Project Title:
In-Depth Analysis of Supermarket Sales Across Three Branches

Project overview

Description:
This project involves an in-depth analysis of sales data from three supermarket branches over a three-month period (January to March). The goal is to explore sales trends, customer behavior, and branch performance to uncover valuable insights that can drive business improvements. The analysis covers various aspects, including sales by product line, customer demographics, peak shopping times, and customer satisfaction.

The dataset, which is available on Kaggle, contains detailed information about sales transactions, including customer types, product categories, pricing, payment methods, and customer ratings. This analysis highlights key insights on gross income, sales performance, and customer feedback, helping businesses optimize their strategies.

Key Analyses:

1. **Customer Behavior Analysis:**
    1. Product line popularity by customer type and gender.
    2. Identification of peak shopping times by day and hour.
    3. Deeper insights by combining customer type, gender, and time.
2. **Branch Performance Analysis:**
    1. Total sales performance by branch.
    2. Gross income comparison.
    3. Customer satisfaction analysis across branches.
3. **Product Insights:**
    1. Sales volume and gross income contribution by product line.
    2. Average gross income per sale and the relationship between sales volume and gross income.
    3. Comparison of product line performance by customer type.

1. Data Preparation

1.1 Libraries and Tools

To perform the analysis, the following Python libraries were utilized:

- **Pandas** and **NumPy** for data manipulation and statistical analysis.
- **Seaborn** and **Matplotlib** for creating visualizations.

1.2 Dataset Loading

The dataset was sourced from Kaggle and imported into the environment for analysis. The file contains sales data from a supermarket company operating across three branches, spanning three months of recorded transactions.

1.3 Initial Exploration

The dataset was inspected to understand its structure and key attributes:

- **Previewing the Dataset:** The dataset includes **1,000 rows** and **17 columns**, representing various aspects of sales data, including product information, customer demographics, and financial details.

- **Statistical Overview:** A summary of the numerical columns revealed consistent values for key statistics such as minimums, maximums, and quartiles, confirming the dataset's integrity.
- **Data Types and Completeness:** Each column was verified to ensure proper data types (e.g., numerical, categorical, or date/time). Additionally, no missing values were identified, ensuring a clean dataset for analysis.

At this stage, the dataset was deemed complete and well-structured, ready for deeper exploration and analysis.

2. Exploratory Data Analysis (EDA)

2.1 Overview of Data Relationships

The initial step in the exploratory analysis involved assessing the relationships between the numerical features using a **pairplot**. The pairplot revealed several notable insights:

- **Total**, **Tax (5%)**, and **Gross Income** exhibit strong linear correlations, which align with their derived relationships, such as the fact that the **Total** is directly influenced by the sum of **Tax (5%)** and **Gross Income**.
- **Gross Margin Percentage** displayed no variability, indicating it remained constant across transactions, thus offering no significant predictive value.
- The **Quantity** variable exhibited a grid-like distribution, signifying that it is a discrete variable.
- Variables such as **Unit Price** and **Rating** did not show any significant trends or correlations, which suggests weak relationships between these variables and other features.
- Some **outliers** were observed in the **Unit Price** and **Total** variables, suggesting that further examination may be necessary for these extreme values.

2.2 Correlation Matrix

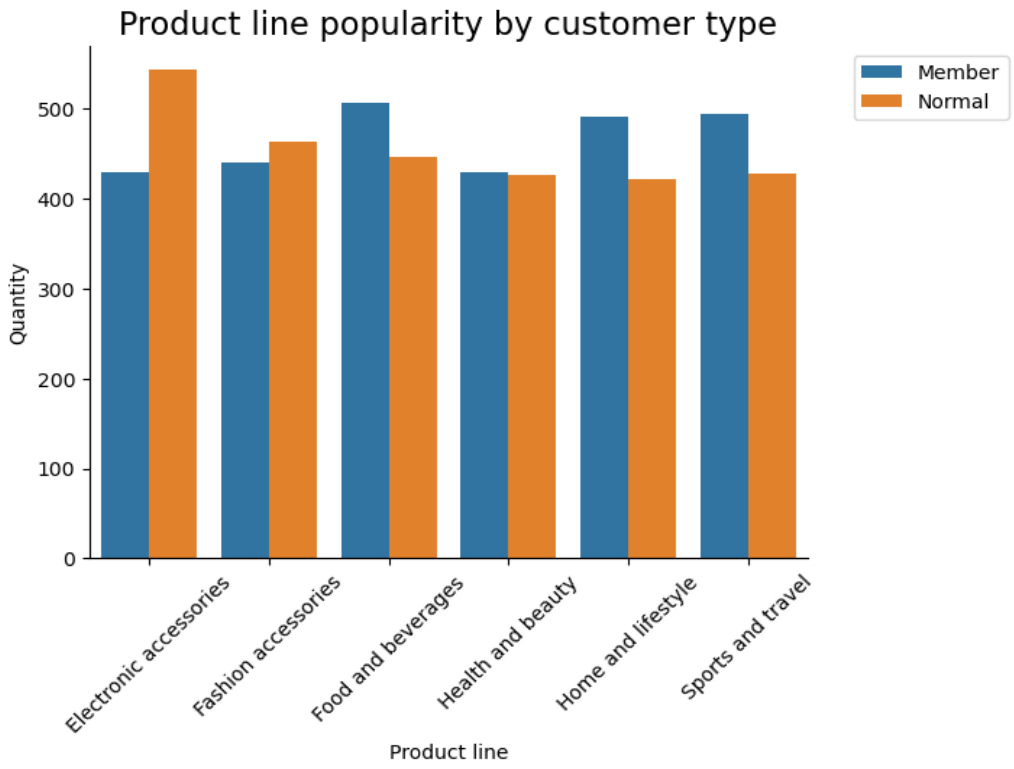
The correlation matrix analysis highlighted expected patterns, with most of the variables (except **Rating**) being strongly dependent on one another. Notably, **Rating** was found to be an isolated variable, showing little to no correlation with other features.

2.3 Customer Behavior Analysis

2.3.1 Customer Type and Spending Patterns

The spending patterns across customer types were examined, revealing the following:

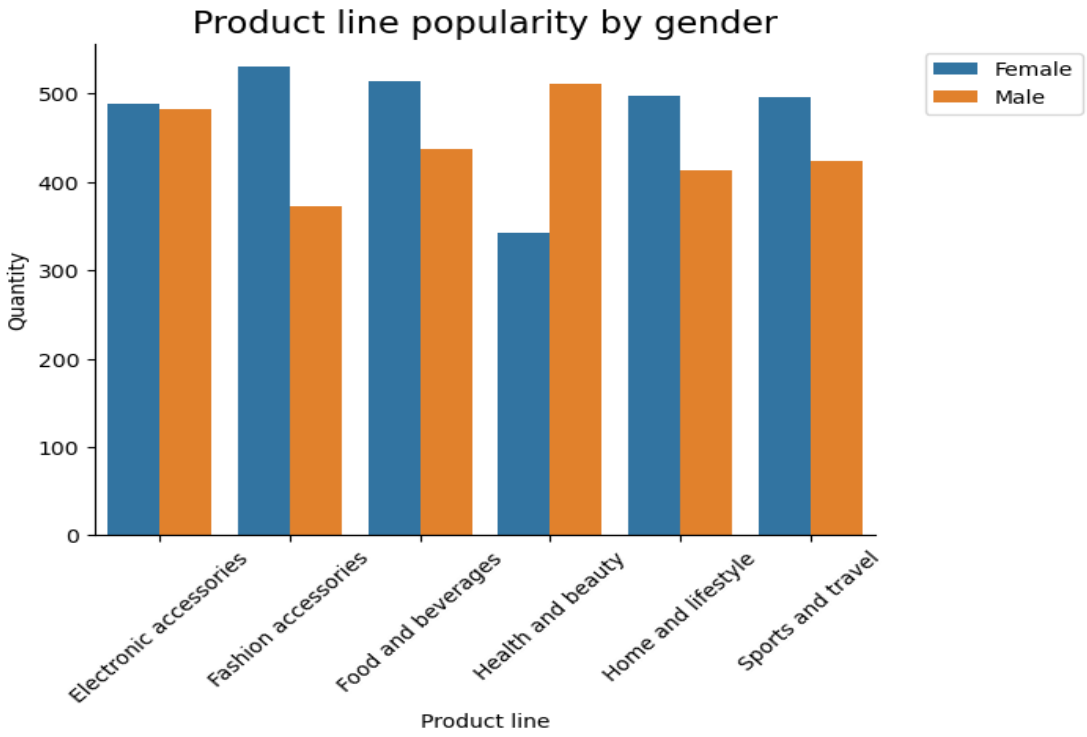
- **Members** tended to spend more on **Food and Beverages**, while **Normal Customers** favored **Electronic Accessories**. Interestingly, members spent more than non-members in almost all categories, except for **Fashion** and **Electronic Accessories**.



2.3.2 Gender and Product Preferences

The analysis of **customer gender** provided these insights:

- **Females** generally spent more across most categories, with the highest spending observed in **Fashion Accessories** and **Food and Beverages**.
- **Males**, on the other hand, tended to spend more on **Health and Beauty** and **Electronic Accessories**, while they spent the least on **Fashion Accessories**.

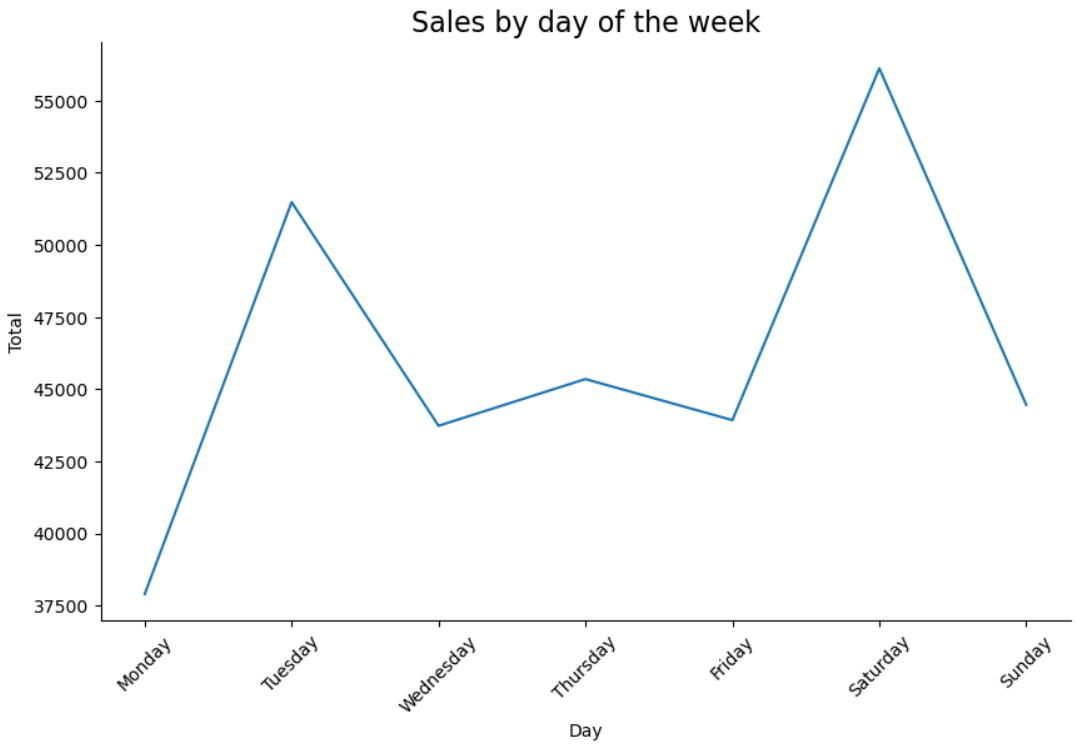


2.4 Peak Shopping Times

2.4.1 Shopping Patterns by Day

The analysis of shopping frequency by day revealed the following:

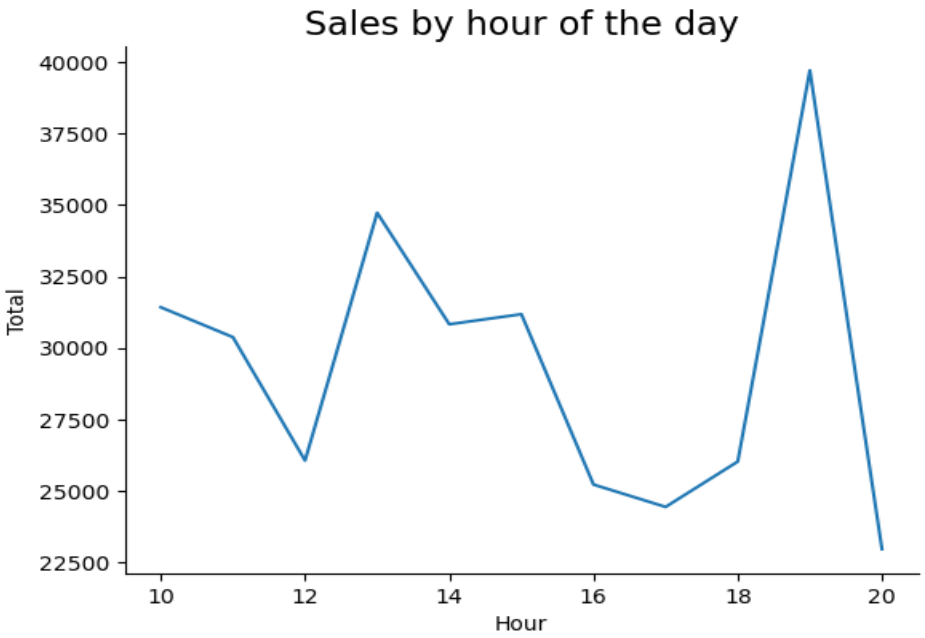
- **Saturdays** and **Tuesdays** emerged as the busiest days for shopping, while **Mondays** had the lowest foot traffic.
- Other weekdays showed similar levels of activity, suggesting consistent shopping behavior across the week.



2.4.2 Shopping Patterns by Hour

The peak shopping times by hour were identified as:

- **7 PM** was the busiest hour, followed by **1 PM**.
- Shopping activity was fairly consistent during **10 AM**, **11 AM**, **2 PM**, and **3 PM**.
- The lowest levels of shopping occurred at **4 PM**, **5 PM**, **6 PM**, and **after 8 PM**.



2.5 Branch Performance Analysis

2.5.1 Sales Performance by Branch

In terms of **sales performance**, the analysis revealed that:

- **Branch C** marginally outperformed **Branches A and B**, which showed similar performance levels.
- The differences were small but suggest that **Branch C** is slightly more successful in driving sales.

2.5.2 Customer Satisfaction (Ratings) by Branch

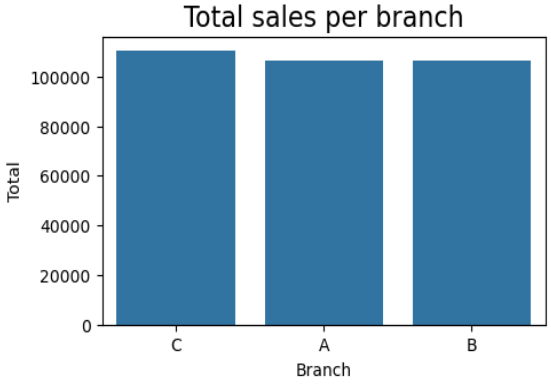
Looking at customer satisfaction, measured through **ratings** by branch:

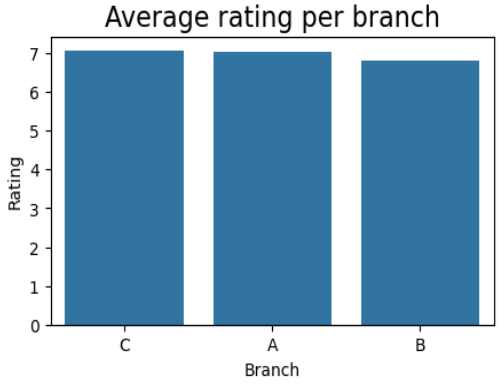
- **Branch C** and **Branch A** had similar customer satisfaction ratings.
- **Branch B**, however, performed slightly lower in customer satisfaction, indicating areas for improvement.

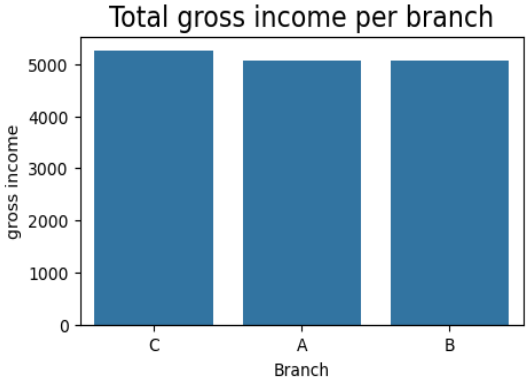
2.5.3 Gross Income by Branch

The analysis of **gross income** showed:

- **Branch C** generated the highest gross income compared to **Branches A and B**, where the performance was almost identical.





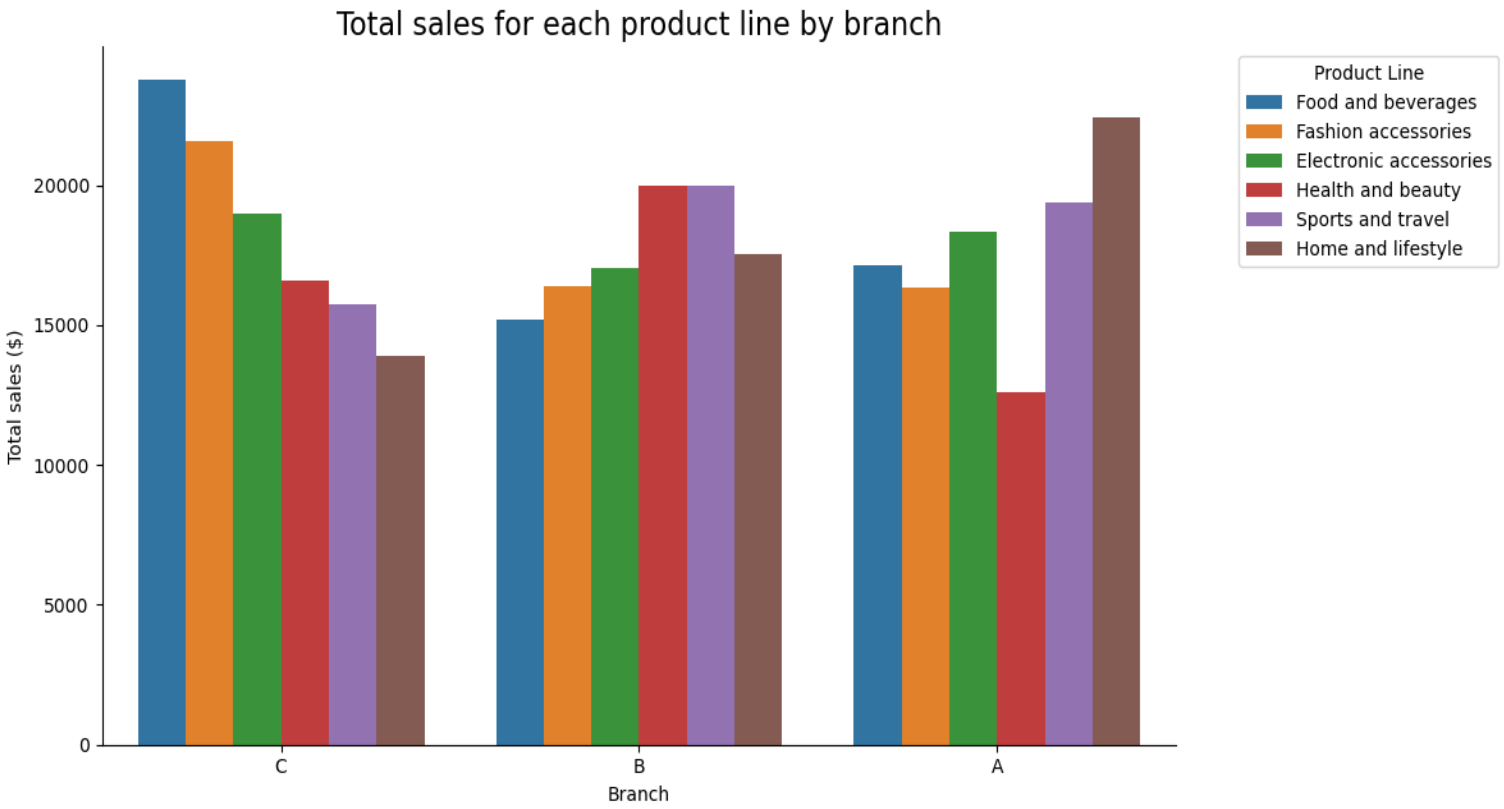


2.6 Performance by Product Line and Customer Type

2.6.1 Sales by Product Line for Each Branch

The analysis by **product line** revealed:

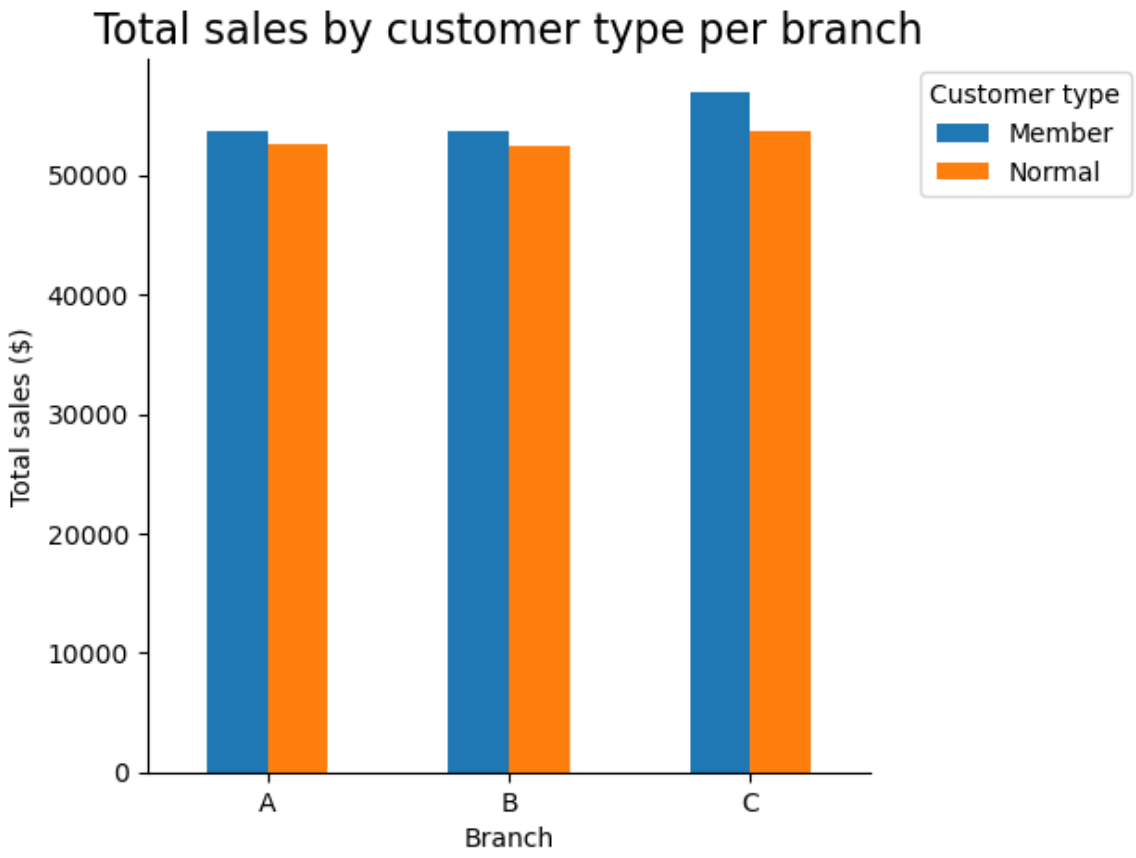
- **Branch A** excelled in **Home and Lifestyle** with the highest overall sales, but underperformed in **Health and Beauty**.
- **Branch B** showed steady, mid-level performance across all categories without significant highs or lows.
- **Branch C** dominated in **Food and Beverages** and **Fashion Accessories**, but had weaker performance in **Home and Lifestyle**.



2.6.2 Sales by Customer Type

When comparing **sales by customer type**, it was found that:

- **Members** spent more than **Normal Customers** across all branches.
- **Branch C** had the highest sales from both **members** and **normal customers**, while **Branches A and B** performed similarly for both customer types.

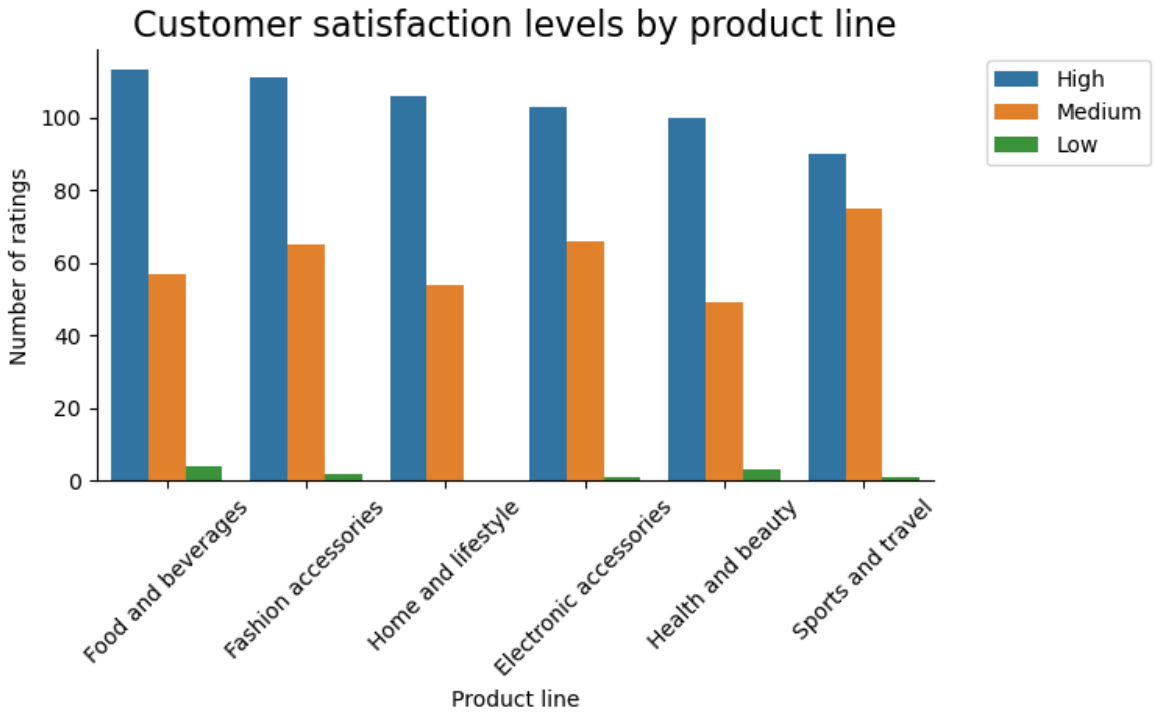


2.7 Customer Satisfaction

2.7.1 Satisfaction Levels by Product Line

For **product line vs satisfaction level**, customer ratings were categorized into three groups:

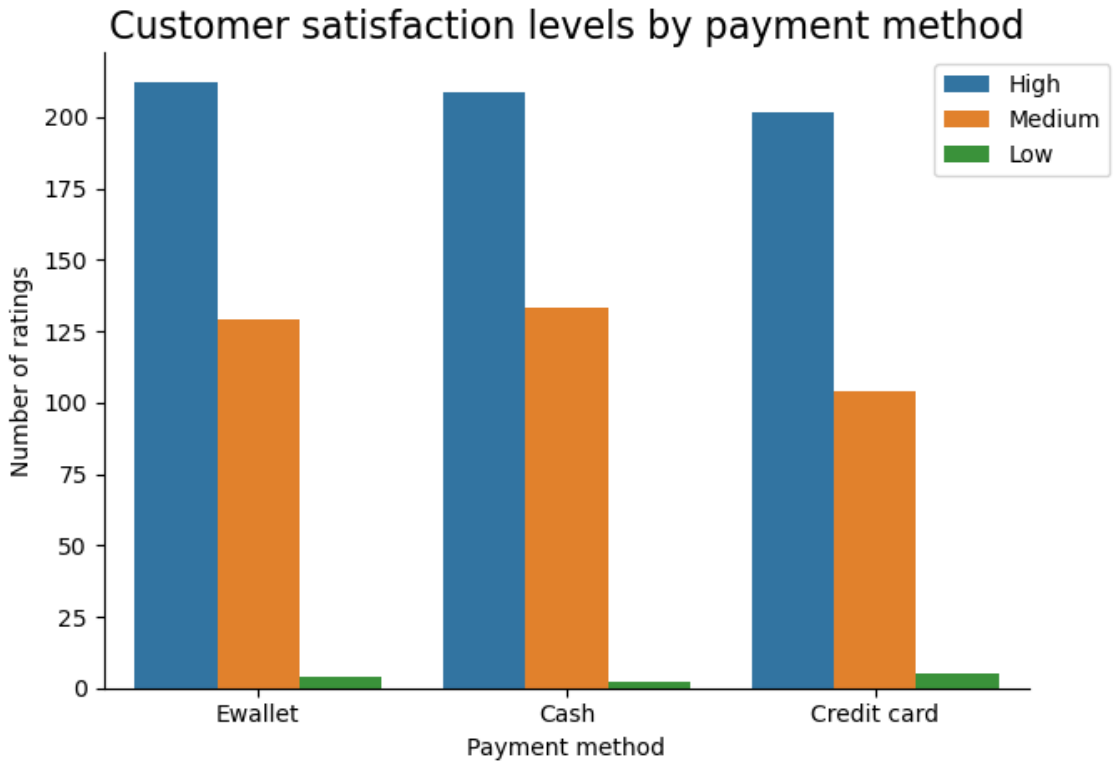
- **Food and Beverages** received the highest number of **high ratings** (113), but also the most **low ratings** (4), indicating a polarized customer experience.
- **Fashion Accessories** and **Electronic Accessories** had similar customer satisfaction levels with a high number of **high ratings** (111 and 103, respectively).
- **Sports and Travel** had the highest number of **medium ratings** (75), suggesting a more balanced customer experience.
- **Health and Beauty** and **Home and Lifestyle** had fewer **high ratings**, but **Home and Lifestyle** stood out for having no **low ratings**, indicating stability in customer satisfaction.



2.7.2 Satisfaction Levels by Payment Method

For **payment method vs satisfaction level**, the findings were:

- **Ewallet** had the highest number of **high ratings**, while **Credit Card** had the lowest number of **high ratings** and the highest number of **low ratings**.
- **Cash** transactions showed the lowest **low ratings** and the highest **medium ratings**, indicating more balanced satisfaction.

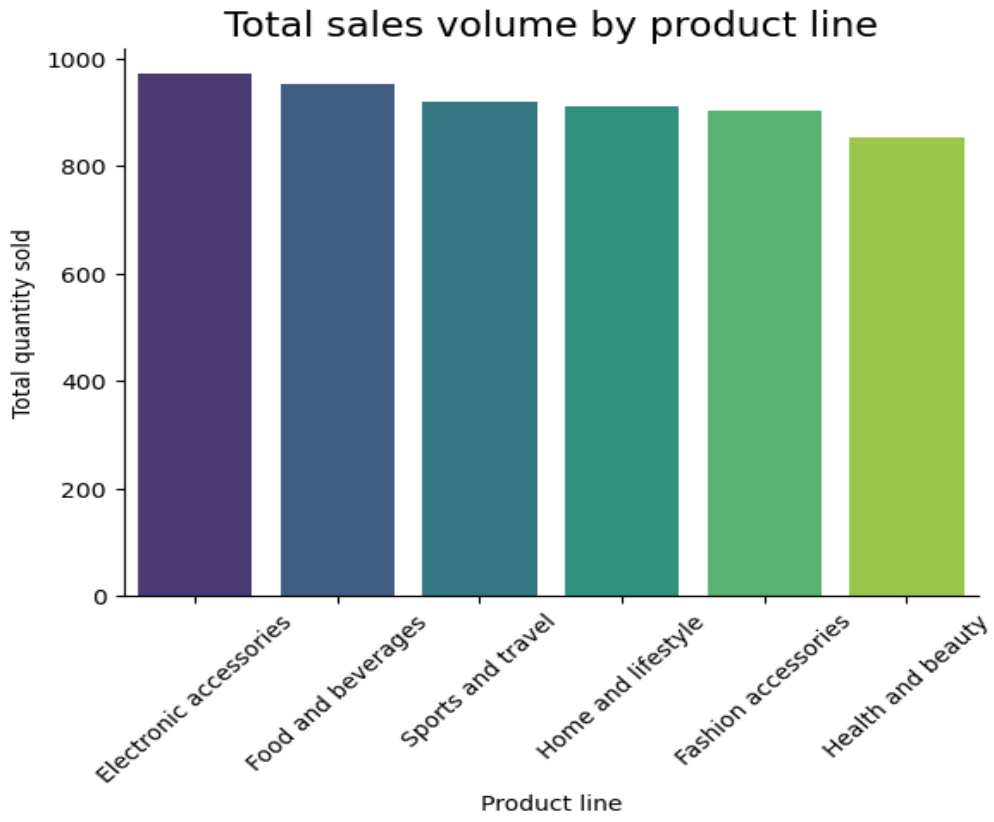


2.8 Product Insights

2.8.1 Sales Volume by Product Line

The sales volume by **product line** revealed that:

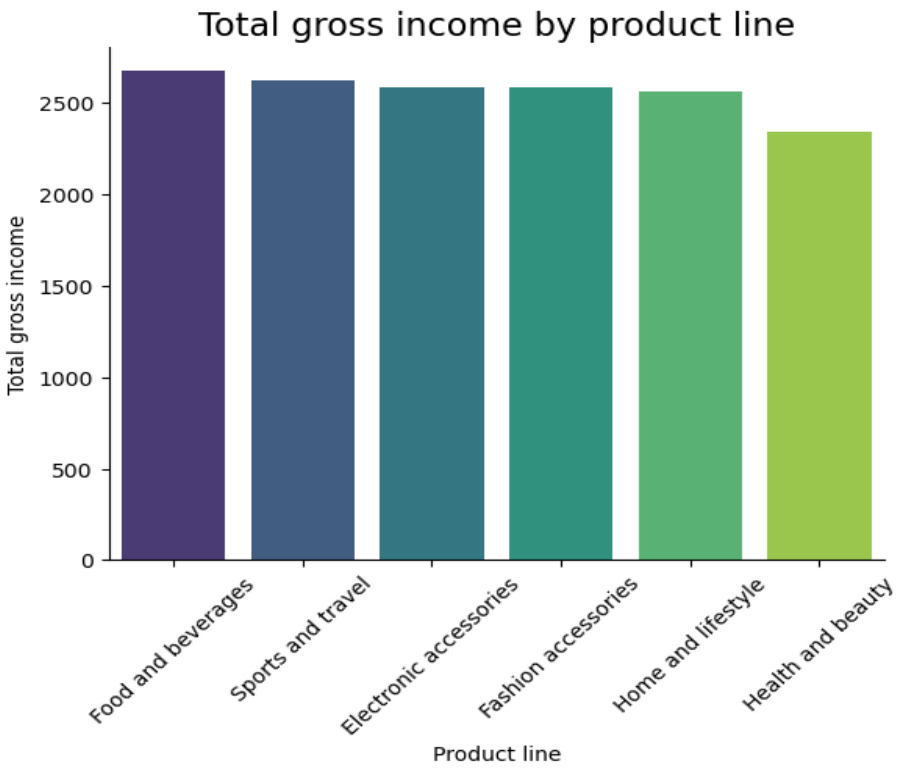
- **Electronic Accessories** was the highest-selling product line, followed by **Food and Beverages**, **Sports and Travel**, and **Home and Lifestyle**.
- **Health and Beauty** had the lowest sales volume.



2.8.2 Gross Income Contribution by Product Line

The **gross income** contribution by product line showed that:

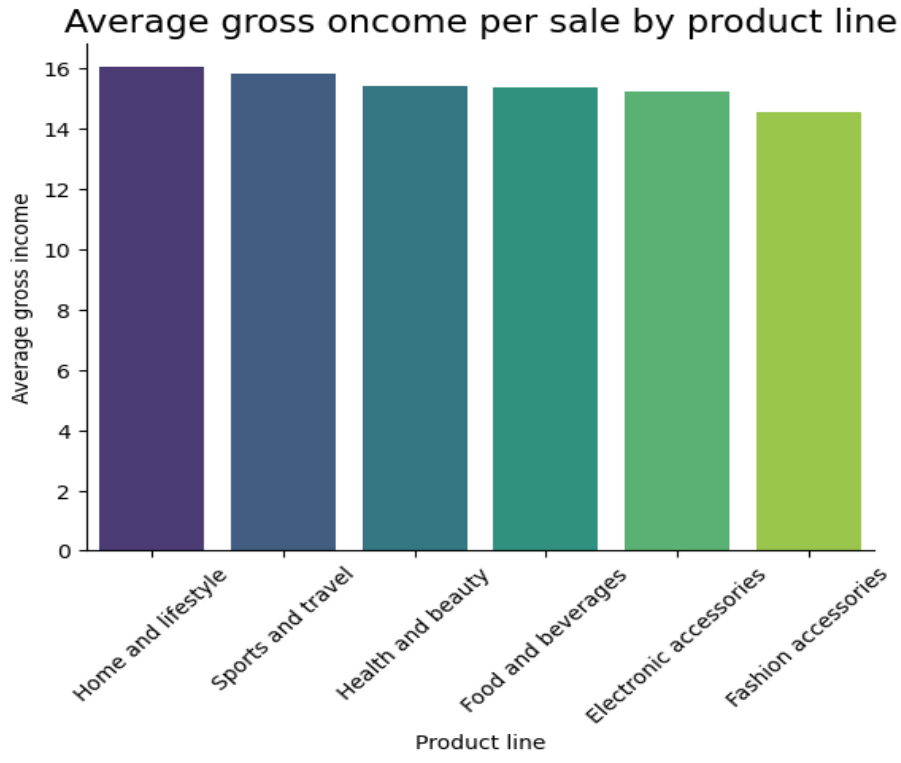
- **Food and Beverages** generated the highest gross income, while **Health and Beauty** contributed the least.
- Other product lines, such as **Sports and Travel**, **Electronic Accessories**, and **Fashion Accessories**, showed similar levels of income.



2.8.3 Average Gross Income per Sale by Product Line

When examining the **average gross income per sale**, it was found that:

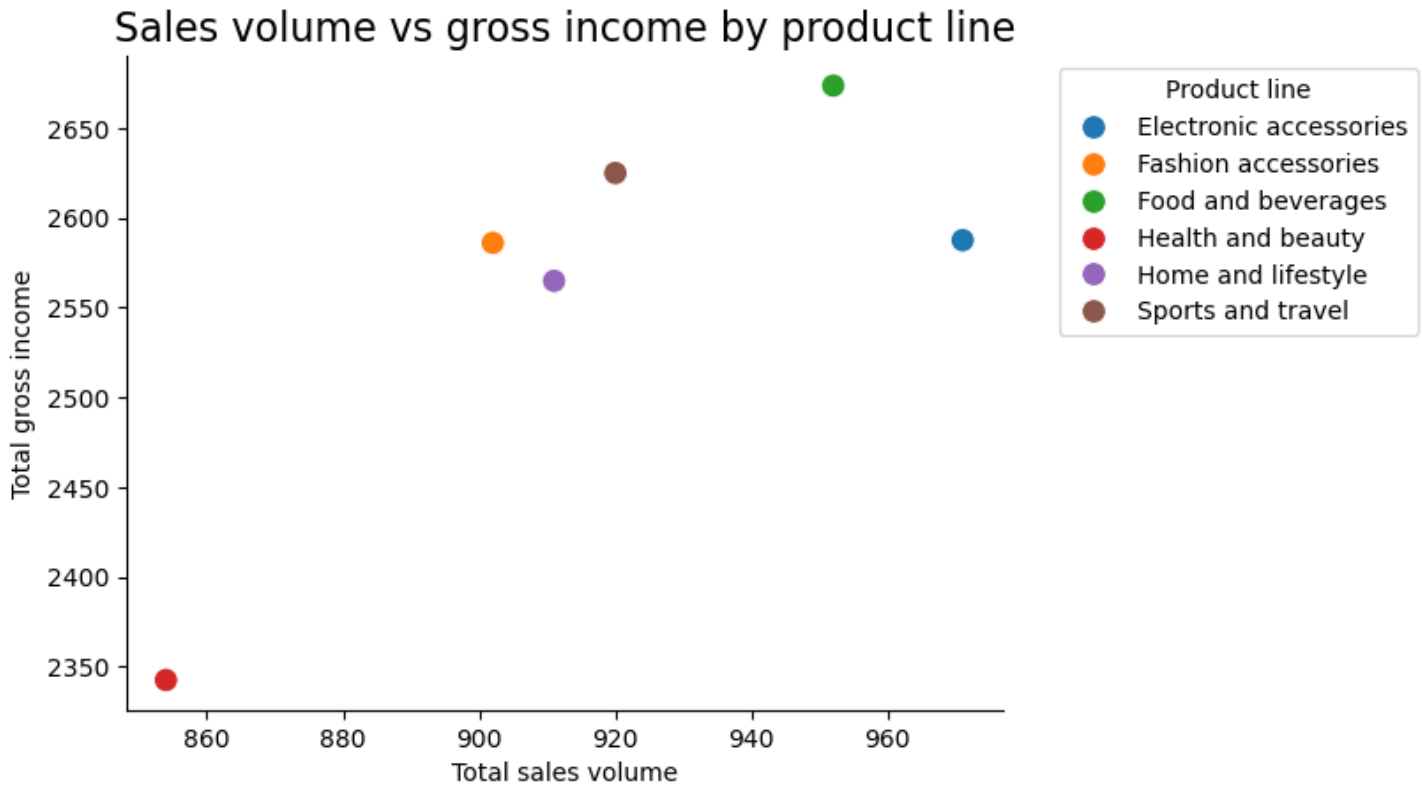
- **Home and Lifestyle** had the highest average, followed by **Sports and Travel** and **Health and Beauty**.
- **Food and Beverages** and **Electronic Accessories** had similar average gross incomes, with **Fashion Accessories** having the lowest.



2.8.4 Correlation Between Sales Volume and Gross Income

A correlation analysis between **sales volume** and **gross income** revealed:

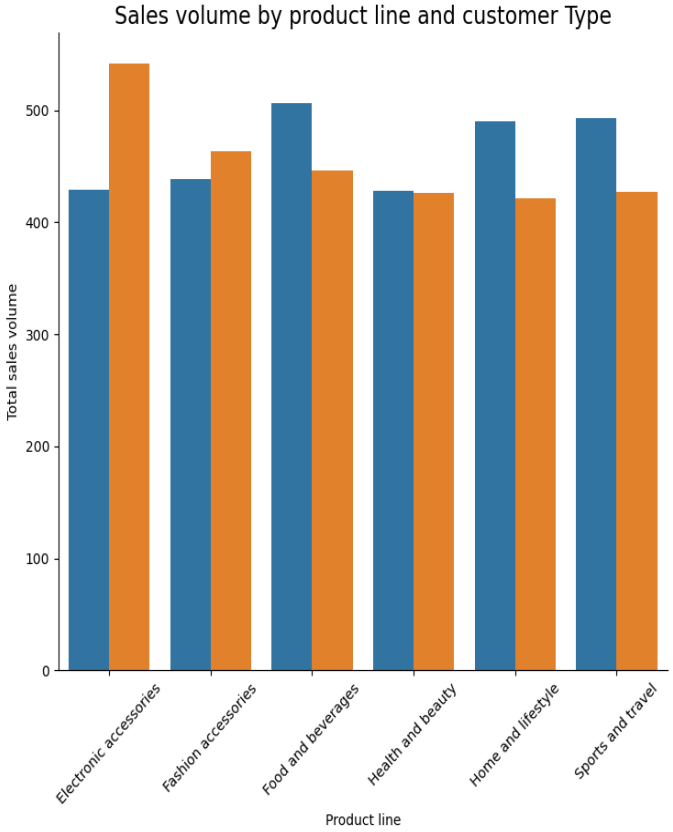
- **Food and Beverages** was the top performer in both sales and gross income.
- **Electronic Accessories**, despite having the highest sales volume, had a lower gross income, suggesting a lower profit margin.
- **Health and Beauty** had both the lowest sales and gross income, indicating a potential area for improvement.
- **Sports and Travel** performed well in terms of profit despite lower sales.

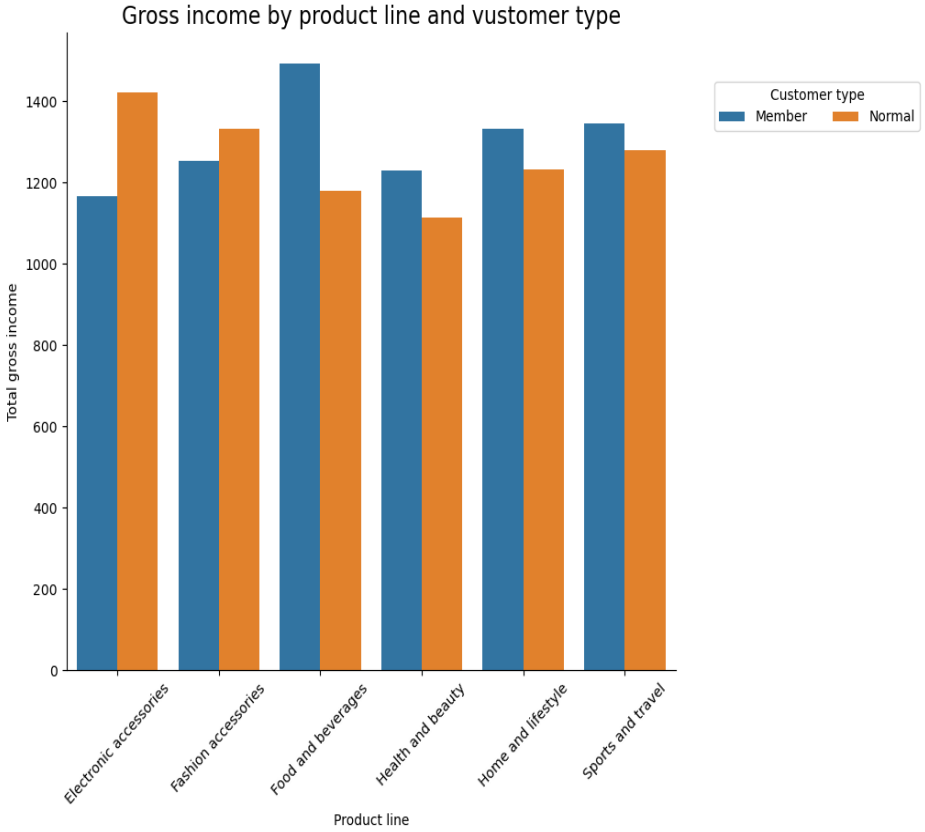


2.8.5 Product Line Comparison by Customer Type

In terms of **sales and gross income by customer type**, it was found that:

- **Electronic Accessories** generated more sales from **Normal Customers**, while **Food and Beverages** performed better with **Members**, suggesting that membership drives higher loyalty in certain categories.
- **Health and Beauty** saw similar sales from both customer types, but members spent more per purchase.


Sales volume by product line and customer Type
Total sales volume
0
100
200
300
400
500
Electronic accessories
Fashion accessories
Food and beverages
Health and beauty
Home and lifestyle
Sports and travel
Product line


Gross income by product line and vustomer type
Total gross income
0
200
400
600
800
1000
1200
1400
Customer type
Member
Normal
Electronic accessories
Fashion accessories
Food and beverages
Health and beauty
Home and lifestyle
Sports and travel
Product line