From this initial look, we can see various columns such as Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax, Total, Date, Time, Payment method, Cost of goods sold (cogs), Gross margin percentage, Gross income, and Rating.

To predict sales analysis and suggest improvements, we'll need to perform the following steps:

1. **Data Cleaning**: Check for any missing or inconsistent data.
2. **Exploratory Data Analysis (EDA)**: Analyze the sales trends, customer demographics, and product performance.
3. **Sales Forecasting**: Use historical data to predict future sales.
4. **Identify Patterns**: Look for patterns in customer behavior, product popularity, and sales performance across different branches.
5. **Recommendations**: Based on the analysis, suggest strategies for business improvement.

Let's start with data cleaning and exploratory data analysis to understand the sales trends and patterns.

**CODE**

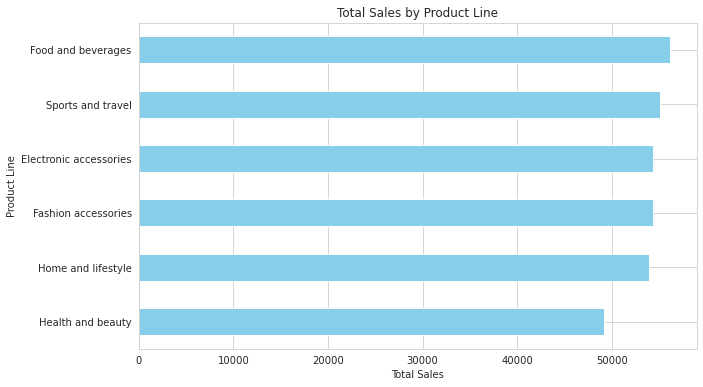
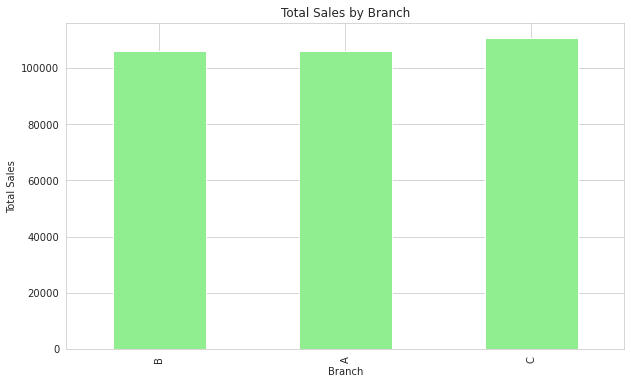
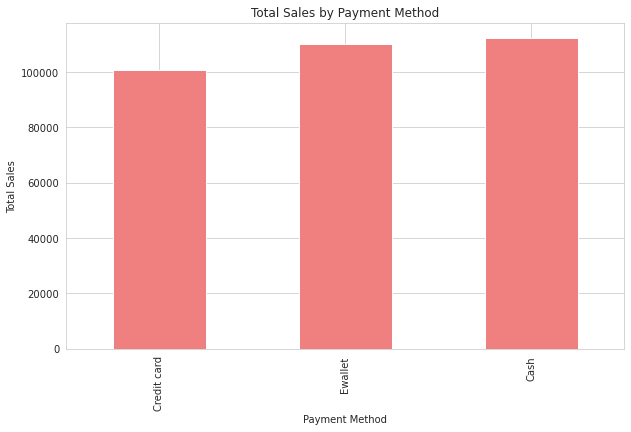
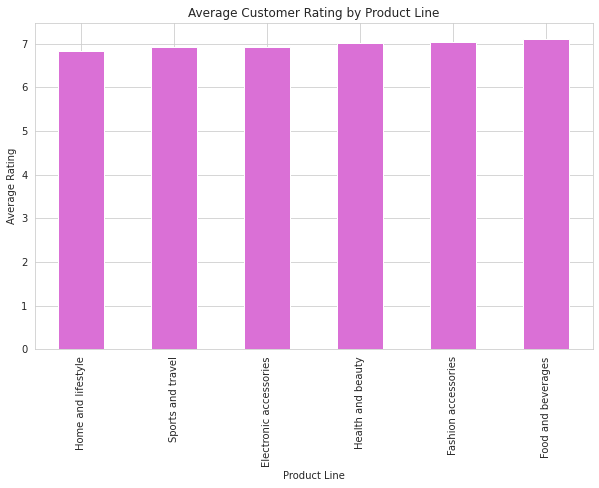
The dataset has no missing values and no duplicate rows, which is excellent for analysis. Here's a summary of the dataset:

* There are 1000 entries, and each entry has a unique Invoice ID.
* The dataset covers 3 branches, located in 3 different cities.
* Customer types are categorized into 'Member' and 'Normal'.
* There are entries for both 'Male' and 'Female' genders.
* The product line is diversified into 6 categories.
* The dataset includes various continuous variables like 'Unit price', 'Quantity', 'Tax 5%', 'Total', 'cogs', 'gross margin percentage', 'gross income', and 'Rating'.
* There are 89 unique dates and 506 unique times when purchases were made.
* Payments were made using 3 different methods.

Next, we will perform exploratory data analysis to uncover sales trends and customer behaviors.

**CODE**

The exploratory data analysis has provided some insights into the sales trends and customer preferences:

* **Total Sales by Product Line**: Food and beverages lead in sales, followed closely by sports and travel, and electronic accessories. Health and beauty products have the lowest total sales. 
* **Total Sales by Branch**: Branch C has the highest total sales, while Branches A and B are almost equal in their sales performance. 
* **Total Sales by Payment Method**: Cash payments account for the highest total sales, followed by Ewallet and credit card payments. 
* **Average Customer Rating by Product Line**: Food and beverages have the highest average customer rating, indicating customer satisfaction in this category. Home and lifestyle products have the lowest average rating. 

Based on these findings, here are some suggestions to improve business:

* **Product Line Focus**: Increase marketing efforts for health and beauty products to boost sales in this category.
* **Branch Performance**: Investigate why Branch C performs better than A and B to replicate successful strategies across all branches.
* **Payment Incentives**: Since cash payments are most popular, consider offering incentives for using alternative payment methods to increase their usage.
* **Customer Satisfaction**: Explore the reasons behind the lower ratings in the home and lifestyle category to improve customer satisfaction.

Next, we can perform a more detailed analysis of customer demographics, time-based sales trends, and correlations between variables such as gross income and customer ratings.

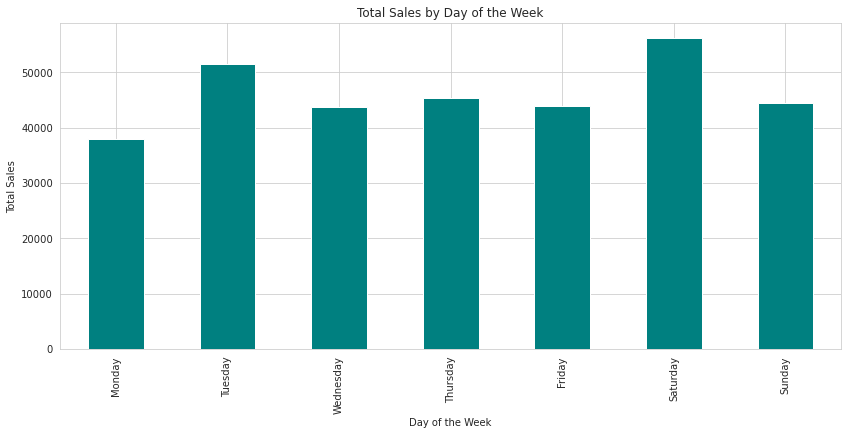
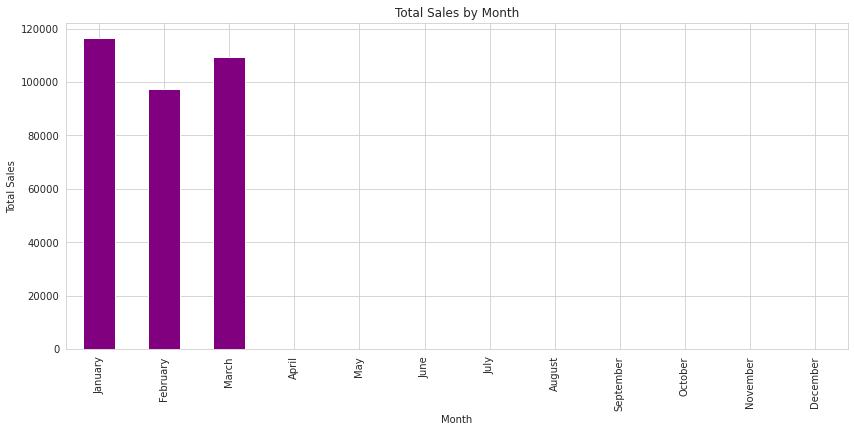
**CODE**

The analysis of sales trends by day of the week and by month has been completed.

* **Sales Trends by Day of the Week**: The data shows that Saturday has the highest total sales, followed by Tuesday, indicating that these might be the busiest shopping days of the week for the supermarket. Sales trends by day of the week:
* **Sales Trends by Month**: The analysis reveals that January has the highest total sales, followed by March and then February. It's important to note that the dataset might not cover the entire year, as there are no sales data for months April through December.

|  | **Total** |
| --- | --- |
| **DayOfWeek** |  |
| **Monday** | 37899.078 |
| **Tuesday** | 51482.2455 |
| **Wednesday** | 43731.135 |
| **Thursday** | 45349.248 |
| **Friday** | 43926.3405 |
| **Saturday** | 56120.8095 |
| **Sunday** | 44457.8925 |

Here are the visual representations of the sales trends:

* **Total Sales by Day of the Week**: 
* **Total Sales by Month**: 

Based on these findings, the supermarket could consider the following actions:

* **Promotional Activities**: Plan more promotional activities or sales on Saturdays and Tuesdays to capitalize on the higher customer footfall.
* **Seasonal Analysis**: Investigate the reasons behind higher sales in January and March to understand seasonal trends and plan inventory accordingly.

**CODE**

Continuing with the analysis, I will now explore the sales data further by examining the following:

* The relationship between the product line and gross income to see which products are generating the most revenue.
* The average unit price and quantity sold for each product line to understand the sales volume and pricing strategy.
* The distribution of payment methods used by customers to identify any trends in how customers prefer to pay.

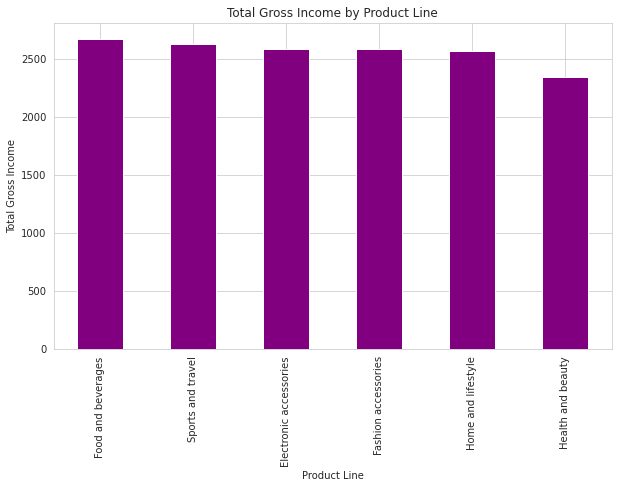
I will conduct this analysis and provide visual representations to aid in understanding the trends. Let's proceed with the analysis.

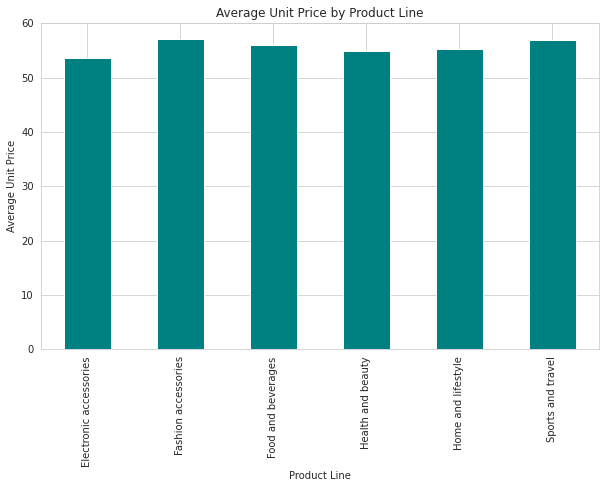
**CODE**

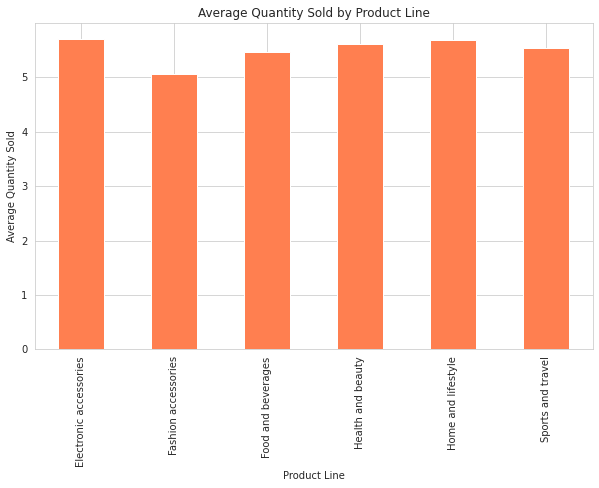
The analysis of the supermarket sales data has provided some interesting insights:

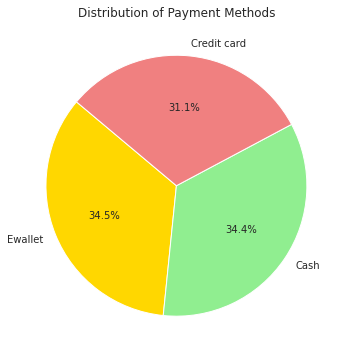
* The product line "Food and beverages" generated the highest gross income, followed closely by "Sports and travel" and "Electronic accessories".
* The average unit price was relatively consistent across different product lines, with "Fashion accessories" having the highest average price and "Electronic accessories" the lowest.
* The average quantity sold was also quite consistent, with "Home and lifestyle" having a slightly higher average quantity sold than other product lines.
* Payment methods were distributed fairly evenly among customers, with "Ewallet" being the most used, followed by "Cash" and "Credit card".

Let's take a look at the visual representations of these findings:

Total Gross Income by Product Line: 

Average Unit Price by Product Line: 

Average Quantity Sold by Product Line: 

Distribution of Payment Methods: 

These visuals help to better understand the sales dynamics within the supermarket. The data suggests that while some product lines generate more revenue, the pricing and quantity sold are quite balanced across the board. Additionally, the preference for payment methods does not show a significant bias towards any particular method.

I will delve deeper into the data to uncover additional insights:

* Analyze the sales performance by city to see which locations are generating the most revenue.
* Investigate the relationship between the time of day and sales to determine if there are peak sales periods.
* Explore the impact of customer type (Member vs. Normal) on sales performance.

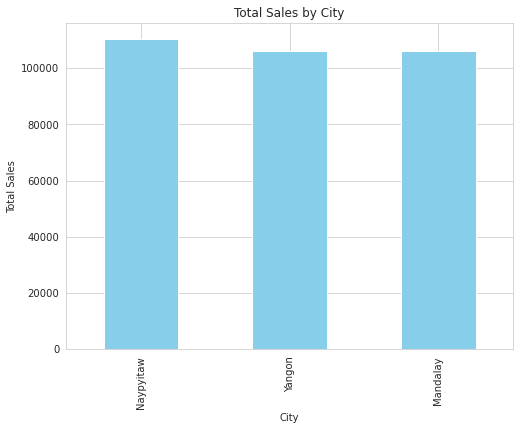
I will execute these analyses and provide visualizations to enhance the understanding of these aspects. Let's proceed with this next step.

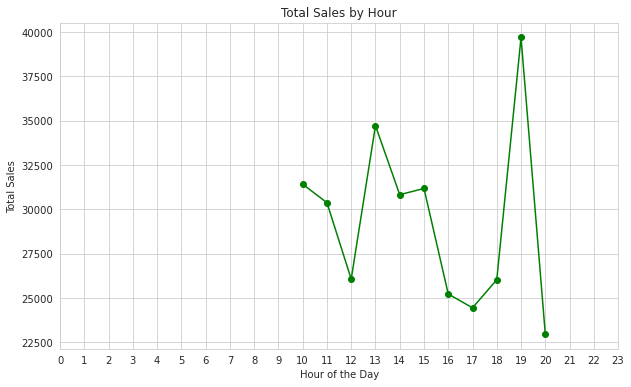
**CODE**

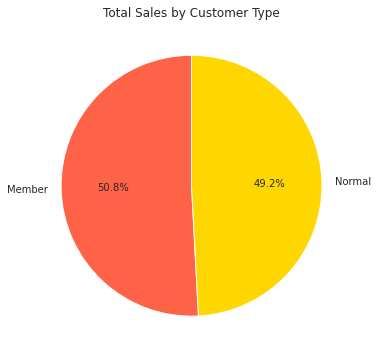
The further analysis has provided additional insights into the supermarket sales:

* The city of Naypyitaw has the highest total sales, followed by Yangon and Mandalay, which are quite close in figures.
* Sales tend to peak around 1 PM and 7 PM, indicating these could be the busiest times for shopping.
* Members contribute slightly more to total sales compared to normal customers.

Here are the visual representations of these findings:

Total Sales by City: 

Total Sales by Hour: 

Total Sales by Customer Type: 

These visuals suggest strategic opportunities for targeted marketing and staffing adjustments to accommodate peak shopping hours. Additionally, the slight difference in sales between member and normal customers might indicate the effectiveness of loyalty programs.