average Profit per Sale: ₹-47.89
Average Profit Margin: -18.80%
```

Fig(5) Python lib panda

Python library, panda was used which allows direct operations like summation, averaging, filtering, and grouping on entire columns and reduces the chances of manual errors and is significantly faster than using loops or basic Python lists.

#### 4.2. Method for Financial Overview

To address the challenge of providing a comprehensive financial overview and sales analysis for the dataset under study, a systematic approach was undertaken.

This involved calculating various key performance indicators to gain insights into the business's sales performance, pricing strategy, and customer purchasing behavior across different product categories and payment modes.

- **Total Cost and Pricing Calculation:**

1. **Base Price (Price in Rs.):**

This represents the original price or cost value associated with each product before any discounts are applied. It reflects the base market value from which profit margins and discounts are derived.

2. **Discount (%):**

The discount percentage indicates the promotional or seasonal reductions offered on each product. This value directly impacts customer buying patterns and overall revenue generation.

3. **Final Price (Rs.):**

The final selling price is calculated after applying the discount on the base price using the formula:

$$\text{Final Price} = \text{Price (Rs.)} \times (1 - \text{Discount (\%)} / 100)$$

This parameter provides a realistic measure of actual sales revenue per transaction.

User_ID	Product_ID	Category	Price (Rs.)	Discount (%)	Final_Price(Rs.)
337c166f	f414122f-e	Sports	36.53	1?	=D2*(1-E2/100)
d38a19bf	fde50f9c-5	Clothing	232.79	20	186.23
d7f5f0b0	0d96fc90-3	Sports	317.02	25	237.76
395d4994	964fc44b-d	Toys	173.19	25	129.89
a83c145c	d70e2fc6-e	Beauty	244.8	20	195.84

Fig (6) Formula computed in excel sheet to calculate Final\_Price(Rs.)

- **Revenue and Sales Performance:**

1. **Revenue Generation:**

The total revenue was derived by summing up all *Final Price (Rs.)* values, representing the total earnings from customer purchases.

Comparative analyses were made between *Category*, *Payment Method*, and *Time-based* attributes (Month, Weekday, and Year) to identify trends in sales and consumer

preferences.

## 2. Category-wise Analysis:

The dataset includes multiple product categories such as *Electronics, Clothing, Beauty, Toys, and Sports*.

Distribution across categories was studied to determine which segment contributes most to revenue and which requires strategic improvement.

```
df["Profit"] = df["Final_Price(Rs.)"] - df["Price (Rs.)"]
df["Profit_Margin"] = (df["Profit"] / df["Price (Rs.)"]) * 100
category_margin = (
    df.groupby("Category")["Profit_Margin"]
    .mean()
    .sort_values()
)
plt.figure(figsize=(8, 5))
category_margin.plot(kind="bar")
plt.xlabel("Product Category")
plt.ylabel("Average Profit Margin (%)")
plt.title("Category-wise Average Profit Margin")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Fig(7) Category-wise Average Profit Margin

Here, the dataset is grouped by product category using `groupby()`. The average profit margin for each category is calculated using `.mean()`, and the results are sorted using `.sort_values()` for better comparison. The bar chart is then plotted for data visualization.

## 3. Payment Mode Analysis:

Various payment methods—such as *UPI, Credit Card, Debit Card, and Net Banking*—were analyzed to identify the most commonly used and customer-preferred modes of transaction.

```

import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv("bdmcsv.csv")

# Create payment method distribution chart
payment_counts = df['Payment_Method'].value_counts()

plt.figure(figsize=(6, 6))
payment_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90, colors=plt.cm.Paired.colors)

plt.title('Payment Method Distribution', fontsize=14)
plt.ylabel('') # Remove default y-label
plt.tight_layout()
plt.show()

```

Fig(8) Python library Pandas was used to load the dataset, process categorical data, and compute frequency counts for analysis.

The data then was visualized though a pie chart using the python library matplotlib

- **Time-Based Sales Evaluation:**

1. Monthly Trends:

Purchase dates were converted into structured time components (*Month*, *Weekday*, *Year*) to analyze how sales fluctuated throughout the year.

This enabled identification of peak business months and low-performance periods.

Purchase_Date	Purchase_Date	YEAR	MONTH	WEEKDAY
12/11/2024	2024-12-11	? =YEAR(I2)	December	Wednesday
9/2/2024	2024-09-02	2024	September	Monday
1/9/2024	2024-01-09	2024	January	Tuesday
1/4/2024	2024-01-04	2024	January	Thursday

Fig(9) Year, Month, Weekday extracted from purchase date

The purchase date was first converted to date format using a format tool, then year, month, weekday were extracted using the functions, YEAR(date), MONTH(date),

WEEKDAY(date).

```
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv("bdmcs.csv")

# Convert 'Purchase_Date' to datetime
df['Purchase_Date'] = pd.to_datetime(df['Purchase_Date'])

# Extract time components
df['Year'] = df['Purchase_Date'].dt.year
df['Month'] = df['Purchase_Date'].dt.month_name()
df['Weekday'] = df['Purchase_Date'].dt.day_name()
```

Fig(10)Python program to extract year, month and weekday from dates.

Purchase dates were converted from string format to datetime format using Pandas to enable time-series based evaluation.

## 2. Customer Behavior Patterns:

Patterns in purchase frequency and payment methods across weekdays and weekends were examined to assist in designing future marketing and pricing strategies.

### 4.3. Profit Margin Analysis

Since the listed price is treated as the cost price, the absence of a markup implies that any level of discounting directly results in negative profit margins. Profit margin analysis serves as a critical indicator of a company's financial health and operational efficiency. It measures how much profit is earned from each unit of cost, helping businesses identify whether their pricing and discounting strategies are sustainable. In this project, **Price (Rs.)** has been considered as the **Cost Price**, and **Final\_Price (Rs.)** as the **Selling Price** after discounts. Profit and margin are derived using the following formulas:

$$\text{Profit (Rs.)} = \text{Final Price (Rs.)} - \text{Price (Rs.)}$$

This analysis aims to assess the profitability of various product categories and understand the relationship between discount rates and profitability, ensuring that the business maintains a balance between competitive pricing and sustainable profit.

The profit margin analysis reveals a significant insight into the company's pricing and discounting strategy.

## 5. Results and Findings

To derive meaningful insights from the dataset, various visual and statistical analyses were conducted to interpret the financial and sales performance across key business parameters.

This section focuses on evaluating *category-wise performance, payment behavior, discount strategy impact, and price distribution* to provide a clear understanding of customer trends and business profitability.

### 1. Category-Wise Distribution Analysis

The product data was grouped by the **Category** attribute to examine the relative contribution of each product type to total revenue.

The **Clothing** segment emerged as the **highest-earning category**, followed by **Electronics** and **Beauty** products.

This indicates a diverse product portfolio, with fashion-related goods dominating overall sales volume and profitability.

#### a. Category wise Count of Products

The category-wise distribution of products shows a fairly balanced dataset.

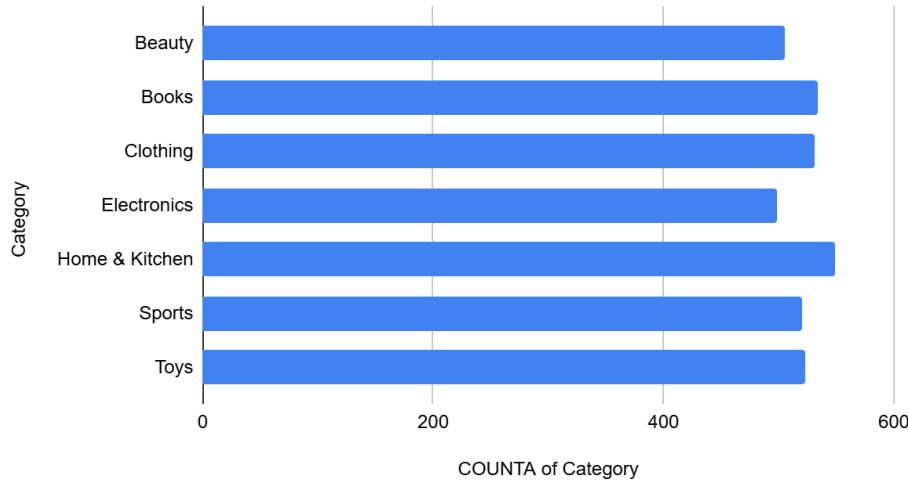
Category	COUNT of Category
Beauty	505
Books	534
Clothing	531
Electronics	498
Home & Kitchen	549
Sports	520
Toys	523
<b>Grand Total</b>	<b>3660</b>

- **Key Observation:**

**Home & Kitchen** is the "Mode" of this dataset. This category has the highest volume of

transactions. However, the gap between the most popular category (Home & Kitchen: 549) and the least popular (Electronics: 498) is only 51 units. This indicates a **low standard deviation** in category volume. The business is diversified; this a general retailer with equal traction across all seven segments.

COUNTA of Category vs. Category



Fig(11) Category Distribution using bar graph

#### b. Category wise Overall financial Summary

Overall Financial Summary across different categories were computed using a pivot table.

Category	SUM of Price (Rs.)	SUM of Final_Price(Rs.)	SUM of Profit	AVERAGE of Profit Margin
Beauty	127633.04	104215.1	-23417.94	-18.47544187
Books	138371.71	111149.35	-27222.36	-19.03530586
Clothing	139714.55	115314.84	-24399.71	-17.91905313
Electronics	125320.14	100462.23	-24857.91	-19.26674076
Home & Kitchen	136841.51	110328.08	-26513.43	-19.60847392
Sports	134481.99	108518.79	-25963.2	-19.32670685
Toys	130207.53	107289.69	-22917.84	-18.12665898
<b>Grand Total</b>	<b>932570.47</b>	<b>757278.08</b>	<b>-175292.39</b>	<b>-18.82513333</b>

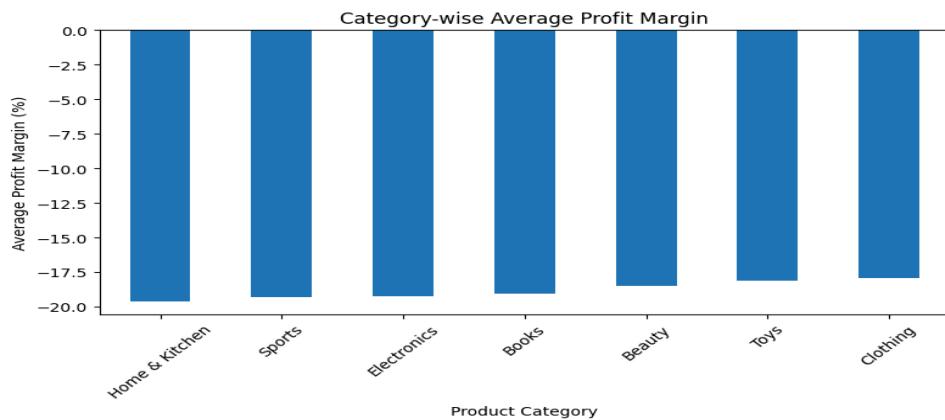
Fig(12) Financial summary (same as shown in the previous section)

Key Observation:

- The **Grand Total Profit** is **-₹175,292.39**, which means the business is operating at an overall loss.
- The **average profit margin** is **-18.83%**, indicating that for every ₹100 in sales, nearly ₹19 is being lost.
- Every category (Beauty, Books, Clothing, Electronics, Home & Kitchen, Sports, Toys) shows **negative profit** and **negative profit margins**. This suggests a **systemic issue**, not a problem limited to one specific category.
- **Home & Kitchen** has the **highest loss (-₹26,513.43)** and the **worst average margin (-19.61%)**. **Sports (-₹25,963.20)** and **Electronics (-₹24,857.91)** also contribute significantly to total losses.
- **Toys** have the **lowest loss (-₹22,917.84)**.
- Clothing** shows the **least negative margin (-17.92%)**, making it comparatively better than others, though still unprofitable.
- In all categories, the **Final Price is significantly lower than the original Price**, which directly leads to losses. This indicates **heavy discounting, high operational costs, or pricing inefficiencies**.

### c. Category Wise Margin Analysis

The **category-wise margin chart** shows that some product categories perform relatively better, though still not achieving positive profitability. This suggests that certain product lines may have more favorable cost structures or lower discount rates.



Fig(13) Category wise Margin visualization using bar graph

Bar chart was constructed to visualize quantitative data, X axis: Product Category, Y axis: Average Profit Margin(%)

## 2. Payment Method Distribution

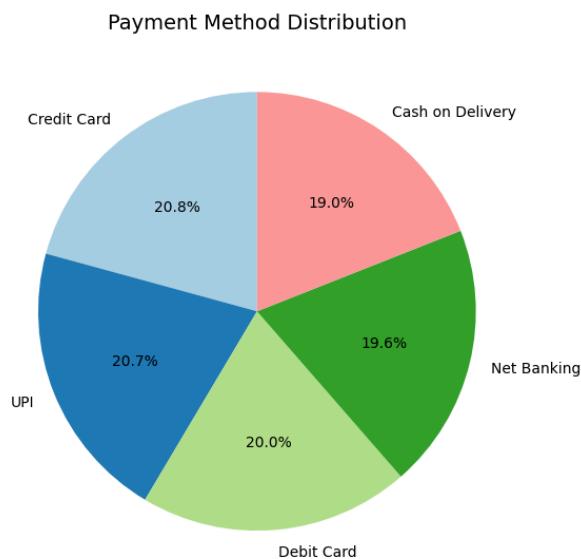
An analysis of the **Payment Method** field revealed distinct customer preferences in transactional behavior.

Among all payment types—*UPI, Credit Card, Debit Card, and Net Banking*—the **Credit Card** option was most frequently used.

This suggests that customers favor flexible payment methods and may respond positively to credit-linked promotions or cashback offers.

- **Key Interpretation:**

The dominance of Credit Card usage indicates a digitally active customer base, highlighting the importance of maintaining seamless and secure online payment systems. Encouraging alternative digital payment modes like UPI could further improve transaction convenience and broaden accessibility.



Fig(14) Pie Chart showing most preferred mode of payment

A pie chart was used to visualize categorical data and shows Credit Card(20.8%) being the most preferred payment method.

### 3. Discount and Pricing Relationship

To evaluate the impact of discounts on sales performance, the relationship between Discount (%) and Final Price (Rs.) was examined.

The Discount (%) vs Price (Rs.) analysis reveals that discounts are applied in fixed slabs (0%–50%) irrespective of product price. This indicates a uniform discounting strategy that does not consider product cost, margin structure, or category-specific

profitability. When linked with the profit analysis, this approach helps explain the persistent negative profit margins across all categories and months. Despite stable final sales values, profits remain consistently negative, particularly during high-discount months such as August to November. The absence of price-sensitive or margin-based discounting results in revenue erosion without sufficient compensation through increased sales volume. Furthermore, while moderate discounts (15–25%) are effective in sustaining sales volume, they fail to generate positive profits under the current pricing structure. The frequent use of deep discounts (30%–50%), intensifies losses, especially in categories with higher costs such as Electronics, Home & Kitchen, and Sports.

### **Key Insight:**

While moderate discounts help sustain sales volume, the analysis shows that they **do not ensure positive profit margins under the current pricing structure**. This highlights the importance of balancing promotional strategies with profitability objectives rather than relying solely on discount-driven sales.

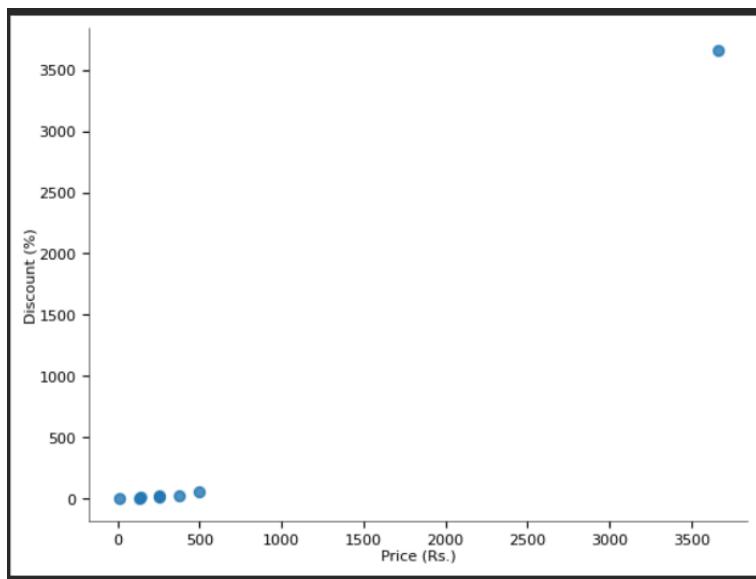


Fig (15) Scatter plot plotted using python library matplotlib

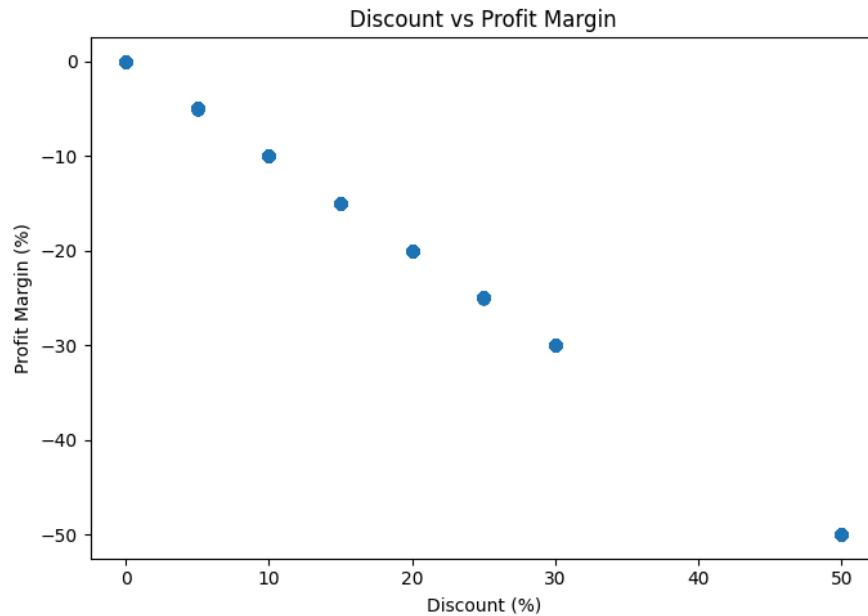
The scatter plot shows no strong correlation between price and discount percentage. Most products cluster at lower price levels with moderate discounts, while an extreme outlier indicates a possible data error. This suggests inconsistent discounting practices, which may contribute to overall losses.

### **4. Discount and Profit Relationship**

To evaluate the impact of discounts on profitability, the relationship between **Discount (%)** and **Profit Margin (%)** was examined.

The analysis revealed a clear negative correlation, indicating that profit margins decline steadily as discount levels increase.

Higher discount rates significantly reduce profitability, demonstrating that aggressive discounting erodes margins when not supported by cost reductions or increased customer acquisition.



Fig(16) Scatter plot between Discount(%) and Profit Margin(%) plotted using Matplotlib

The **scatter plot between discount and profit margin** demonstrates a clear negative correlation as discounts increase, profit margins decrease sharply. This trend highlights that while discounting might be used to drive sales volume, it adversely impacts profitability when not paired with a corresponding reduction in costs or increase in customer acquisition.

### Key insights

- Profit margins show a strong negative relationship with discount levels, confirming that higher discounts directly reduce overall profitability.
- Moderate discounting helps sustain sales volume but does not translate into improved profit margins.
- Excessive discounts lead to sharp margin erosion, indicating that discount-driven sales growth is financially unsustainable without parallel cost optimization.

## 5. Final Price Distribution

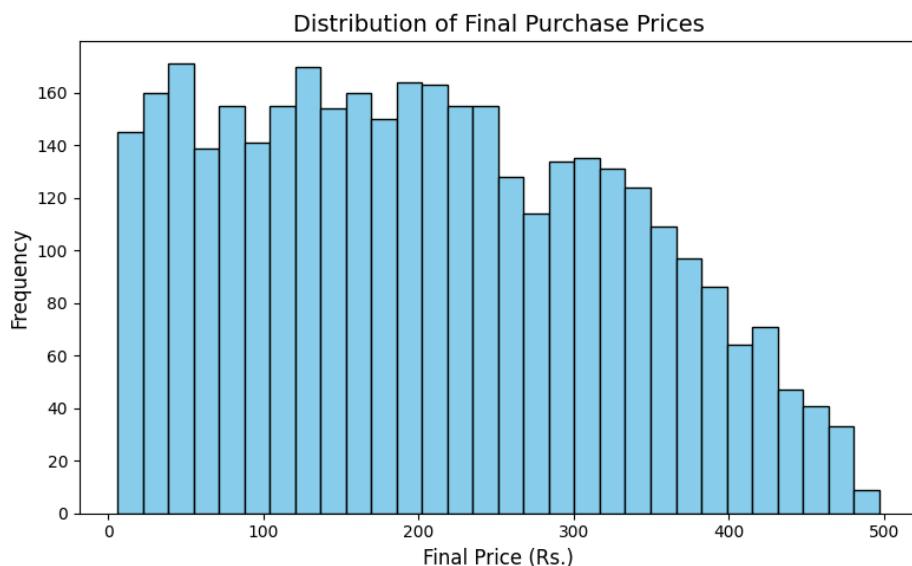
The **Final\_Price(Rs.)** distribution exhibited a concentration of sales in the low-to-mid price range (₹100–₹300), with fewer high-value transactions.

This pattern suggests that the business primarily caters to mass-market segments rather than premium consumers.

The steady demand in this price bracket reflects affordability-driven purchasing behavior and price-conscious customers.

- **Key Interpretation:**

Strategic product diversification toward slightly higher-value offerings may help in expanding revenue without losing affordability appeal.



Fig(17) Histogram showing Distribution of final purchase prices

This histogram illustrates the distribution of final purchase prices after discounts. Most transactions are concentrated in the ₹100–₹300 range, indicating strong customer preference for mid-priced products, while higher price ranges show declining purchase frequency due to price sensitivity. Histogram was plotted as it is easy to visualize the distribution and frequency of final purchase prices, helping identify common price ranges and customer buying patterns.

## 6. Time-Based Sales Insights

Using the **Purchase\_Date**, **Month**, and **Weekday** columns, temporal sales trends were analyzed.

The month of **July** demonstrated peak sales performance, indicating a strong mid-year buying pattern, possibly due to seasonal promotions or festival-related shopping.

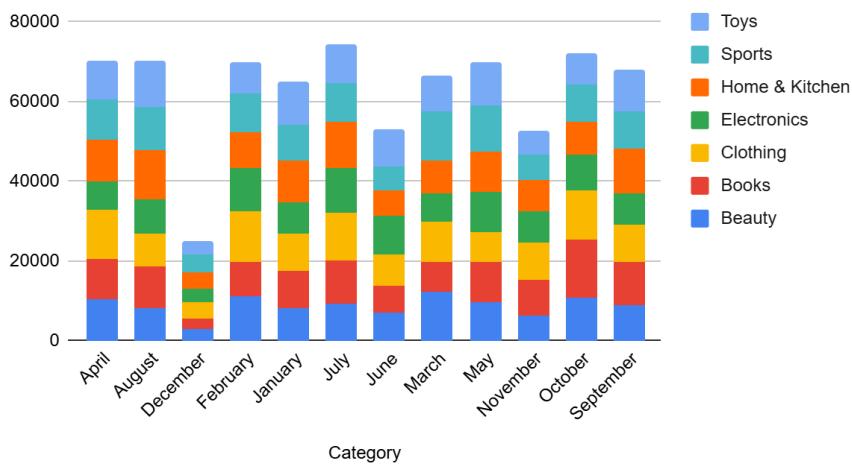
Sales were found to be relatively stable throughout weekdays, suggesting consistent consumer activity across the week.

Monthly sales across various categories were also analyzed to analyze which category was sold most in the peak month.

SUM of Profit	MONTH										
Category	April	August	December	July	June	May	November	October	September	Grand Total	
Beauty	-492.83	-2824.26	-2712.51	-153.39	-1933.47	-1627.28	-4365.62	-5207.38	-4101.2	-23417.94	
Books	-737.17	-3576.65	-2448.99	-1455.01	-3039.48	-1337.31	-4945.21	-5137.04	-4545.5	-27222.36	
Clothing	-246.99	-3123.83	-2636.99	-824.07	-1210.03	-1141.47	-5350.33	-5717.41	-4148.59	-24399.71	
Electronics		-3070.03	-2150.4	-1868.21	-2555.04	-1351.4	-4315.45	-5414.91	-4132.47	-24857.91	
Home & Kitchen	-249.52	-3885.51	-2764.24	-985.41	-1017.21	-2295.38	-5965.71	-4663.03	-4687.42	-26513.43	
Sports	-1481.04	-2638.86	-2627.45	-1683.4	-1967.64	-2053.78	-4558.46	-5354.41	-3598.16	-25963.2	
Toys	-495.22	-2633.1	-2835.4	-505.37	-1194.05	-2239.35	-5029.41	-4508.7	-3477.24	-22917.84	
<b>Grand Total</b>	<b>-3702.77</b>	<b>-21752.24</b>	<b>-18175.98</b>	<b>-7474.86</b>	<b>-12916.92</b>	<b>-12045.97</b>	<b>-34530.19</b>	<b>-36002.88</b>	<b>-28690.58</b>	<b>-175292.39</b>	

Fig(18) Pivot table constructed in excel sheet

Beauty, Books, Clothing, Electronics, Home & Kitchen...



Fig(19) Line chart constructed for monthly revenue analysis across all categories

The chart illustrates monthly final sales contributions of different product categories using a stacked bar format. It shows that total sales remain relatively stable across months, with multiple categories contributing consistently, while noticeable dips and peaks indicate seasonal variations.

- **Key Observation:**

Maximum revenue over all months: Clothing

Minimum revenue over all months: Electronics

Identifying high-traffic periods such as July can assist in inventory planning, targeted marketing, and discount optimization.



Fig(20) Line Chart representing monthly sales trend

This line chart shows the monthly sales trend, with sales remaining relatively stable between ₹65,000 and ₹72,000 for most of the year. A noticeable decline occurs in June and a sharp drop is observed in December, indicating seasonal fluctuations and possible year-end demand or operational constraints.

## 6. Interpretation of Results and Recommendations

The analytical process carried out in this project generated a comprehensive overview of the company's financial performance, sales distribution, and customer behavior.

Through systematic data cleaning, visualization, and interpretation, several key business insights were identified that highlight both strengths and improvement areas within the organization's sales and pricing strategies.

### 1. Interpretation of Results

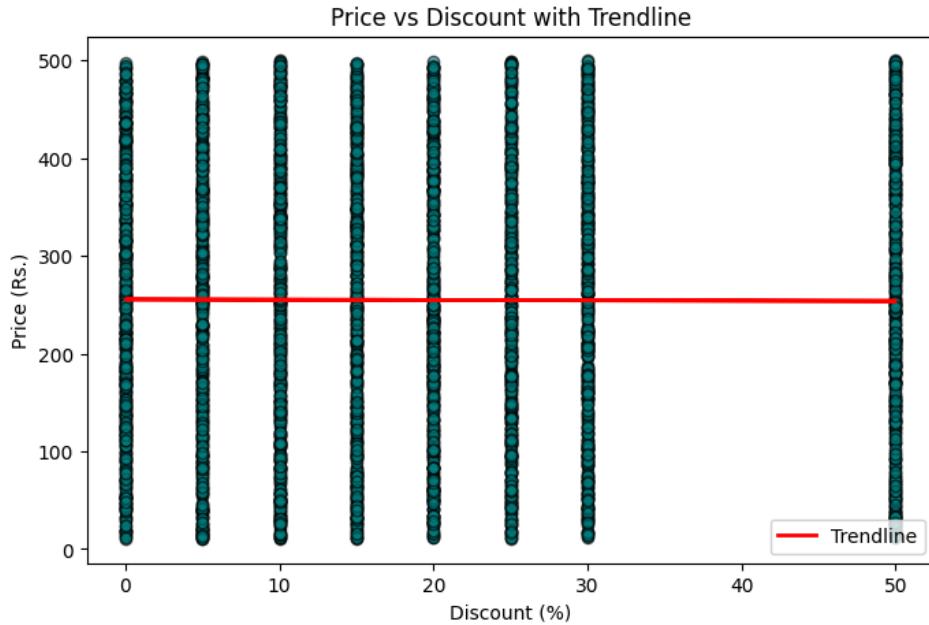
The analysis of the dataset reveals several key patterns that offer valuable insights into business performance, consumer behavior, and pricing efficiency.

### a. Sales and Revenue Trends

- The **total revenue generated (₹7,57,278.08)** reflects an unhealthy business scale, the cost price being higher than the revenue earned.
- The **Clothing category** recorded the **highest revenue (₹1,15,314.84)**, highlighting it as a core segment that drives profitability.
- Other categories such as *Electronics* and *Beauty* contribute steadily, indicating product diversification that helps stabilize overall performance.
- The **month of July** emerged as the **peak sales period**, suggesting seasonal demand influenced by promotional campaigns or festive buying patterns.

### b. Pricing and Discount Insights

- The business offers an average discount of approximately 18.83%; however, this level of discounting is associated with an overall negative profit margin. This indicates that while discounts may attract customers and sustain sales volume, they currently undermine profitability.
- Higher sales volumes do not translate into improved profit margins, highlighting the need to reassess discount effectiveness from a profitability perspective.
- Excessive discounts beyond this percent yield diminishing returns, implying that aggressive markdowns do not always translate into higher profits.



Fig(21) Price vs Discount Correlation Tradeline

This scatter plot shows the relationship between discount percentage and product price, with prices spread uniformly across all discount levels. The nearly flat trendline indicates no significant correlation between price and discount, suggesting that discounts are applied independently of product pricing.

#### c. Customer Behavior and Purchase Pattern

- The **average final purchase price (₹206.91)** suggests a mid-range pricing strategy targeting cost-conscious customers.
- Most customers prefer **affordable products** within ₹100–₹300, supporting the idea that the brand appeals to a broad audience segment.
- **Credit Card** transactions dominate, followed by UPI, showing a **digitally active customer base** that prefers cashless and secure payment methods.
- The data also suggests **steady sales across weeks**, implying consistent customer engagement rather than sporadic or event-driven spikes.

#### d. Profitability and Performance Efficiency

- The relationship between Price (Rs.), Discount (%), and Final Price (Rs.) reveals inefficiencies in the current discount strategy, as aggressive discounting results in selling prices falling below cost.
- The **Profit Margin analysis** suggests that while discounts drive volume, Profitability analysis indicates that discounting must be strategically controlled and aligned with cost structures, as uncontrolled discounts consistently lead to margin erosion.

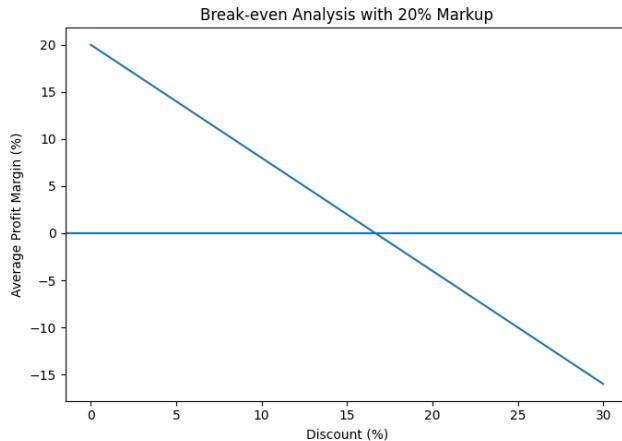
## **2. Business Recommendations**

Based on the analytical insights, the following recommendations are proposed to improve overall business efficiency, profitability, and customer engagement.

### **a. Product and Inventory Management**

- **Focus on High-Performing Categories:**  
Allocate more marketing resources and inventory to top-selling segments like **Clothing** and **Electronics**, as they significantly contribute to revenue.
- **Reevaluate Low-Performing Categories:**  
Conduct surveys or product reviews to understand the low traction in *Toys* and *Sports*, and introduce targeted offers or product bundling to increase demand.

### **b. Pricing and Discount Strategy**



```

import pandas as pd

df = pd.read_csv("bdmcsv.csv")
df.columns = df.columns.str.strip().str.replace(" ", "_")

# Function to simulate profit margin with markup
def simulate_profit_margin(markup, discount):
    marked_price = df["Price_(Rs.)"] * (1 + markup / 100)
    final_price = marked_price * (1 - discount / 100)
    profit_margin = ((final_price - df["Price_(Rs.)"]) / df["Price_(Rs.)"]) * 100
    return profit_margin.mean()

# Test different discount levels at 20% markup
for d in [5, 10, 15, 18, 20]:
    print(f"Avg Profit Margin at {d}% discount:", round(simulate_profit_margin(20, d), 2))

Avg Profit Margin at 5% discount: 14.0
Avg Profit Margin at 10% discount: 8.0
Avg Profit Margin at 15% discount: 2.0
Avg Profit Margin at 18% discount: -1.6
Avg Profit Margin at 20% discount: -4.0

```

Fig(22a top) Negative relationship exists between discount levels and profit margins.

The break-even point occurs at approximately 17–18% discount, beyond which discounts result in negative margins and financial losses.

Fig(22b bottom) Python library pandas used to evaluate average profit margin(%)

- The business operates at an **average profit margin of -18.83%**, indicating that products are sold below cost after discounts due to the absence of a base markup.
- A **strong negative relationship** exists between discount levels and profit margins; higher discounts significantly reduce profitability, even if sales volume is sustained.
- Reducing discounts alone is insufficient to restore profitability, as any non-zero discount leads to losses under the current pricing structure.
- To achieve sustainable profits, the company should introduce a **base markup of approximately 20%** before applying discounts, allowing moderate discounts (up to 15%) without margin erosion.

- **Category-specific discounting**, minimum margin thresholds, and **dynamic pricing strategies** should be implemented to prevent below-cost sales.
- Personalized loyalty discounts can be used selectively to improve customer retention while maintaining financial sustainability.

#### c. Customer Engagement and Retention

- **Digital Payment Incentives:**  
Since Credit Card usage dominates, provide cashback, reward points, or EMI offers to encourage repeat purchases.
- **Personalized Campaigns:**  
Segment customers based on purchase frequency and average spending to send tailored marketing offers.
- **Feedback Integration:**  
Implement post-purchase surveys to gather insights for improving product quality and service satisfaction.

#### d. Marketing and Sales Planning

- **Seasonal Promotions:**  
Capitalize on the **July peak sales trend** by launching planned mid-year and festival campaigns.
- **Omnichannel Marketing:**  
Use social media, e-commerce platforms, and targeted email marketing to boost visibility of underperforming categories.
- **Cross-Selling Opportunities:**  
Promote complementary products (e.g., Electronics accessories with devices, Beauty kits with skincare items) to increase average order value.

#### e. Data-Driven Decision Making

- **Regular Data Monitoring:**  
Maintain dashboards to track sales, discounts, and profits in real time for quick business decisions.

- **Predictive Analytics:**

Use future data analysis techniques like forecasting models to anticipate sales trends and manage inventory efficiently.

- **Business Intelligence Integration:**

Automate reporting systems using tools like Power BI or Tableau to visualize insights dynamically.

## 7. Presentation and Legibility of report

[https://www.canva.com/design/DAG3i61AYUc/NYb4SnNy8WzANaA2Bd6zsA/edit?utm\\_content=DAG3i61AYUc&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAG3i61AYUc/NYb4SnNy8WzANaA2Bd6zsA/edit?utm_content=DAG3i61AYUc&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

Google sheet:

<https://docs.google.com/spreadsheets/d/1Z9bhjbYq8N3Q6zIUWStoyscwvWj-ME3n18l1haZOz8Y/edit?usp=sharing>

Google Colab:

<https://colab.research.google.com/drive/1AB2mBYnsVQOOIUCs32SgEQad9CI4qWB3?usp=sharing>

<https://colab.research.google.com/drive/1ydCuVdmK4yH0mk-Fj5pbnBNzNSfFxXYU?usp=sharing>

