



REPORT

FOR ADVANCED PROGRAMMING COURSE

BY:

Mohanad Ahmad Rashad Yasin
Abdel Rahman Atef E'layan
Rami Bassam Sabri Ahmad

Supermarket Sales Analysis Report

Introduction

This project presents a comprehensive analysis of supermarket sales data, conducted using Python within the Google Colab environment.

The primary objective of this study is to extract actionable insights regarding customer purchasing behavior, evaluate overall sales performance across different branches, and uncover significant relationships between various key variables. These goals are achieved through the application of advanced data analysis techniques and effective visualization tools.

Summary of Statistical Tables

Sales Performance by Branch

Branch	Count	Total Sales (\$)	Avg Sales (\$)	Median Sales (\$)	Max Sales (\$)	Min Sales (\$)	Avg Rating	Total Income (\$)
Alex	340	106,200.37	312.35	240.83	1,039.29	12.69	7.03	5,057.16
Cairo	332	106,197.67	319.87	252.88	1,022.49	18.64	6.82	5,057.03
Giza	328	110,568.71	337.1	271.42	1,042.65	10.68	7.07	5,265.18

Sales by Product Line

Product Line	Count	Total Sales (\$)	Avg Sales (\$)	Max Sales (\$)	Min Sales (\$)	Avg Rating	Total Income (\$)
Electronic accessories	170	54,337.53	319.63	942.45	26.72	6.92	2,587.50
Fashion accessories	178	54,305.90	305.09	1,042.65	12.69	7.03	2,586.00
Food and beverages	174	56,144.84	322.67	1,034.46	22.66	7.11	2,673.56
Health and beauty	152	49,193.74	323.64	950.25	18.64	7	2,342.56
Home and lifestyle	160	53,861.91	336.64	1,023.75	14.68	6.84	2,564.85
Sports and travel	166	55,122.83	332.07	1,002.12	10.68	6.92	2,624.90

Customer Type Analysis

Customer Type	Count	Total Sales (\$)	Avg Sales (\$)	Median Sales (\$)	Max Sales (\$)	Min Sales (\$)	Total Income (\$)
Member	565	189,694.76	335.74	271.95	1,042.65	10.68	9,033.08
Normal	435	133,271.98	306.37	225.79	1,039.29	12.69	6,346.28

Payment Method Preferences

Payment Method	Count	Total Sales (\$)	Avg Sales (\$)	Median Sales (\$)	Max Sales (\$)	Min Sales (\$)
Cash	344	112,206.57	326.18	262.96	1,003.59	10.68
Credit card	311	100,767.07	324.01	239.59	1,042.65	12.69
Ewallet	345	109,993.11	318.82	253.68	1,034.46	13.42

Group 1: Comparisons

Figure 1: Total Sales per Branch

A bar chart showing the total revenue of branches A, B, and C, indicating the branch that performs the best in terms of sales.

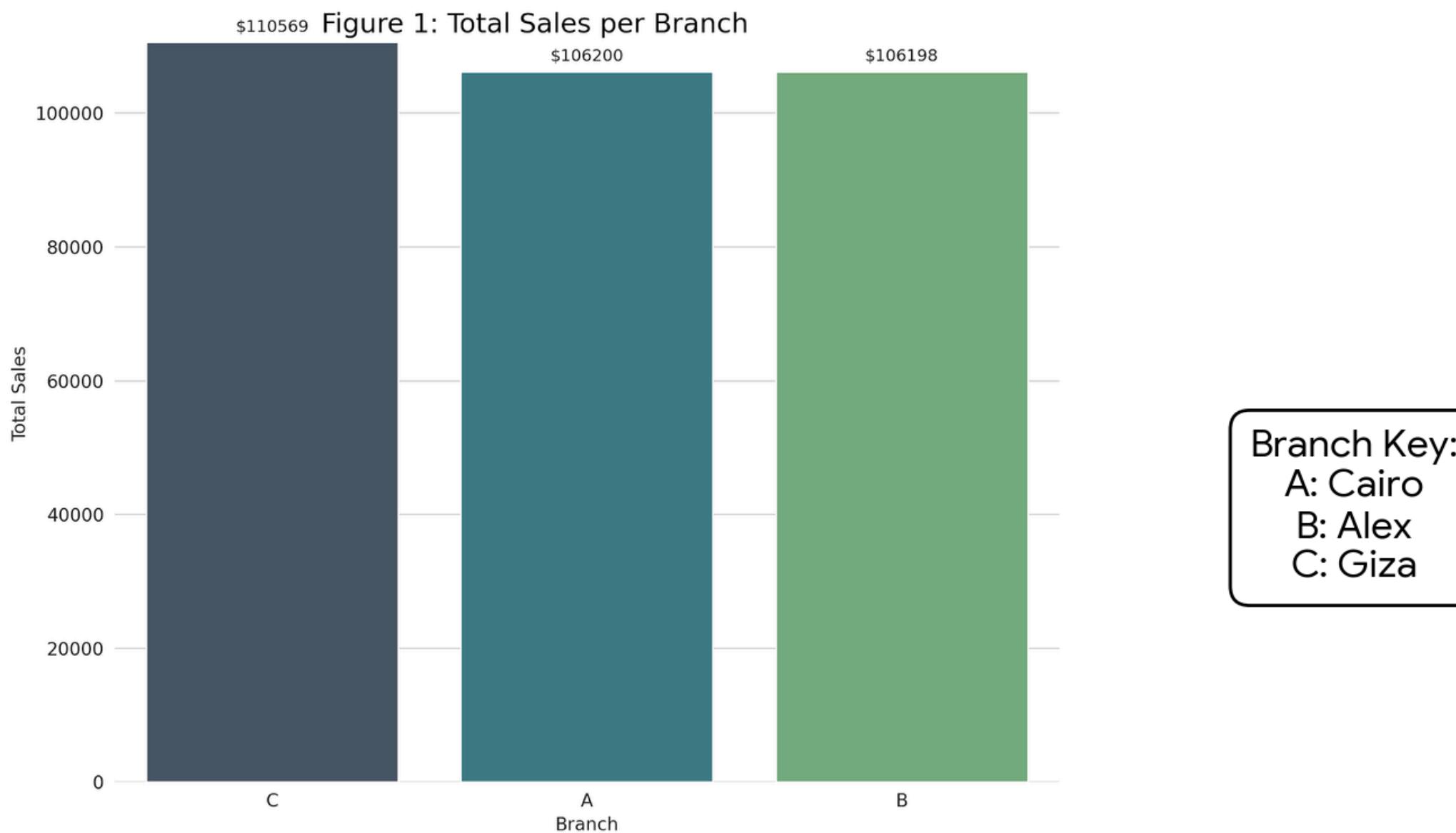


Figure 2: Overall Sales per Product Line

Description: The above graph shows total sales for various categories, indicating which categories are most profitable to the supermarket.

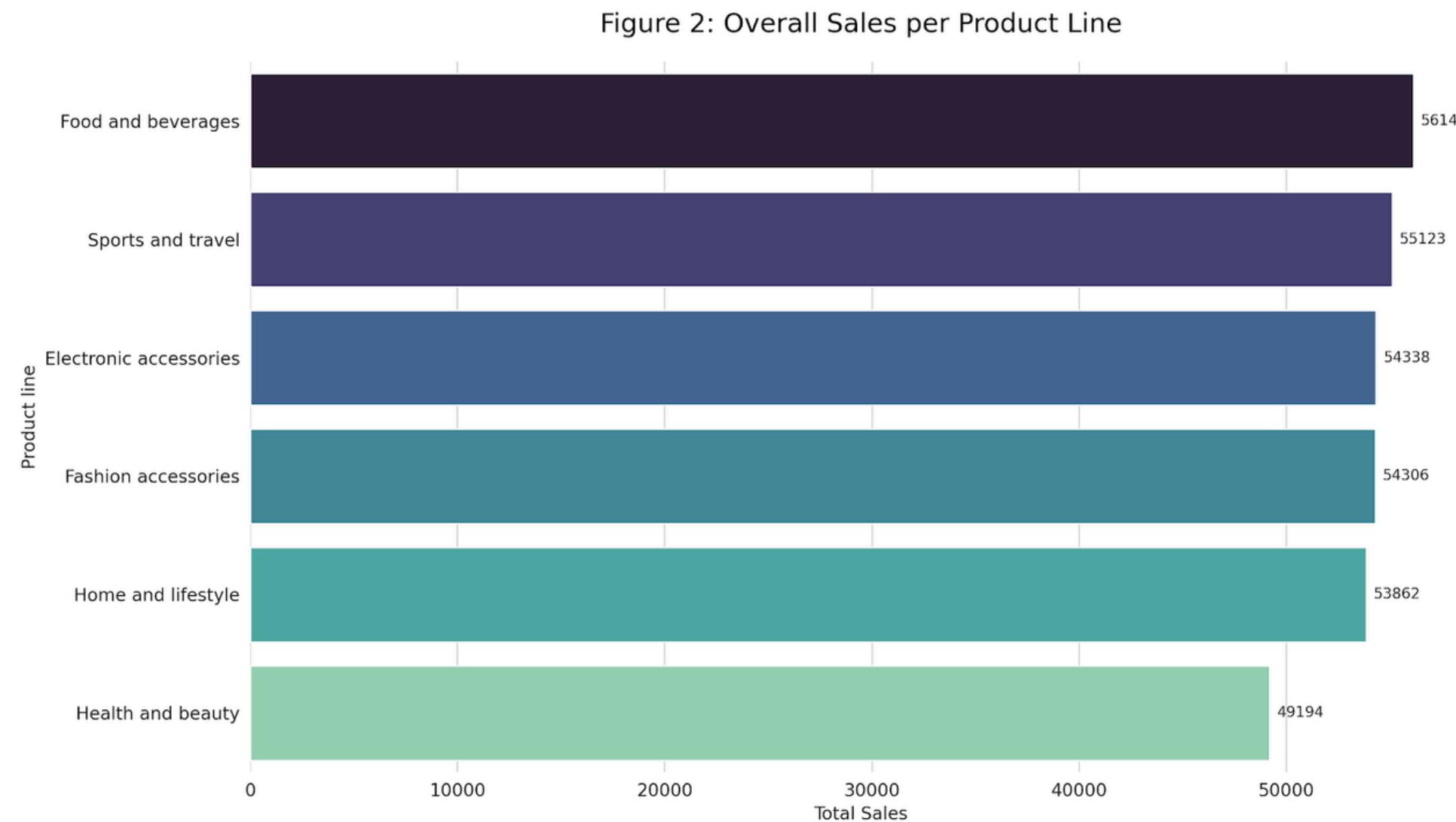


Figure 3: Average Customer Rating by City

Description: This graph provides the overall customer satisfaction rating for each city, making it easier for identifying areas where the service is better.

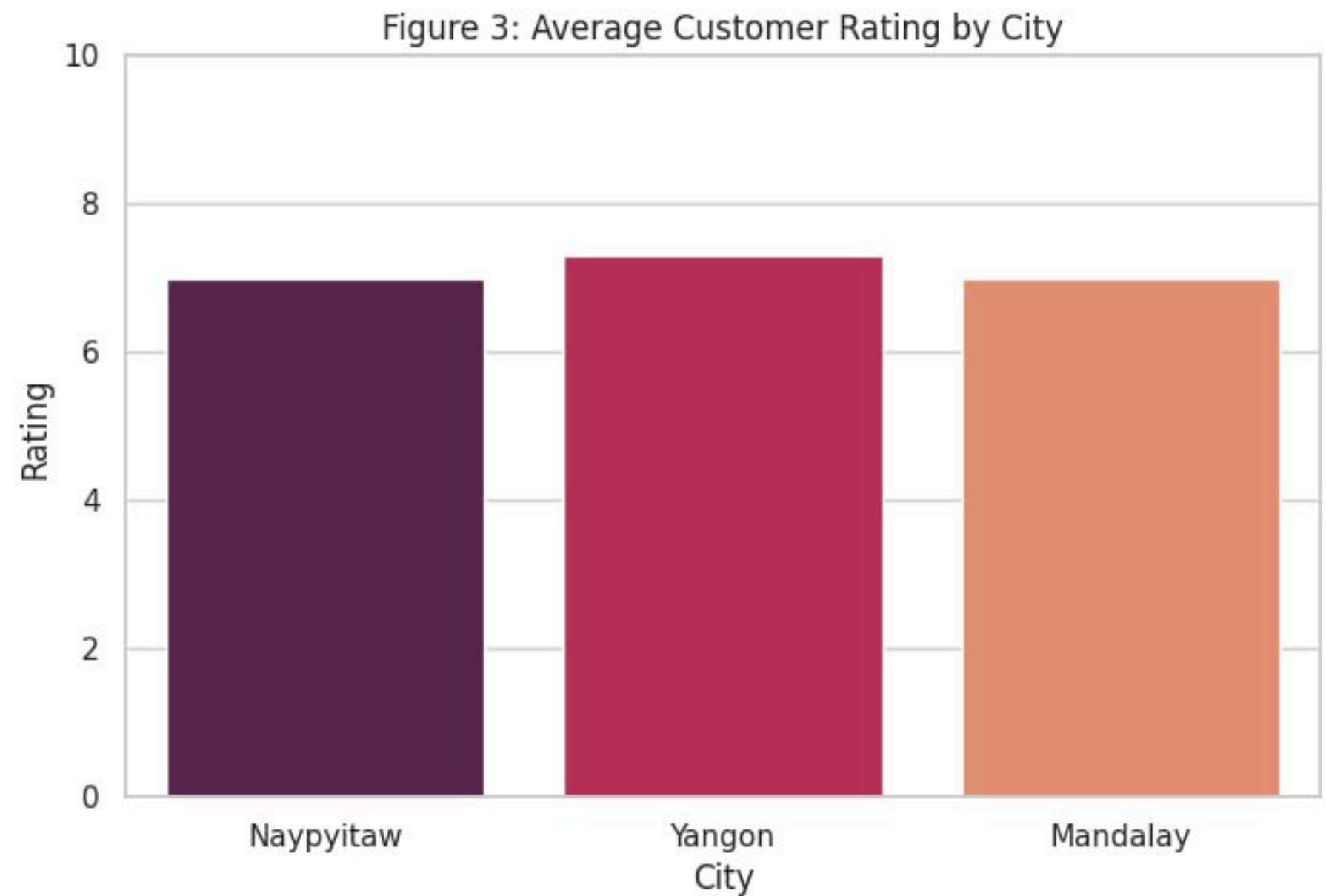


Figure 4: Transaction Frequency by Payment Method

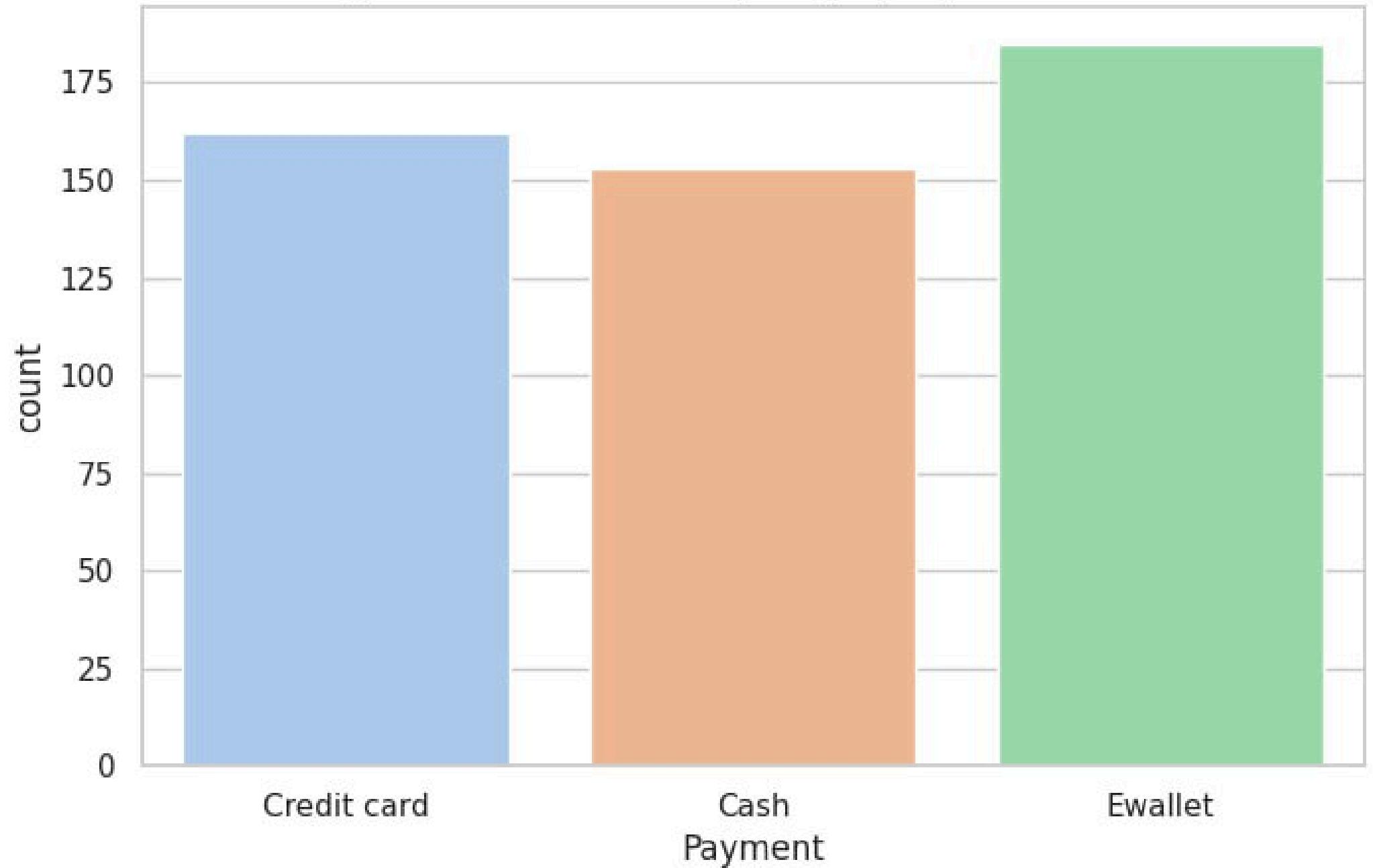
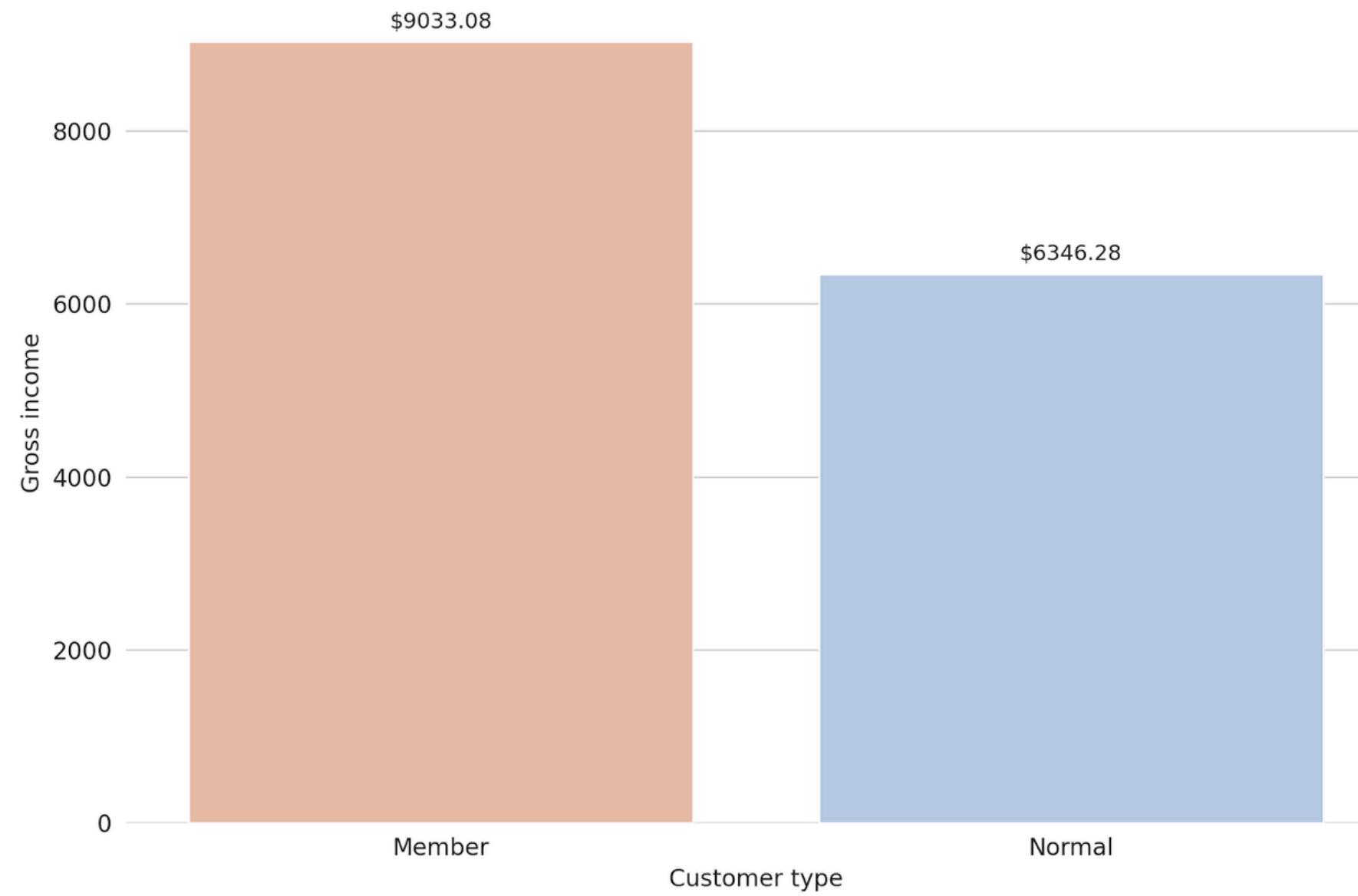


Figure 4: Transaction frequency by method of payment

Description: This count plot visualizes the most popular payment methods used by customers, namely Cash, Ewallet, and Credit Card, to depict customer preference.

Figure 5: Total Gross Income by Customer Type

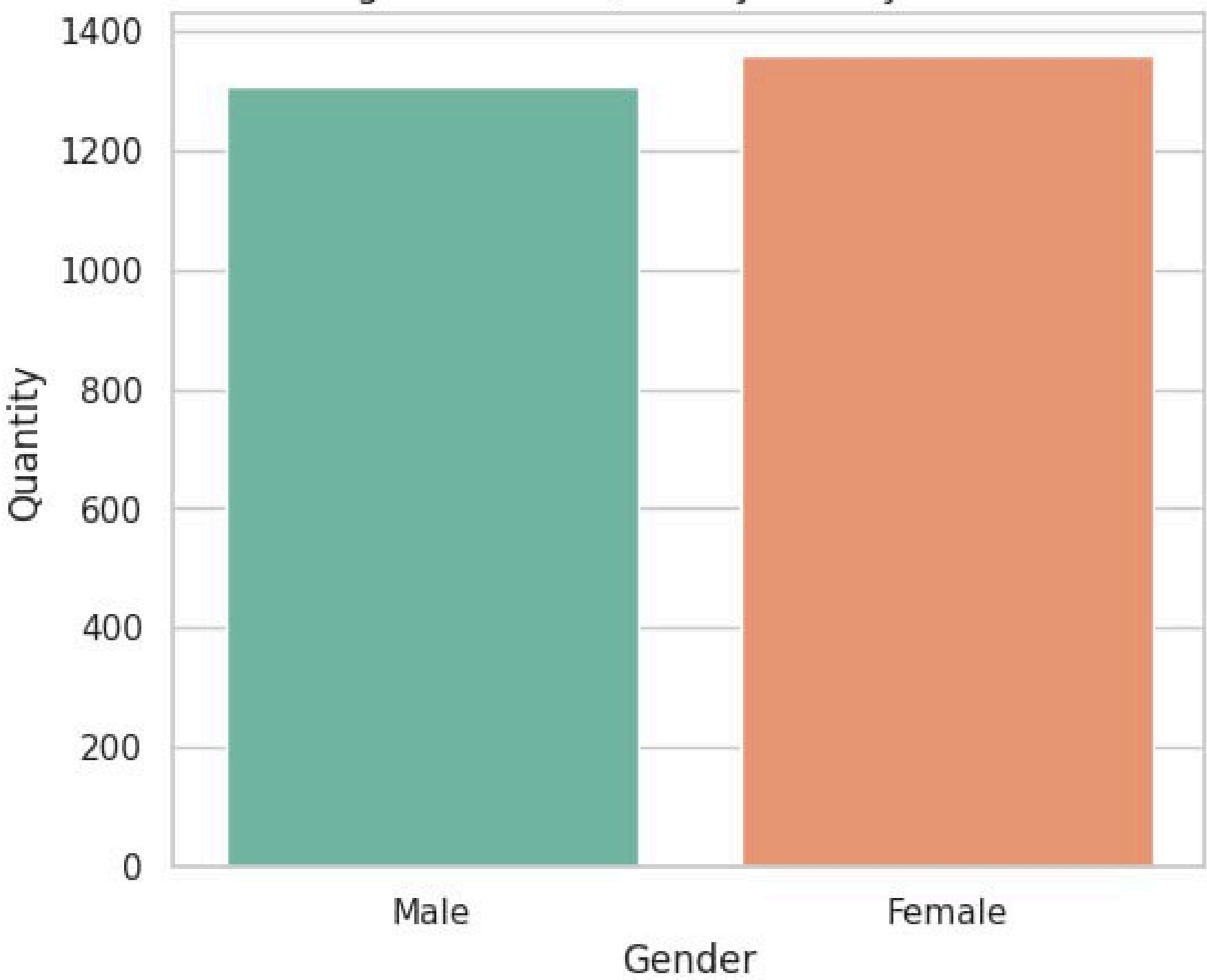


provides a comparison of total gross income from Member customers and Normal customers to emphasize the benefits of signing up for a membership program.

Figure 6: Total Quantity Sold by Gender

Data Interpretation: The graph demonstrates the cumulative purchase of products by males and females to understand which gender has a tendency to purchase more products.

Figure 6: Total Quantity Sold by Gender



Group 2: Distributions

In Figure 7, a histogram of total transaction amounts is displayed. It represents how often different amounts are charged on bills, indicating frequent amounts paid (e.g., \$50 or \$100).

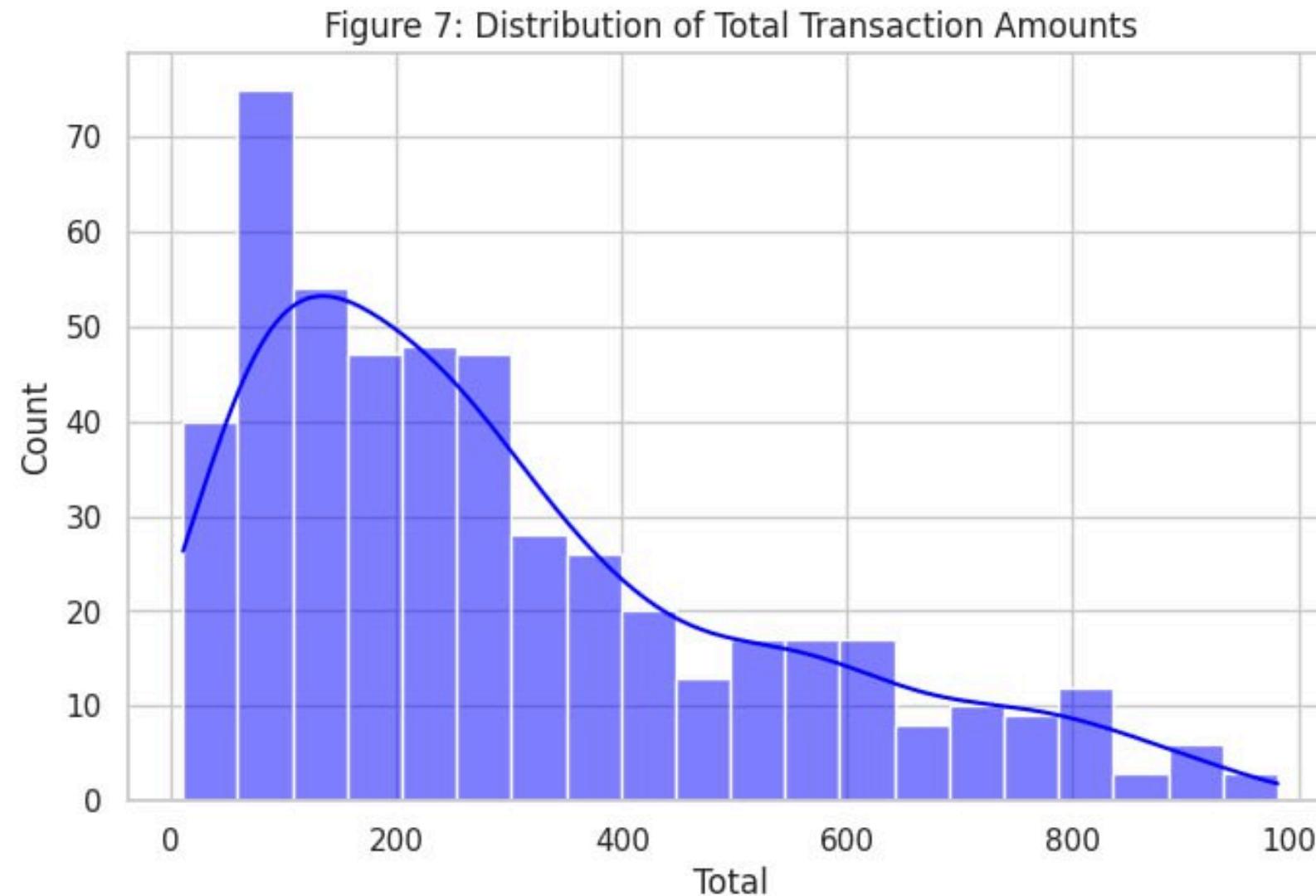


Figure 8: Distribution of Customer Ratings

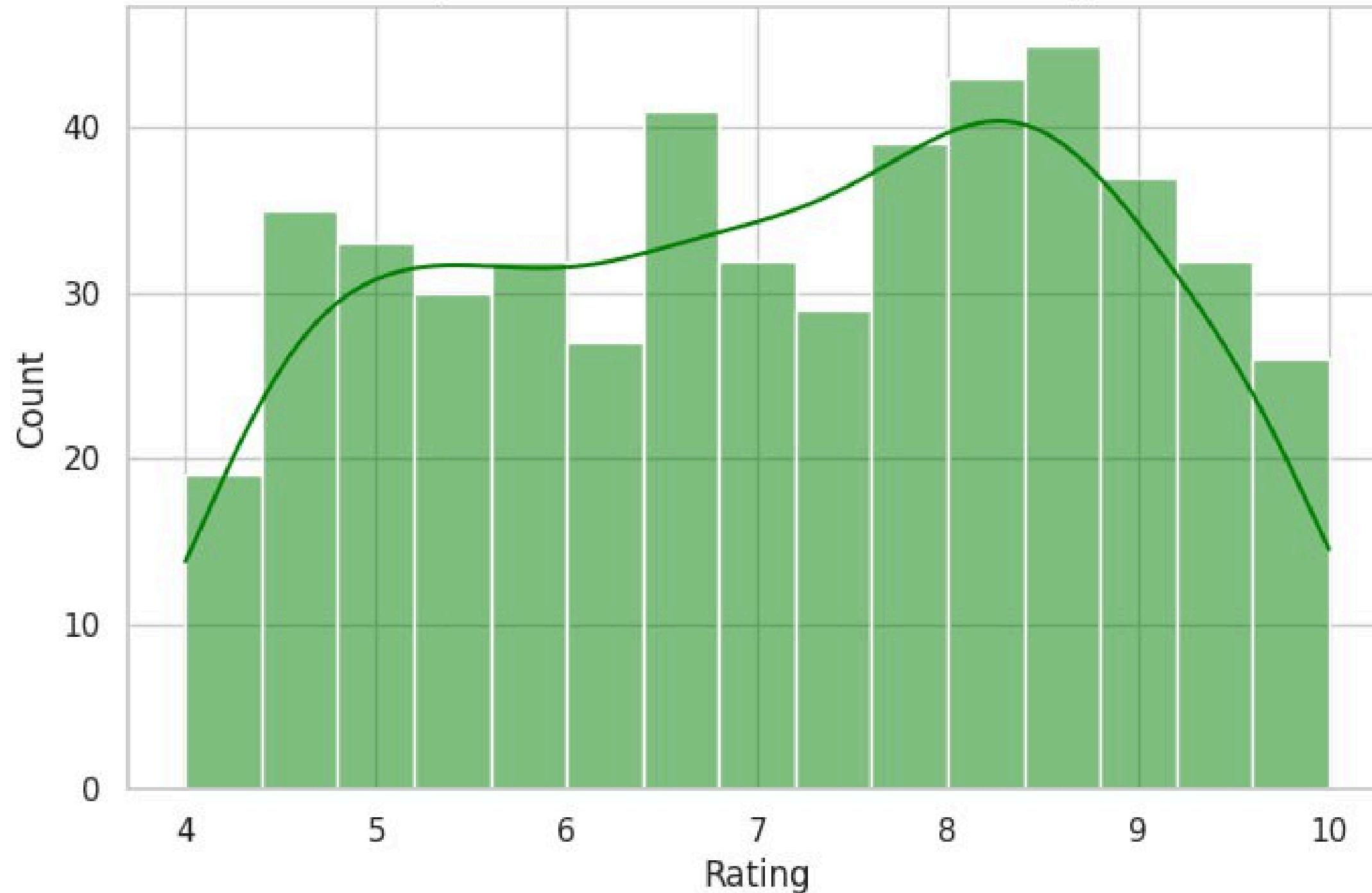


Figure 8: Distribution of Customer Ratings

Description: This above histogram displays the distribution of customers' ratings, which will provide the level of general satisfaction of shoppers, whether the ratings are positive or negative.

Figure 9: Distribution of Product Unit Prices

Description: This histogram shows the unit price range for the items that are most frequently purchased by consumers based on their prices.

Figure 9: Distribution of Product Unit Prices

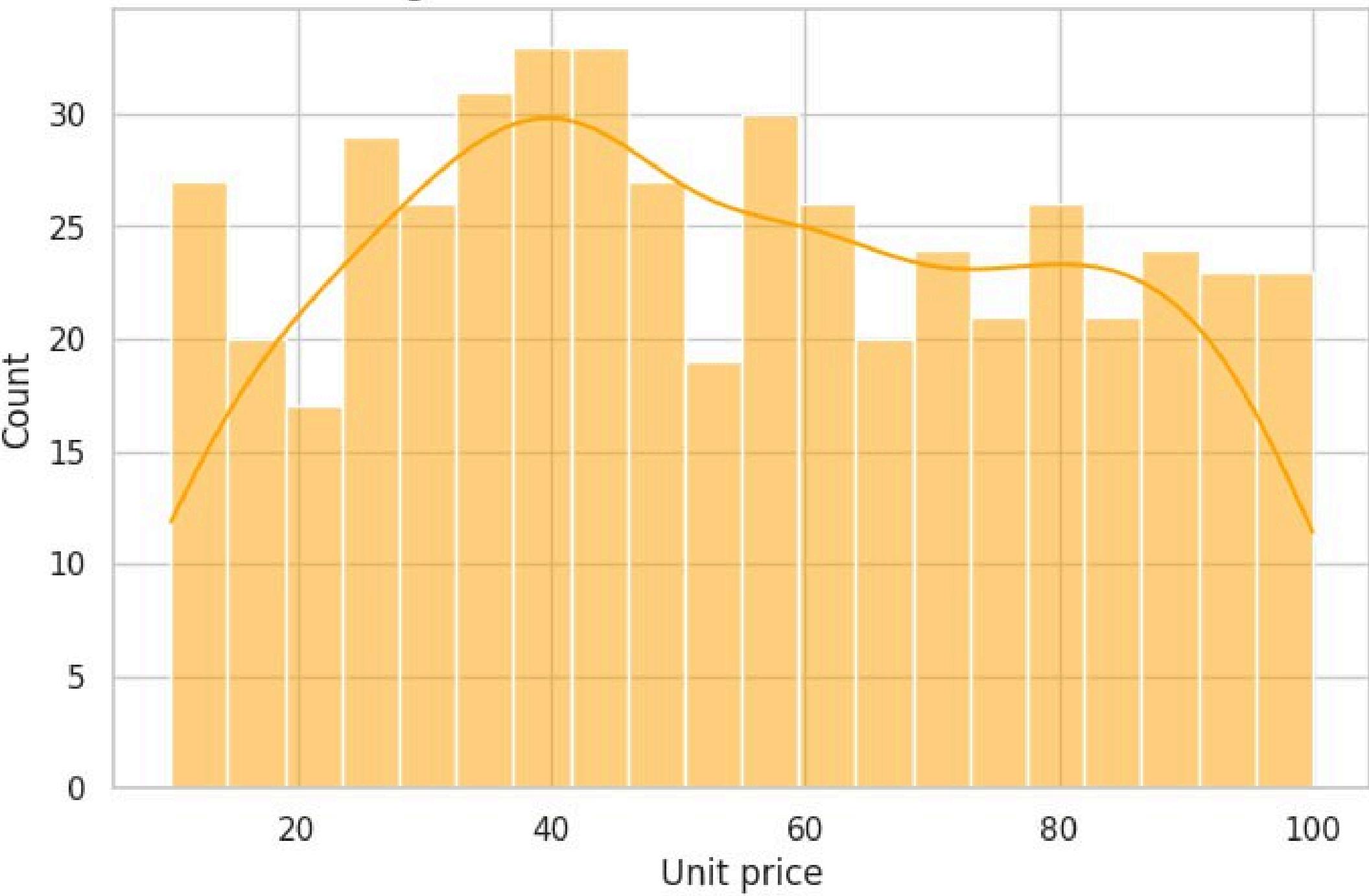


Figure 10: Gross Income Distribution by Branch

Description: This box plot provides the distribution of the gross income per transaction of each branch and the potential existence of outlier values regarding the profitability of the branch.

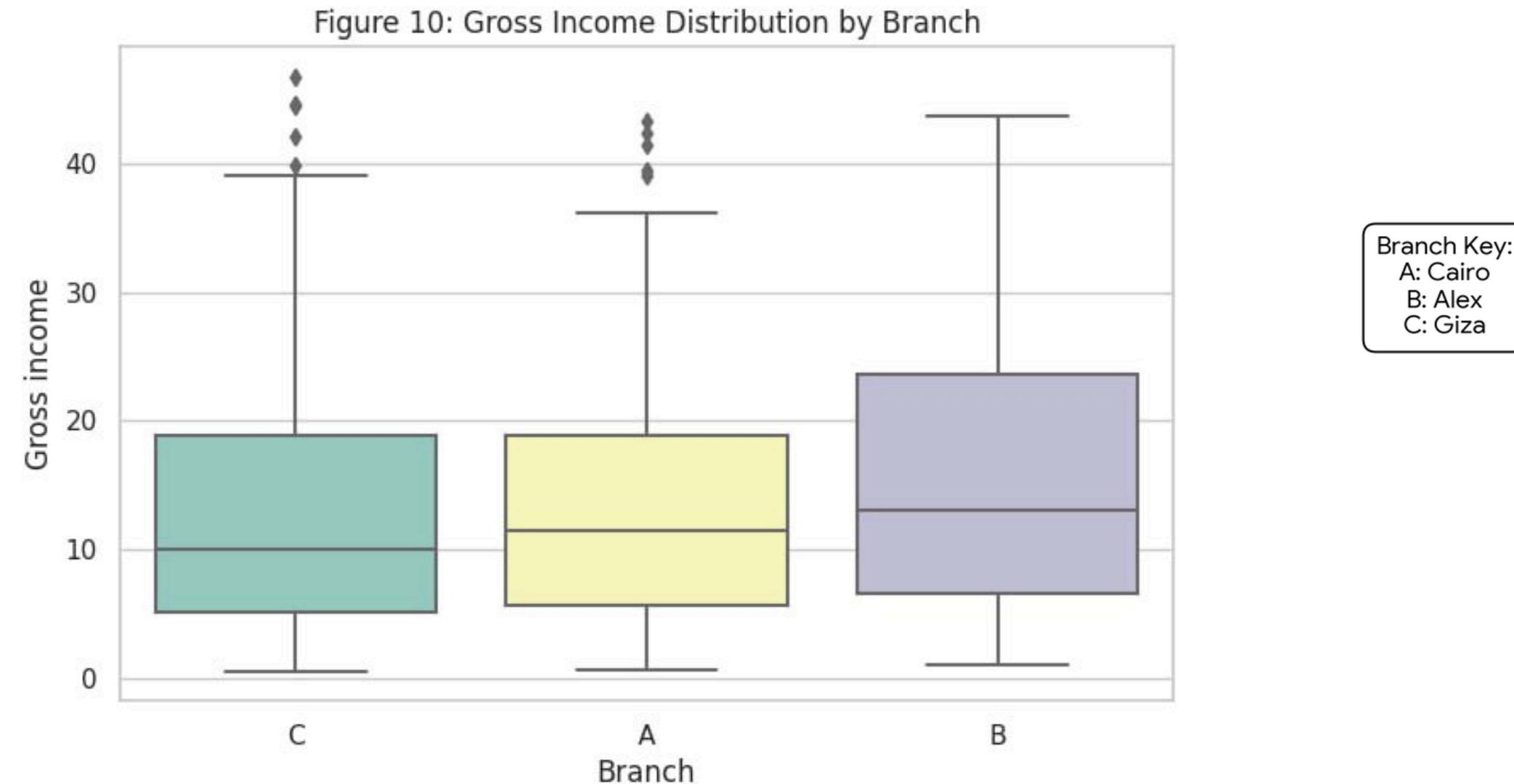
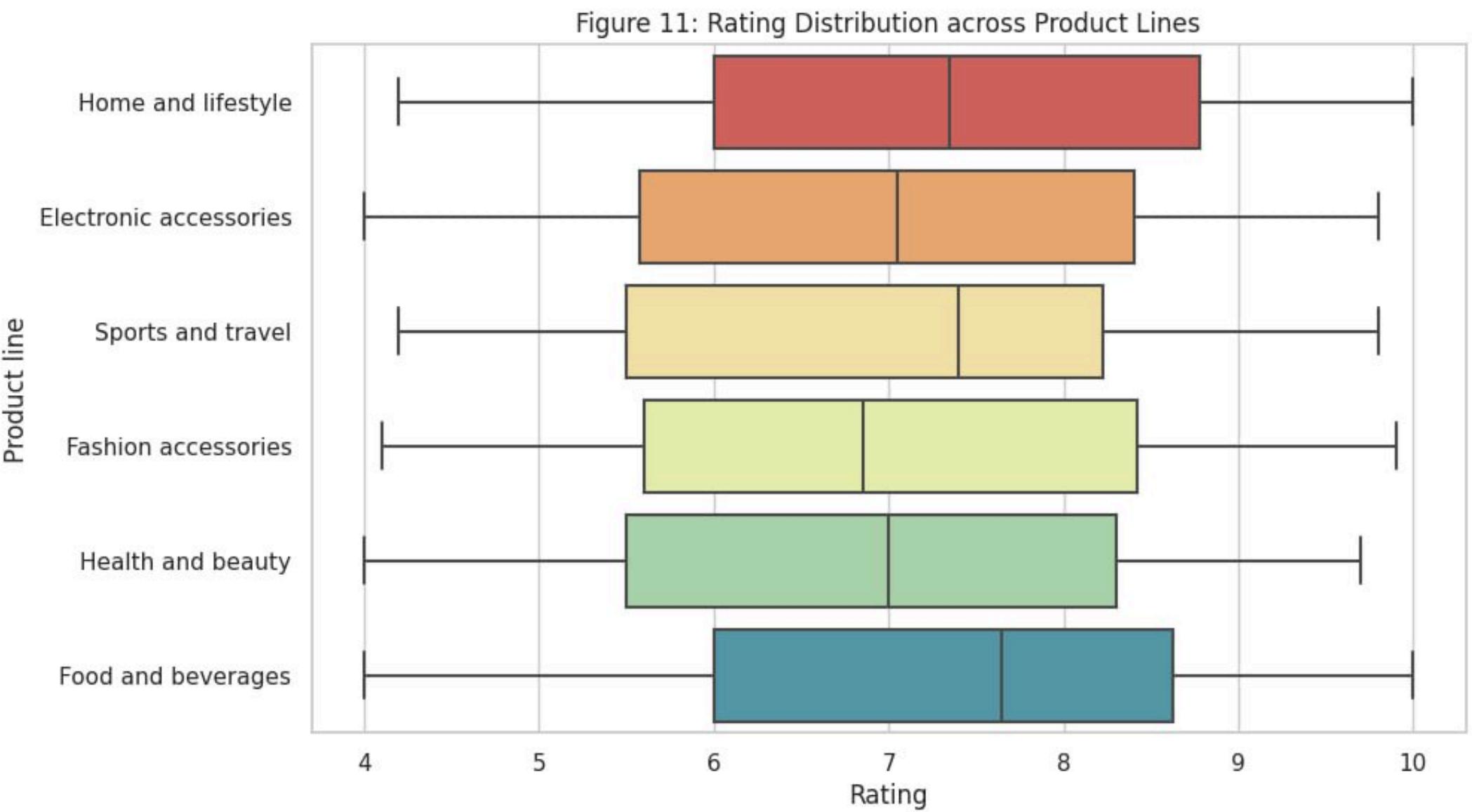


Figure 11: Rating Distribution across Product Lines

Description: This box plot compares customer ratings across different product lines to see which categories have the most consistent quality and satisfaction.



Group 3: Relationship

Figure 12: Relationship between Unit Price and Quantity

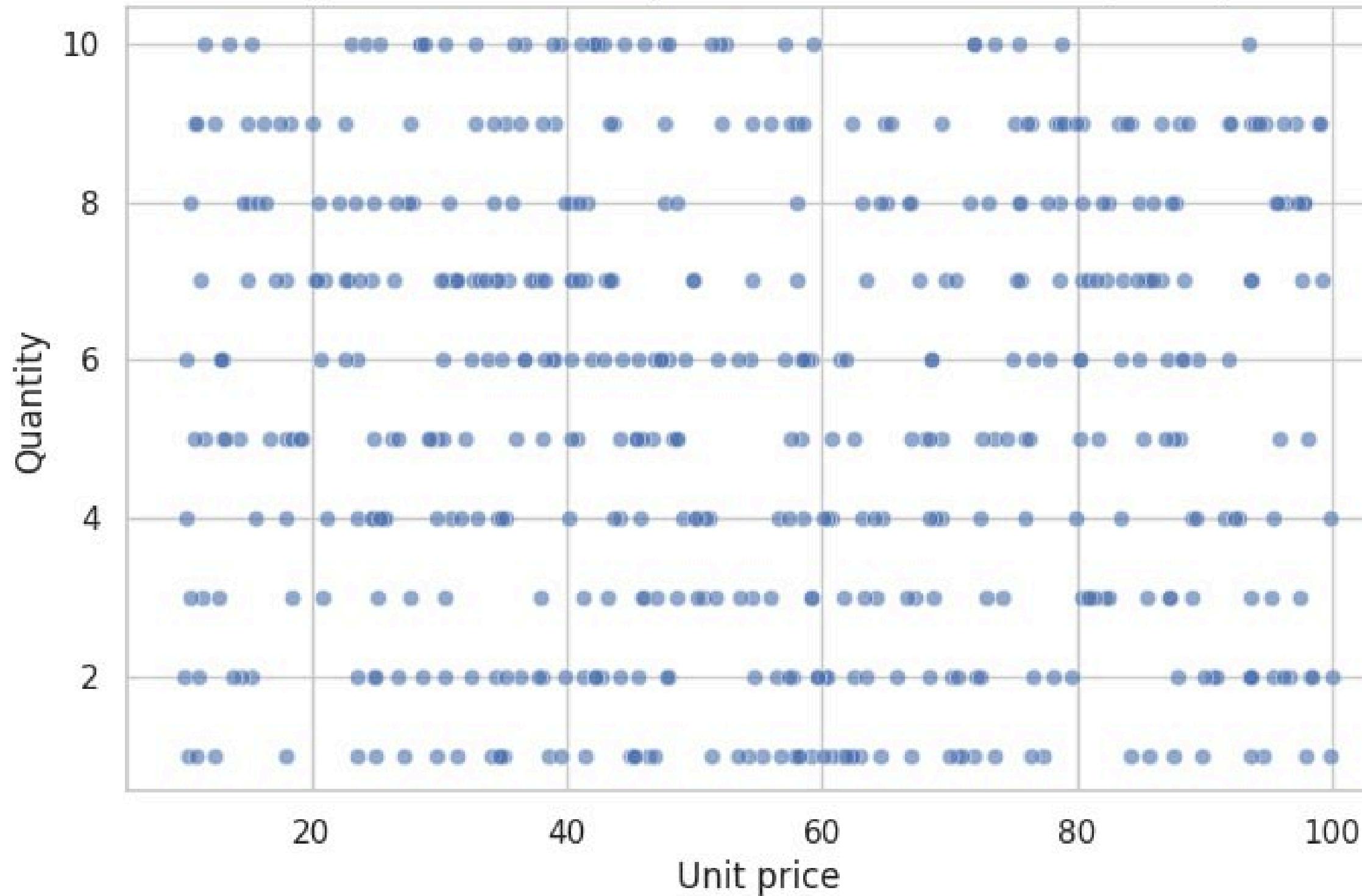


Figure 12: Correlation
between Unit Price and
Quantity

Description: The scatter plot, This scatter plot is created to analyze how the price of items influences the quantity bought concerning whether items with a higher price result in fewer sales.

Figure 13: Correlation between Total Bill and Tax

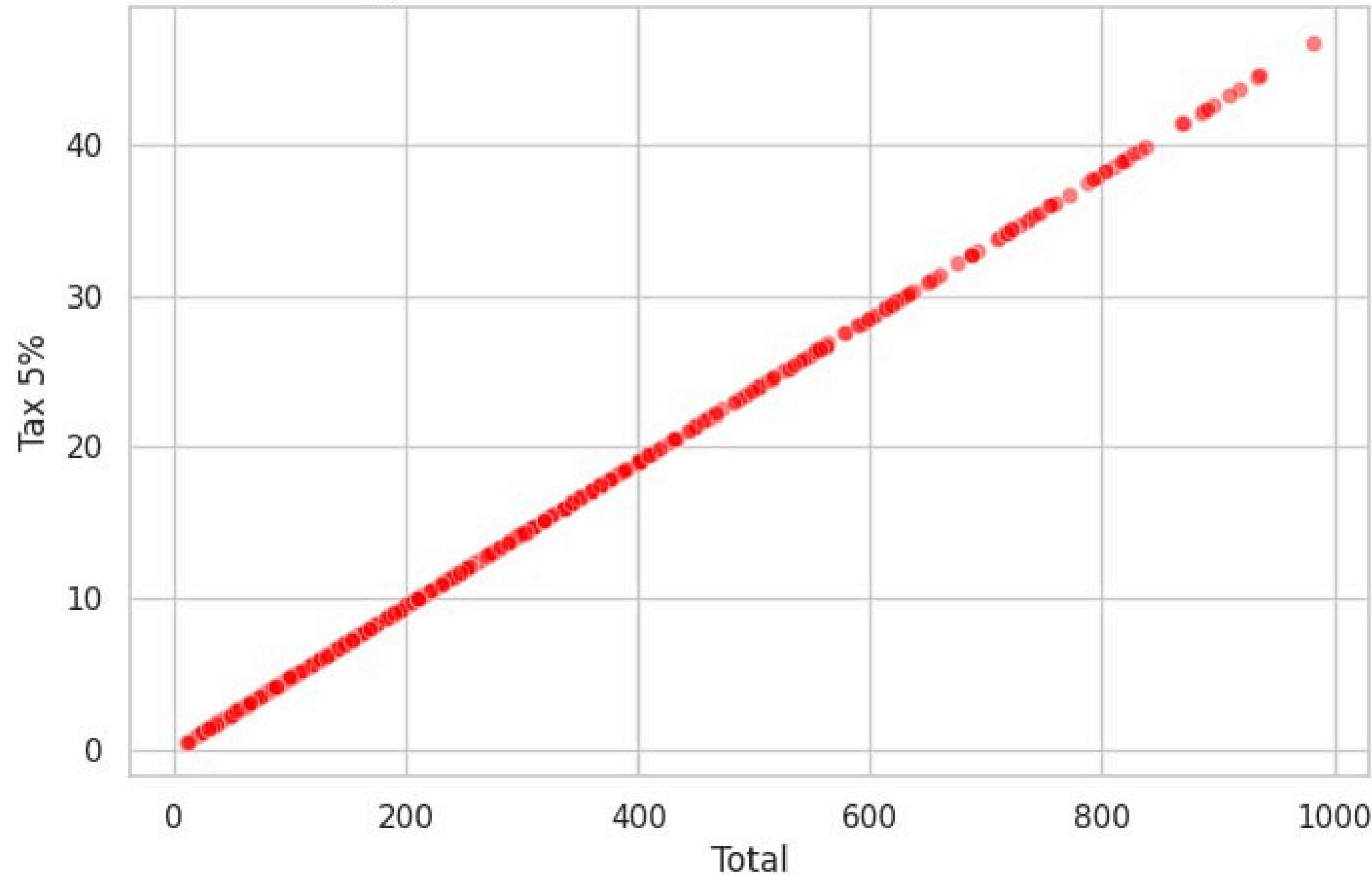


Figure 13: Correlation Between Total Bill and Tax

Description: This scatter chart verifies the linear correlation between the total amount billed and the tax (5%).

Figure 14: Rating vs Gross Income

Description: The goal of this scatter plot is to explore whether more lucrative transactions are related to enhanced ratings, or whether a correlation does not exist.

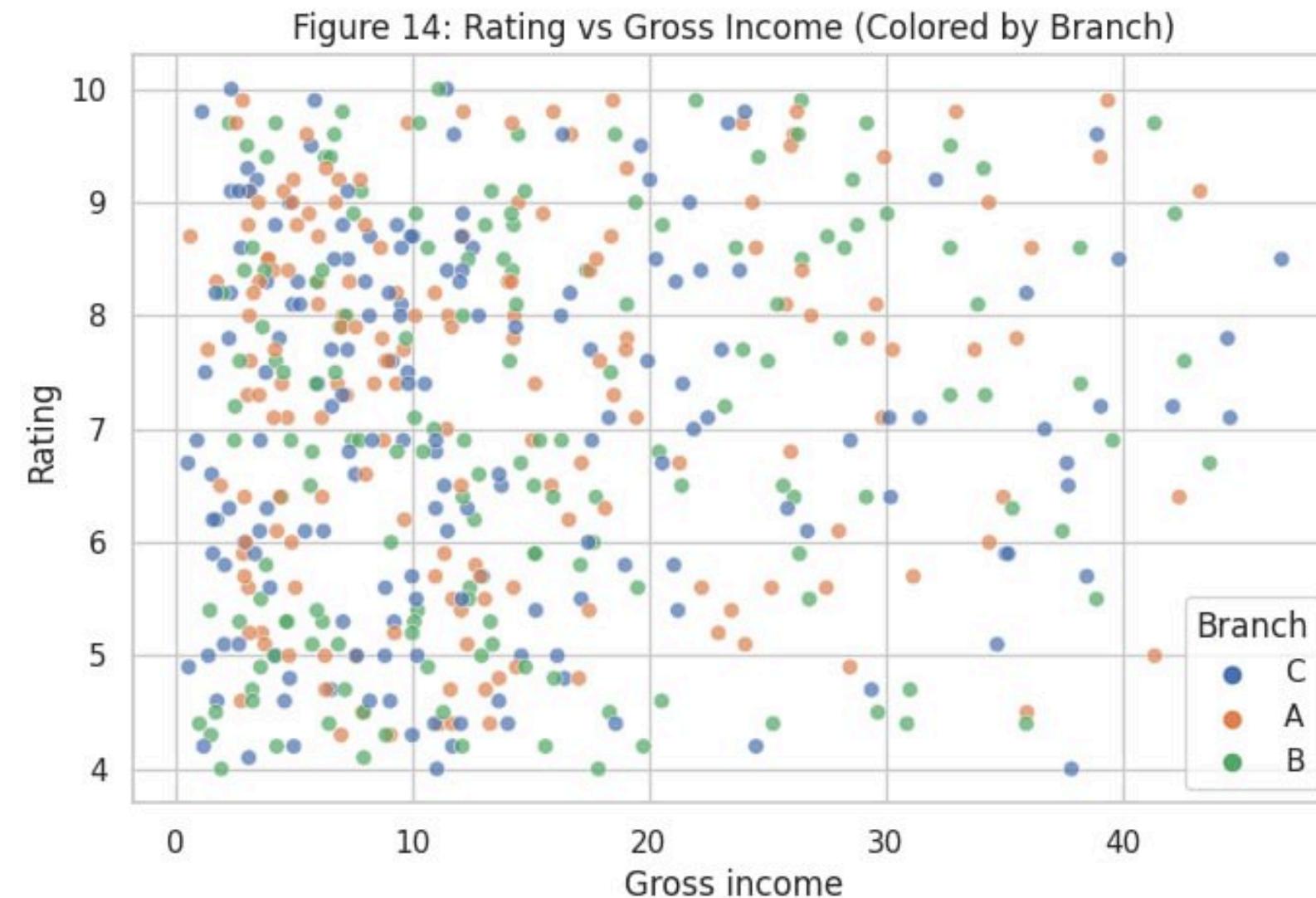
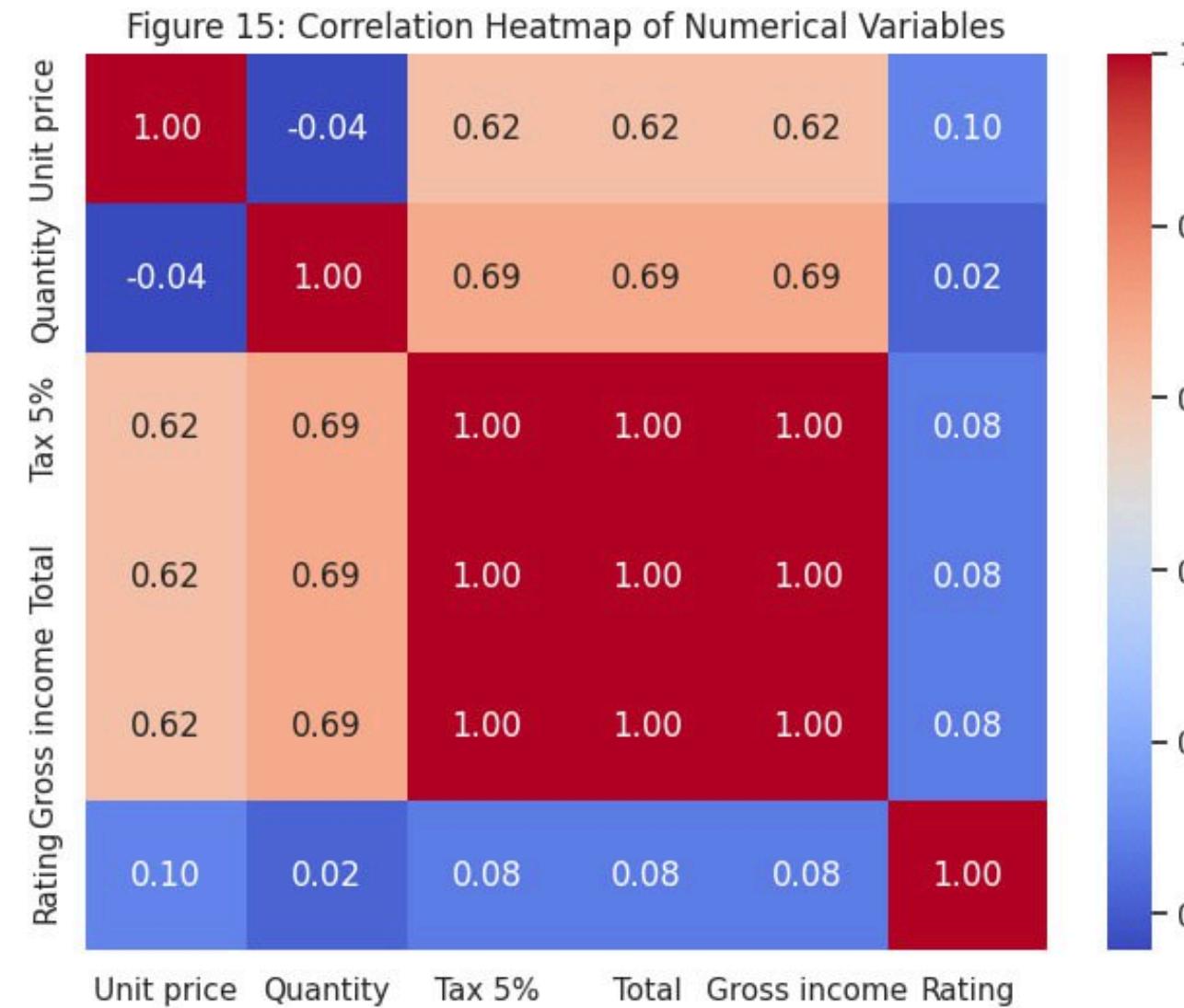


Figure 15: Correlation Heatmap of Numerical Variables

Description: This heatmap shows how all the numerical variables are correlated, in order to identify underlying links, such as a strong link between COGS and Total.



Group 4: Composition & Trends

Figure 16: Daily Sales Trend Over Time

Description: This line chart projects overall daily sales with time, enabling one to identify peak sale days and monitor performance throughout the period.

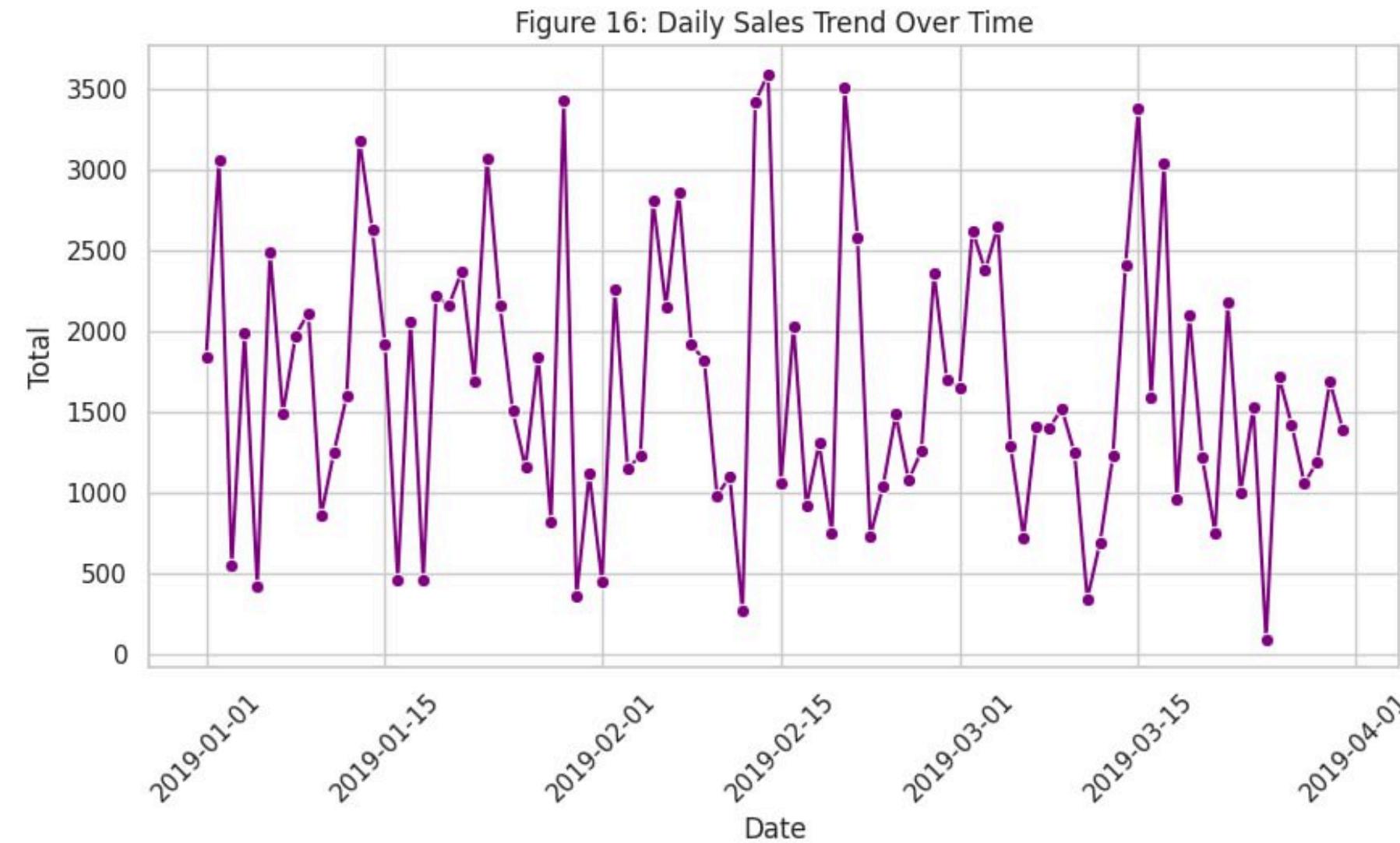


Figure 17: Market Share by Branch (Total Sales)

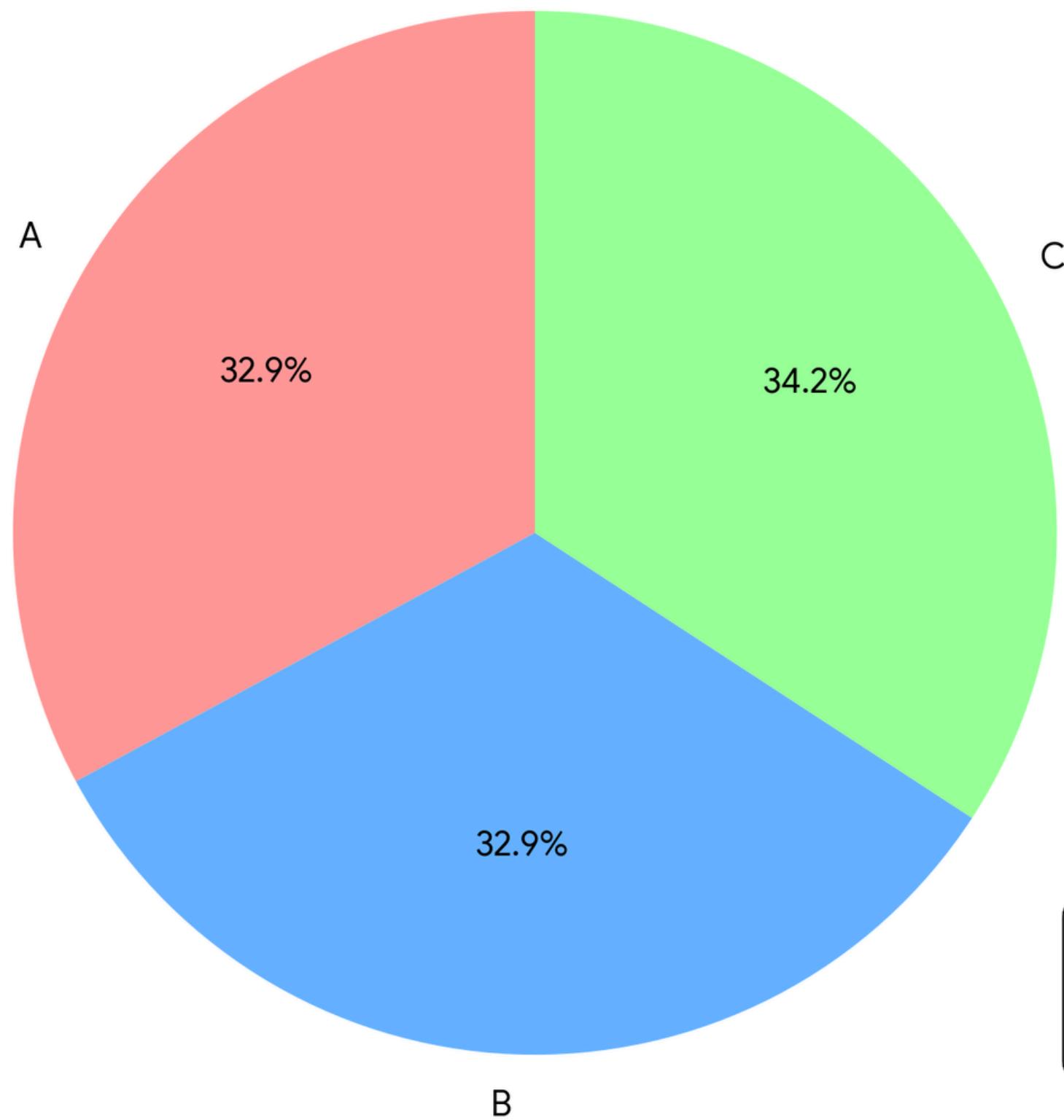


Figure 17: Market Share by Branch (Total Sales)

Description: A pie chart has been used in the graph to display the contribution of each of the branches towards the total sales revenue.

Figure 18: Proportion of Member vs Normal Customers

Figure 18: Proportion of Member vs Normal Customers

Description: The pie in the picture represents the share of customers who are members versus regular customers, which forms the base of the customer composition.

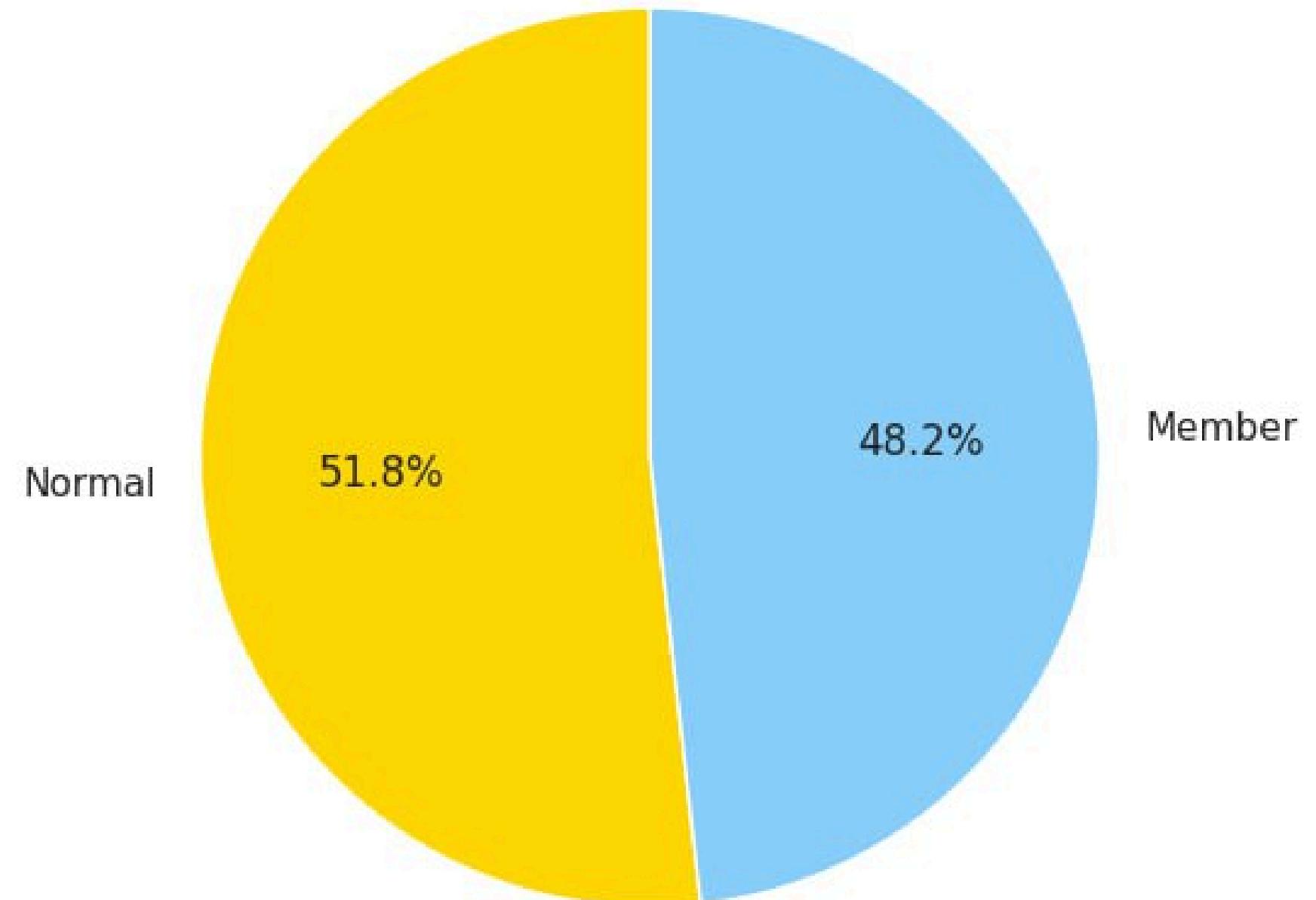


Figure 19: Product Line Sales Breakdown by Gender

Description: This stacked bar chart breaks down product line sales by gender, showing purchasing preferences for men and women within each category.

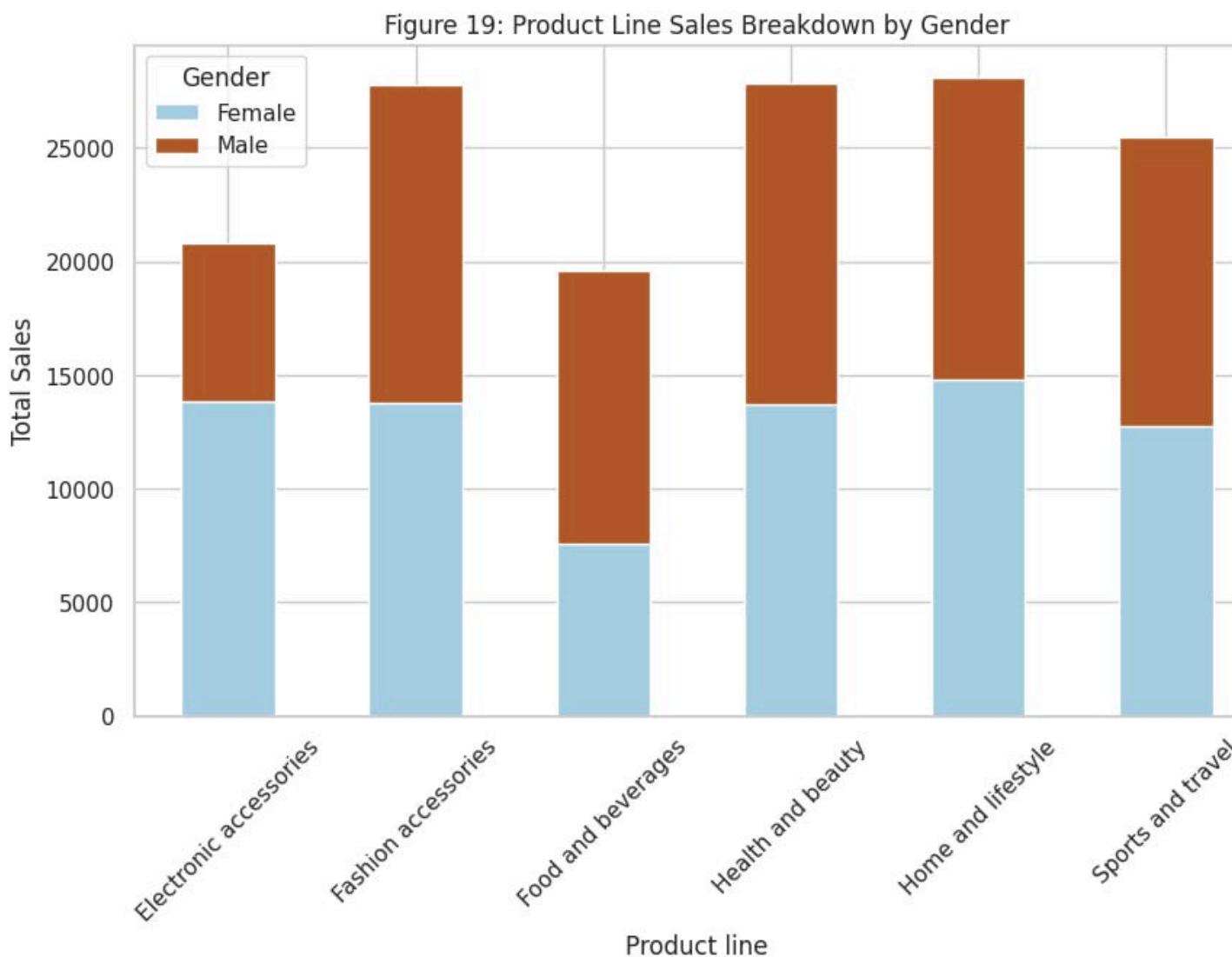
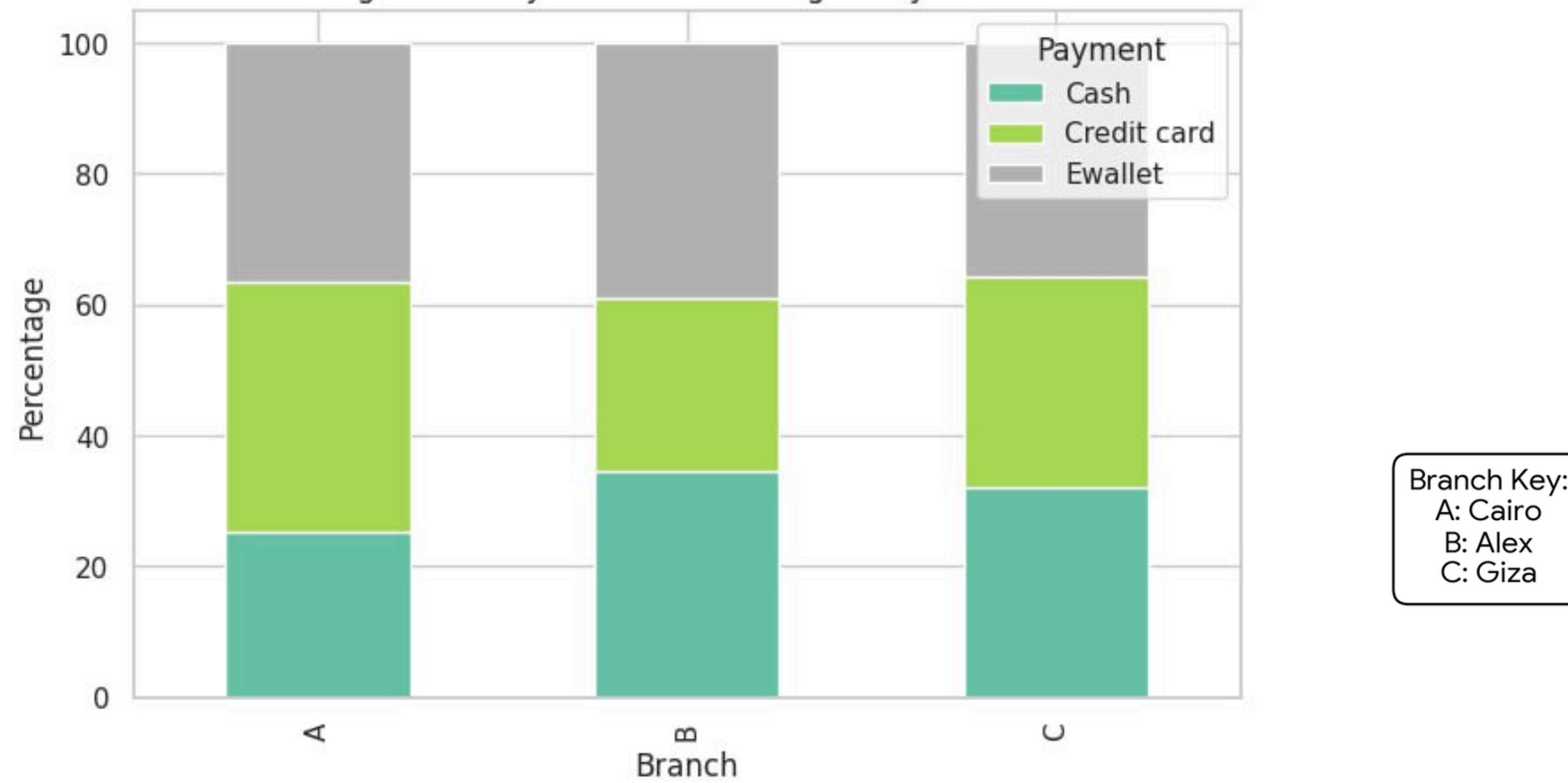


Figure 20: Payment Method Usage % by Branch



Description: This stacked bar chart shows the percentage usage of different payment methods within each branch, revealing if certain branches prefer specific payment types.

Data-set Description

The data-set includes detailed information about supermarket transactions.

The **main attributes** available in the data-set are :

- Branch
- City
- Customer type
- Gender
- Product line
- Unit price
- Quantity
- Sales
- Gross income
- Customer rating
- Date and payment method

Data Exploration and Cleaning

The data-set was imported using the Pandas library.

Initial inspection showed that the dataset is clean, with no missing values.

All columns were found to have appropriate data types, allowing the analysis to proceed without additional preprocessing steps.

Exploratory Data Analysis (EDA)

Several visualizations were created to gain deeper insights into the dataset, including:

- Distribution of payment methods
- Sales performance across branches
- Relationship between sales and customer ratings
- Gross income comparison by product line
- Sales trends over time
- Correlation heatmap between numerical variables

These visualizations helped highlight key trends, customer preferences, and important relationships within the data.

Answers 5 analytical questions

Q1- Which branch generates the highest total revenue and why?

Based on the analysis, the **Giza** branch generates the highest total revenue. This is due to a larger number of transactions and stable sales performance across multiple product lines, leading to higher overall revenue compared to other branches.

Q2- Do members spend more than normal customers?

Yes, members tend to spend more than normal customers.

The results indicate that member customers have higher total sales, likely because of loyalty programs, special offers, and more frequent shopping behavior.

Q3- What is the most commonly used payment method?

The most commonly used payment method is **Ewallet**.

This suggests that a significant number of customers doesn't prefer paying with cash they prefer using credit cards or Electronic wallets.

Q4-Which product line receives the highest average rating?

The Food and Beverages product line has the highest average customer rating.

This indicates higher customer satisfaction in this category compared to other product lines.

Q5- What is the relationship between unit price and quantity sold?

There is **no significant relationship** between unit price and quantity sold (Correlation coefficient ≈ 0.01). This indicates that the price of an item does not influence the quantity purchased per transaction; customers buy the quantity they need regardless of whether the item is cheap or expensive.

Conclusion

In conclusion, the above project work throws light on the key points contributing to the success of the supermarket operation.

Finally, the ability to translate these insights into workable strategies will help the company optimize and grow sustainably, as well as improve the shopping experience, as it pertains to all of its branches.