

SUPER MARKET SALES ANALYSIS REPORT

Problem Statements:

◆ 1. Sales & Revenue

1. *How can the supermarket increase overall revenue while maintaining healthy profit margins across branches?*
 2. *Which product lines contribute the most and least to sales and profit, and how can underperforming categories be improved?*
 3. *What are the key drivers (unit price, quantity, customer type) influencing high-value transactions?*
-

◆ 2. Customer Analytics

4. *Do Members spend more per transaction than Normal customers, and is the loyalty program driving profitability?*
 5. *How do customer ratings vary by demographics (gender, age, income), and what does this imply about customer satisfaction?*
 6. *Which customer segment (based on age, income, type) contributes most to repeat business and long-term value?*
 7. *How can we improve satisfaction for customers giving ratings below 6?*
-

◆ 3. Product Line Performance

8. *Which product lines should the supermarket prioritize for promotions to maximize profitability?*
 9. *Are there product lines that attract younger vs. older customers, and how can marketing be targeted accordingly?*
 10. *Is there a relationship between product line ratings and sales volume?*
-

◆ 4. Branch & Regional Strategy

11. *Which branch is underperforming in terms of sales, profit, and customer satisfaction, and why?*

12. *Do customer preferences differ significantly across cities (Yangon, Mandalay, Naypyitaw), and should product mix be adjusted locally?*
 13. *How does footfall (number of transactions) compare across branches, and what operational changes can balance demand?*
-

◆ **5. Payment & Financial Analytics**

14. *Which payment methods are most popular, and do they correlate with higher spend or satisfaction?*
 15. *How can the supermarket encourage digital payments (Ewallet, Credit card) among low-income and younger customers?*
 16. *Does payment method influence gross income or transaction value?*
-

◆ **6. Time-based Analytics**

17. *What are the peak hours and days of sales, and how can staffing and inventory be optimized accordingly?*
 18. *Do sales vary significantly between weekdays and weekends, and how should marketing campaigns be aligned?*
 19. *Is there a seasonal trend in sales that can guide inventory planning and promotional campaigns?*
-

◆ **7. Strategic / Predictive**

20. *Can we build a predictive model to forecast monthly revenue for each branch?*
21. *Can we segment customers into groups (high spenders, budget shoppers, frequent visitors) for targeted marketing?*
22. *How can we predict customer ratings based on purchase behavior and demographics?*
23. *What is the likelihood of Normal customers converting into Members, and what strategies would improve conversion?*

Data

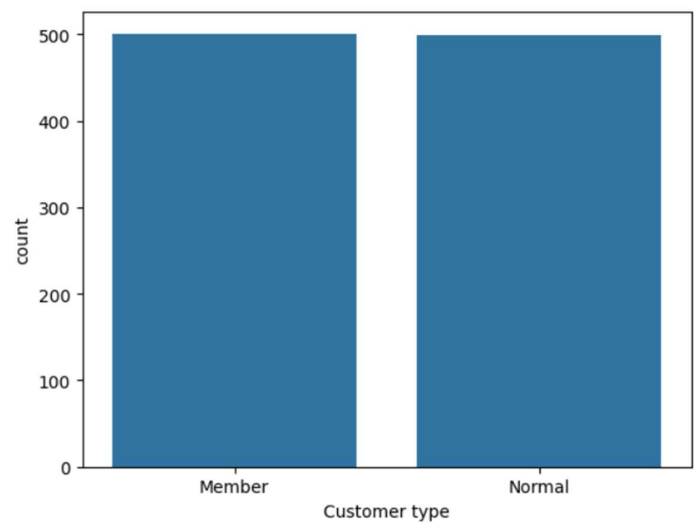
Columns & Their Meaning

1. **Invoice ID** – Unique transaction identifier.
2. **Branch** – Store branch (A, B, or C).
3. **City** – Location (Yangon, Naypyitaw, Mandalay).
4. **Customer type** – Member or Normal.
5. **Gender** – Male or Female.
6. **Product line** – Category of products sold (e.g., Health and beauty, Electronics, Food and beverages).
7. **Unit price** – Price of one unit of the product.
8. **Quantity** – Number of items purchased.
9. **Tax 5%** – 5% VAT charged.
10. **Total** – Final amount including tax.
11. **Date** – Date of purchase.
12. **Time** – Time of purchase.
13. **Payment** – Payment method (Ewallet, Cash, Credit card).
14. **COGS** – Cost of goods sold (before tax).
15. **Gross margin percentage** – Always 4.7619% in this dataset.
16. **Gross income** – Profit earned on the transaction.
17. **Rating** – Customer's rating (1–10 scale).
18. **Age**
19. **Income Level**

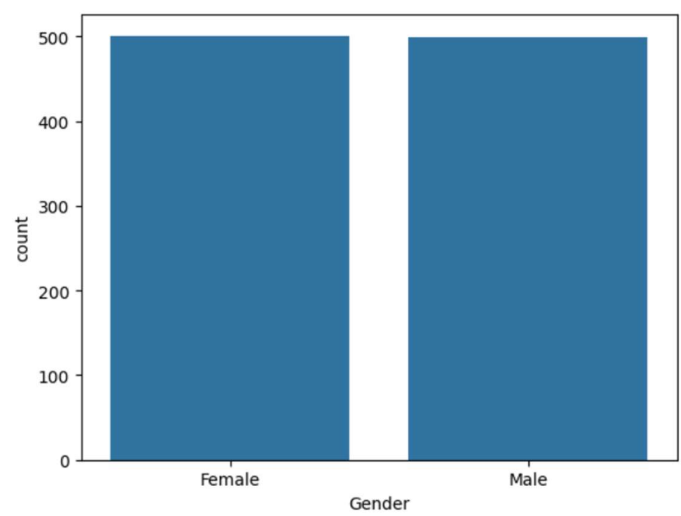
Exploratory Data Analysis:

Customer Analytics:

Customer Type Segmentation -

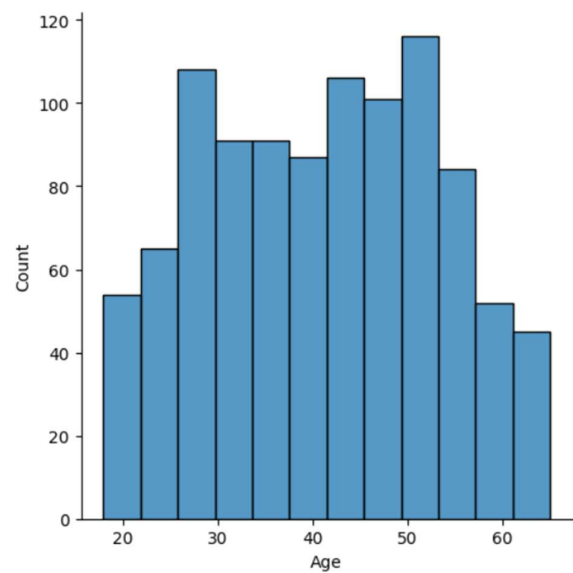


Demographic Distribution –



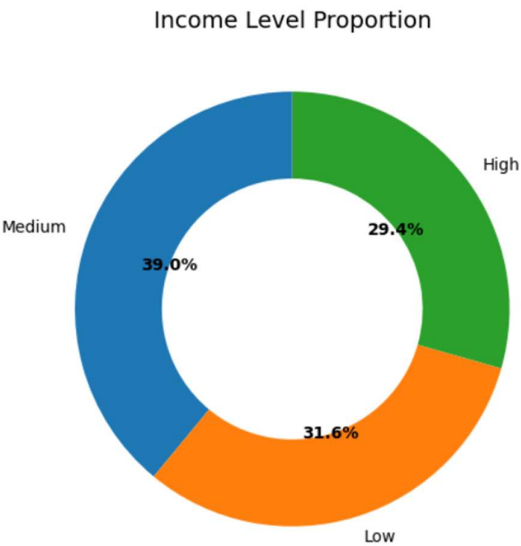
It seems like there are equal amounts of male and female.

Age Distribution –



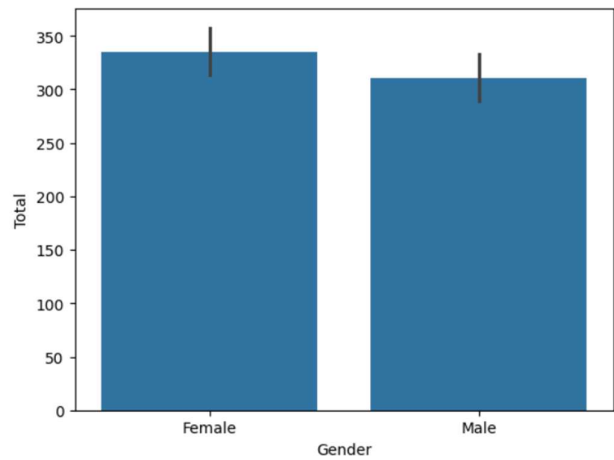
The Age distribution between 40 to 50 is higher.

Income Level Distribution –

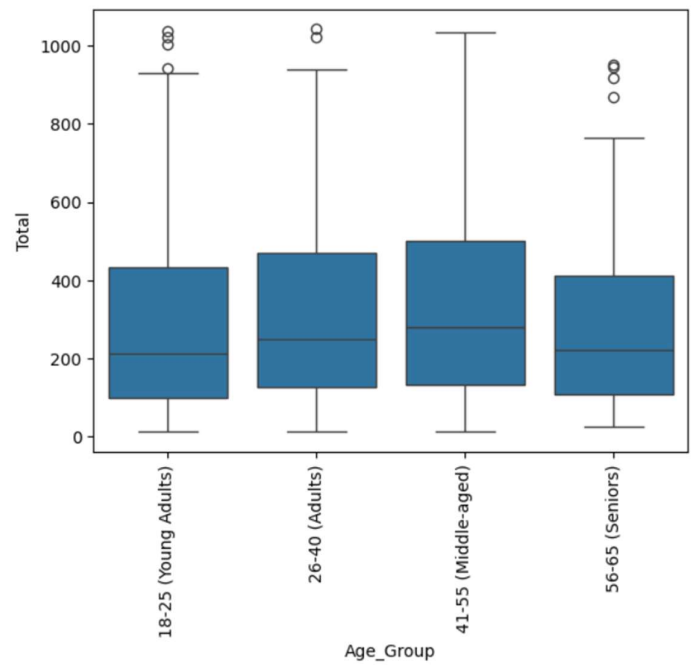


Sales by Demographics

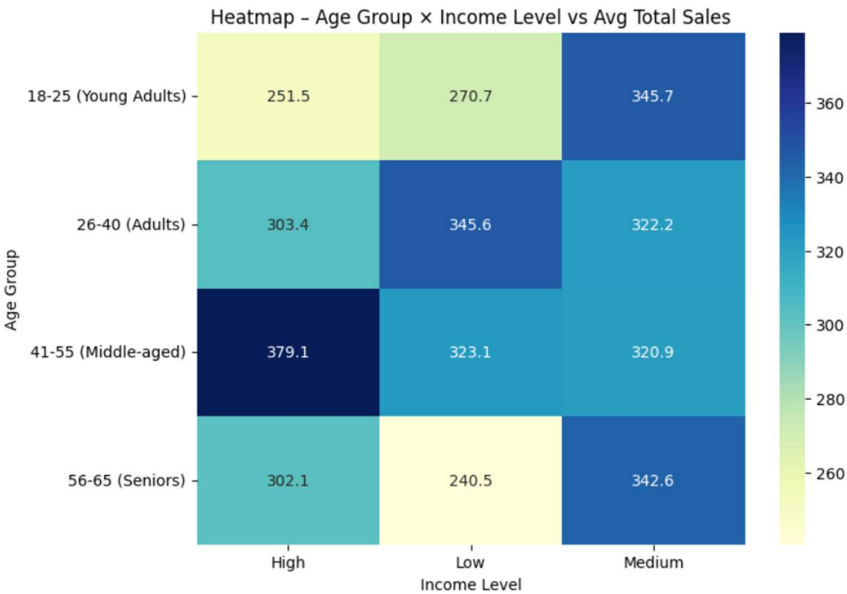
Total Sales by Gender



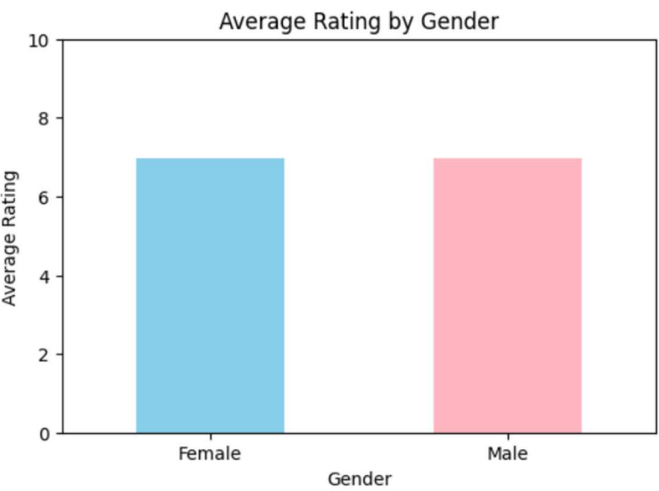
Sales Distribution by Age Group



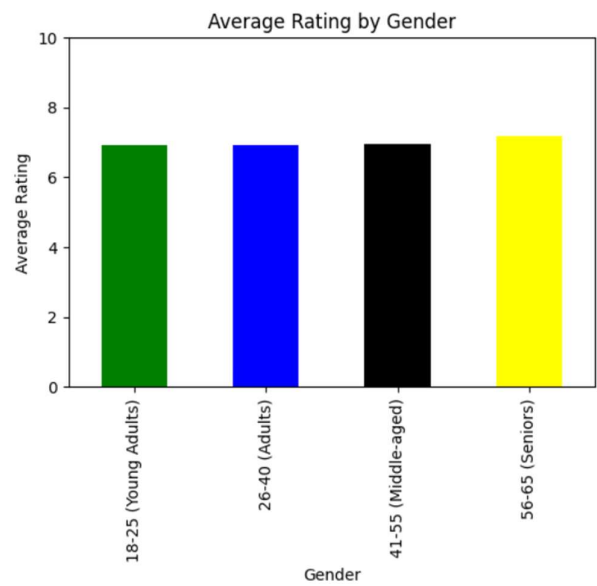
Age Group × Income Level vs Avg Total Sales



Avg Rating by Gender

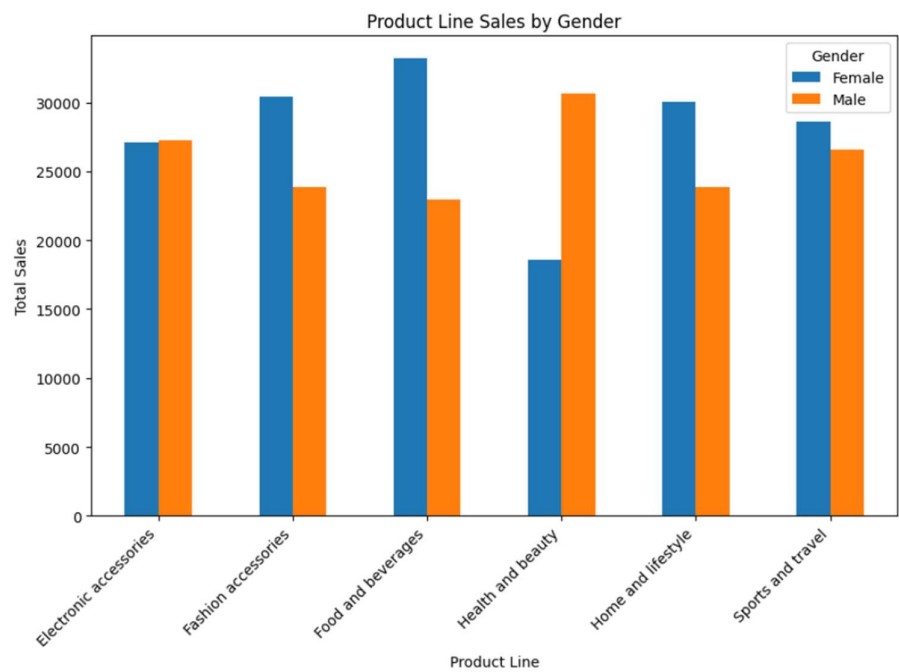


Avg Rating by Age Group

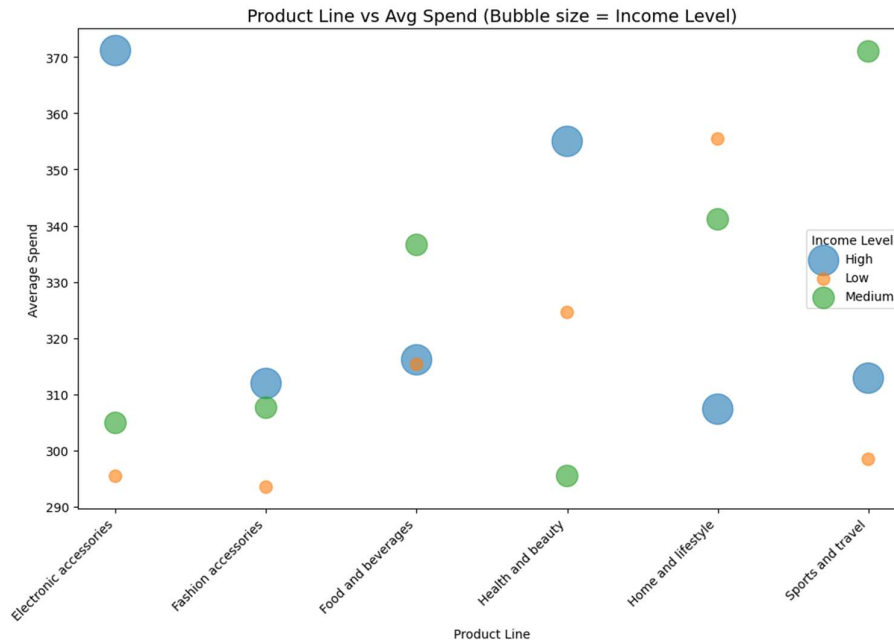


Product Line Preferences

Product Line Sales by Gender



Product Line vs Avg Spend (size = Income Level)



📌 Problem

How can the supermarket increase overall revenue while maintaining healthy profit margins across branches?

🔍 Step 1: Key Metrics to Track

From your dataset, you can analyze:

- **Revenue** = SUM(Total) (sales incl. tax)
- **COGS (Cost of Goods Sold)** = SUM(cogs)
- **Gross Income (Profit)** = SUM(gross income)
- **Profit Margin** = (Gross Income ÷ Revenue) × 100

Then compare across **Branch (A, B, C)** and **City**.

📊 Step 2: Analysis Approach

1. Branch Revenue vs Profitability

- High revenue but low profit margin → possible discounting or high cost structure.
- Low revenue but healthy margin → scaling opportunity.

2. Product Line Contribution

- Identify product lines with **high sales & high profit margins** → promote aggressively.
- Drop or optimize **low-margin product lines**.

3. Customer Segmentation

- Compare **Members vs Normal Customers** spending.
- Higher loyalty = better profitability. Boost loyalty program to convert Normal → Members.

4. Demographic Spending Patterns

- Age × Income × Gender → identify high-spending groups (e.g., 26–40 High-Income Males in Electronics).
- Focus marketing and premium offerings here.

5. Payment Methods

- Analyze profit vs payment choice (Ewallet, Cash, Credit).
- Push for **Ewallet/Credit** if it increases transaction size (convenience, cashback offers).

6. Peak Times

- Use Date & Time fields → find **sales peaks by day/hour**.
- Allocate staff and targeted promotions during high-footfall hours.

Step 3: Revenue Growth Strategies While Preserving Margins

• Branch Optimization

- Replicate **best-performing branch practices** in weaker branches.
- Example: If Yangon branch sells more through Ewallet, encourage the same in Mandalay with cashback tie-ups.

• Product Mix Strategy

- Bundle **high-margin products** with popular low-margin ones.
- Promote **cross-selling** (e.g., Fashion Accessories + Health & Beauty).
- **Customer Loyalty & Retention**
 - Reward Members with discounts on high-margin items.
 - Offer personalized coupons (based on Age/Income group preferences).
- **Dynamic Pricing & Promotions**
 - Adjust pricing for different branches (urban vs semi-urban).
 - Use **time-based discounts** during off-peak hours to spread demand.
- **Upselling & Cross-Selling**
 - Train staff or digital prompts at checkout to recommend add-ons.
 - Example: “Customers buying Sports & Travel often add Fashion Accessories.”

Problem Statement

Which product lines contribute the most and least to sales and profit, and how can underperforming categories be improved?

Step 1: Metrics to Use from Your Dataset

- **Sales Revenue (Total)** → overall sales volume per product line.
 - **Gross Income (Profit)** → profit per product line.
 - **Profit Margin %** = $(\text{Gross Income} \div \text{Revenue}) \times 100$.
 - **Quantity Sold** → units contribution.
 - **Avg Spend per Customer (Revenue ÷ Transactions)** → customer willingness to pay.
-

Step 2: Analysis Approach

1. Rank Product Lines by Revenue

- Identify top contributors (e.g., *Food & Beverages*, *Electronic Accessories*).
- Spot weakest contributors (e.g., *Fashion Accessories*).

2. Rank Product Lines by Profit

- A category may have high sales but low profit (possible discounts/high cost).
- Another may have moderate sales but strong margins (hidden gem).

3. Cross Revenue vs Profit

- High Sales + High Profit = ★ “Star Category” (boost more).
 - High Sales + Low Profit = ⚠ “Volume Trap” (optimize costs/price).
 - Low Sales + High Profit = 💎 “Niche Profit” (expand marketing).
 - Low Sales + Low Profit = ✖ “Underperformer” (fix or phase out).
-

💡 Step 3: Recommendations for Underperforming Categories

◆ If Low Sales but High Profit Margin

- Example: *Sports & Travel* sells less but earns good margins.
- ✅ Action: Increase visibility (promotions, shelf placement, bundles).

◆ If High Sales but Low Profit Margin

- Example: *Food & Beverages* may sell a lot but profit is squeezed.
- ✅ Action: Adjust pricing, reduce discounts, negotiate supplier terms.

◆ If Low Sales and Low Profit

- Example: *Fashion Accessories* often falls in this quadrant.
- ✅ Action:
 - Refresh product range.
 - Pair with high-demand items (*cross-sell* with *Health & Beauty*).
 - Target specific demographics (young females with fashion campaigns).

◆ If High Sales and High Profit

- Example: *Health & Beauty* often excels.
- ✅ Action:
 - Prioritize inventory.

- Push more marketing spend.
- Use loyalty rewards to make customers stick.

Problem Statement

What are the key drivers (unit price, quantity, customer type) influencing high-value transactions?

Step 1: Define "High-Value Transaction"

From your dataset:

- **Total** = Sales including tax.
 - We can define **High-Value** = transactions in the **top 20% of Total sales** (or above a threshold like >500).
-

Step 2: Potential Drivers

1. Unit Price

- Higher unit price products (e.g., *Sports & Travel*, *Home & Lifestyle*) often create larger transactions even at low quantity.
- Correlation test: Does Unit Price $\uparrow \rightarrow$ Total $\uparrow$ significantly?

2. Quantity

- Some categories rely on bulk purchases (e.g., *Food & Beverages*, *Electronic Accessories*).
- High quantity at moderate price can still push transactions into high-value.

3. Customer Type (Member vs Normal)

- Members might spend more due to loyalty benefits, or
- Normal customers might buy in bulk occasionally.

4. Branch/City Factor (secondary)

- Branches in urban centers (Yangon, Naypyitaw) may have higher-ticket sales.
-

Step 3: Analytical Methods

1. Correlation Analysis

- Correlate Unit Price, Quantity, and Customer Type (encoded) with Total.
- Check which driver has strongest relationship.

2. Regression Model

- Build a simple multiple regression:

$$\text{Total} = \beta_0 + \beta_1(\text{UnitPrice}) + \beta_2(\text{Quantity}) + \beta_3(\text{CustomerType}) + \varepsilon$$

- Coefficients tell how much each factor influences sales.

3. Segmentation

- Slice transactions by High vs Low Value and compare average Unit Price, Quantity, Customer Type distribution.

Step 4: Possible Findings (Hypothesis)

- **Unit Price:** Likely a strong driver for categories like *Sports & Travel*, *Health & Beauty*.
- **Quantity:** Big driver in *Food & Beverages* and *Electronic Accessories*.
- **Customer Type:** Members may have more medium-value frequent purchases, while Normals might spike into high-value with bulk buys.

Step 5: Business Actions

- If **Unit Price is key** → Upsell premium products, highlight quality.
- If **Quantity is key** → Offer bulk discounts, family packs.
- If **Customer Type is key** → Expand loyalty program or target occasional big spenders.