Pattern Discovery of Supermarket Sales Transactions using EDA

BY: ARYAN DAHIYA

Abstract: *In order to identify trends in sales, profitability indicators, and customer behaviour, this study offers a thorough examination of supermarket sales data. Using a range of visualization strategies and thorough Exploratory Data Analysis (EDA), the goal is to pinpoint important information like customer evaluations, gender-based spending, payment preferences, and product line success. Analysing data distributions, spotting long-term trends, and investigating correlations between various variables are all part of the EDA process. The findings shed light on important factors including popular product lines, peak sales times, and gender-specific buying patterns. This knowledge is helpful for making strategic decisions in retail operations. Patterns that can assist optimize marketing tactics, improve inventory management, and boost customer happiness are revealed through in-depth visualizations and statistical analysis. Furthermore, temporal patterns in sales are investigated, including the busiest times of day and the most lucrative days of the week. Retailers may increase their operational efficiency and profitability by utilizing the actionable insights provided by this comprehensive investigation.*

Key Words: Exploratory Data Analysis, Correlations, Lucrative

INTRODUCTION:

Any retail organization that wants to improve operations and customer happiness must understand sales statistics. In order to derive important insights that can inform corporate strategy, this analysis focuses on sales data from supermarkets. Data on product lines, sales transactions, consumer demographics, and other pertinent information are all included in the dataset. Through the examination of these aspects, patterns and trends that can guide the development of marketing strategies, inventory control systems, and customer service enhancements are discovered. In order to provide useful data for strategic decision-making, the analysis attempts to unearth important insights pertaining to consumer behaviour, sales trends, and profitability measures. Through an extensive Exploratory Data Analysis (EDA), the research explores a number of topics, including customer ratings, payment preferences, gender-based spending patterns, and the effectiveness of various product lines. Businesses can also maximize personnel and inventory levels by analysing temporal sales patterns to pinpoint peak sales periods. With the use of these insights, retailers are better equipped to make data-driven decisions that increase profitability, optimize operational efficiency, and raise customer satisfaction levels.

EXPLORATORY DATA ANALYSIS:

The gender distribution is shown in the pie chart, where each gender makes up exactly 50% of the data. Depending on the context, this balanced representation may indicate a fairly even client base or study group. Although this offers a foundation for analysis, more research is necessary to comprehend the consequences for product development, marketing, and other pertinent aspects. Accurate conclusions must also take into account the sample size and any potential biases in the data collection process.
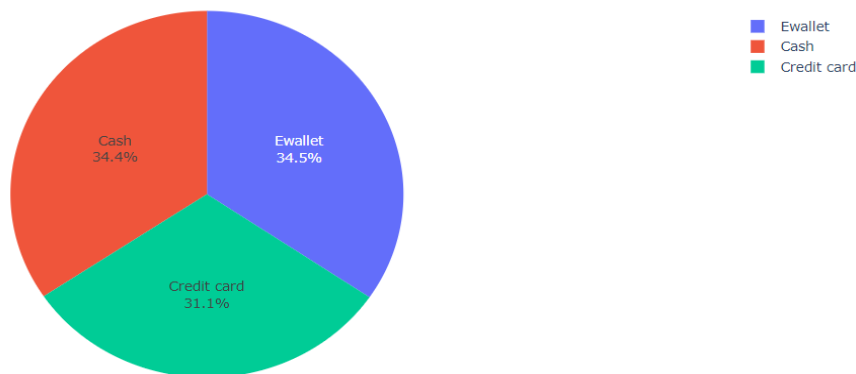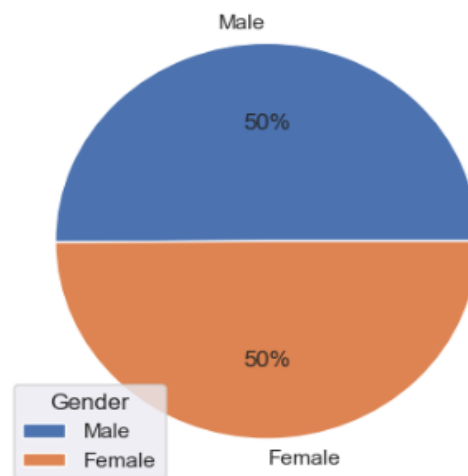

The distribution of consumer payment methods is shown in the pie chart. The most popular payment method is e-wallets, which overtake credit card and cash transactions in usage. This suggests that customers strongly favor digital payment options. Although it is still an option, cash is used far less frequently than technological alternatives. Among the three payment methods, credit cards have the lowest utilization rate despite being a conventional option.

These three scatter plots show an association between customer rating for various payment methods (Cash, Credit Card, and E-wallet) and gender (male and female). A cluster of data points representing specific clients is shown in each plot. However, there appears to be a broad distribution of scores for each payment method between the genders.
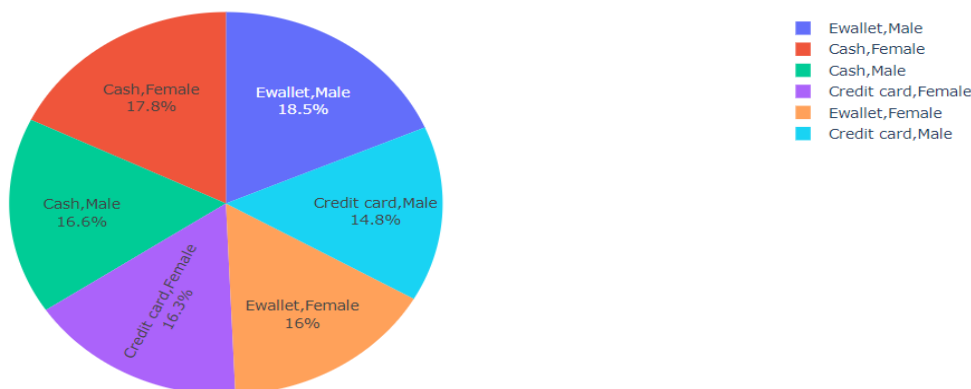
The payment preferences of each gender are visually represented in a pie chart that divides transactions into six categories: cash and e-wallets for both genders, and credit card usage for both. Notably, the most common payment methods—Cash, Female, and Ewallet, Male—account for almost 18% of all transactions. Credit card transactions seem to be less popular across both genders, with a combined share of about 31%. This illustration provides a basic grasp of payment patterns and raises the possibility of gender variations in payment behaviour

Customer Rating Distribution
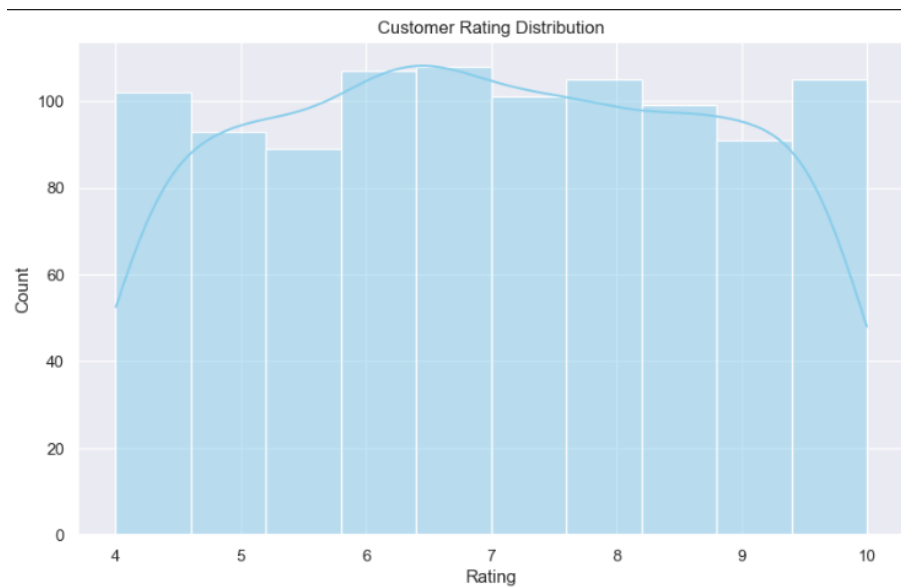
The distribution of customer ratings for a supermarket is shown by this histogram. The rating scale, which runs from 4 to 10, is represented by the x-axis, and the number of customers who gave that rating is shown by the y-axis. Plotting indicates a largely positive emotion, with a cluster of ratings in the 6–8 range. This implies that a sizable percentage of clients are content with their purchasing encounter. A smoother depiction of the distribution is provided by the overlay curve, a kernel density estimate (KDE), which highlights probable peaks and valleys in customer sentiment.



Aggregate Sales Across Branches
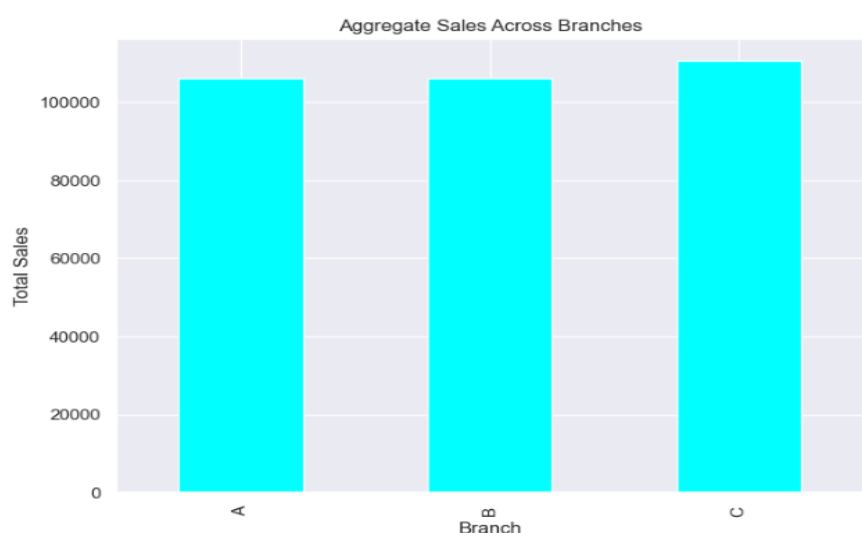
A visual comparison of the total sales for the three branches, A, B, and C, is shown in the bar chart. Branch C shows the largest overall sales compared to others, outperforming them significantly. While

Branch C leads by a significant margin, Branches A and B show sales figures that are comparatively close. Among the most important findings is the significant discrepancy in sales between Branch C and the other two branches, indicating possible causes affecting sales output. These could be related to a number of factors, including customer demographics, product assortment, store size, location, and marketing tactics.
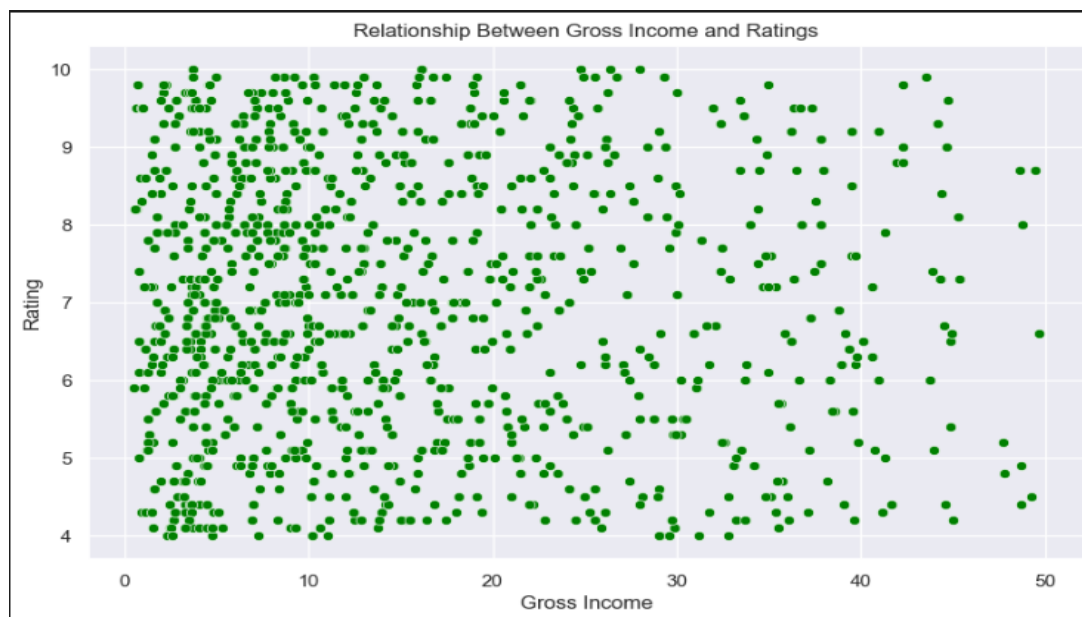

The association between gross income and customer ratings is shown graphically in the scatter plot. A data point is represented by each dot, which shows the gross income (x-axis) and rating (y-axis) of a particular client. The scatter plot's widely dispersed data points, which imply a poor association between gross income and customer ratings, are a striking feature. Stated differently, it appears that there is no discernible trend suggesting that customers with higher incomes consistently provide better or lower evaluations. This lack of a clear pattern suggests that income level in this specific dataset has little bearing on customer happiness as indicated by ratings. Additionally, a clustering of data points within the rating range of 6 to 8 is revealed by the plot, suggesting a general degree of consumer satisfaction. It's difficult to make firm judgments about the general attitude of customers, though, without knowing more details about the rating system and the precise evaluation standards.

Profitability Across Branches
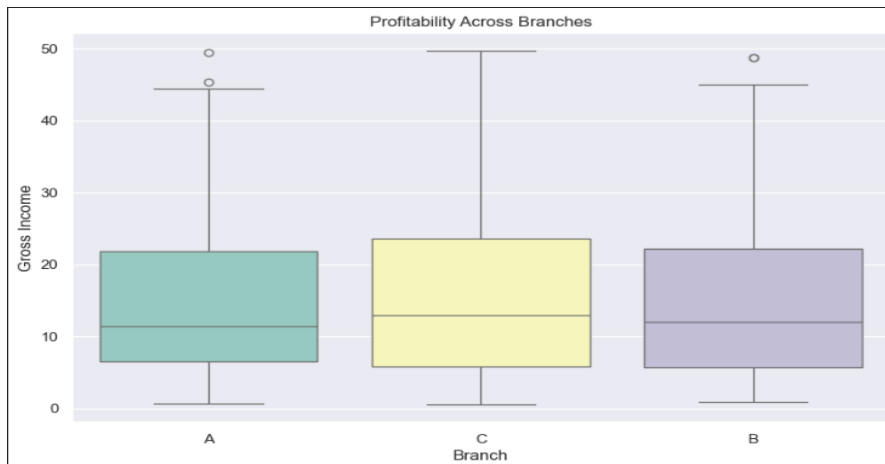
The box plot shows that the three branches' profitability differs significantly from one another. With a larger range of values and the highest median gross revenue, Branch C exhibits higher performance variability. Branch A appears to be more consistent generally, but its median income is lower and its interquartile range is smaller, indicating potentially poorer overall profitability. The performance of Branch B lies in the middle, with a median more in line with Branch A and a larger range more in line with Branch C.

Relationship Between Gender and Gross Income

The box plot shows a potential disparity in gross income between genders. The distribution of values differs between the two genders despite the similarity in median income. The distribution of women's earning levels is wider, with a few exceptions at the top end. Men's income distribution, on the other hand, is more uniform and contains fewer outliers. This suggests that while the average income may remain unchanged, there may be greater income variety among women.

Average Gross Income by Month (Jan-Mar 2019)
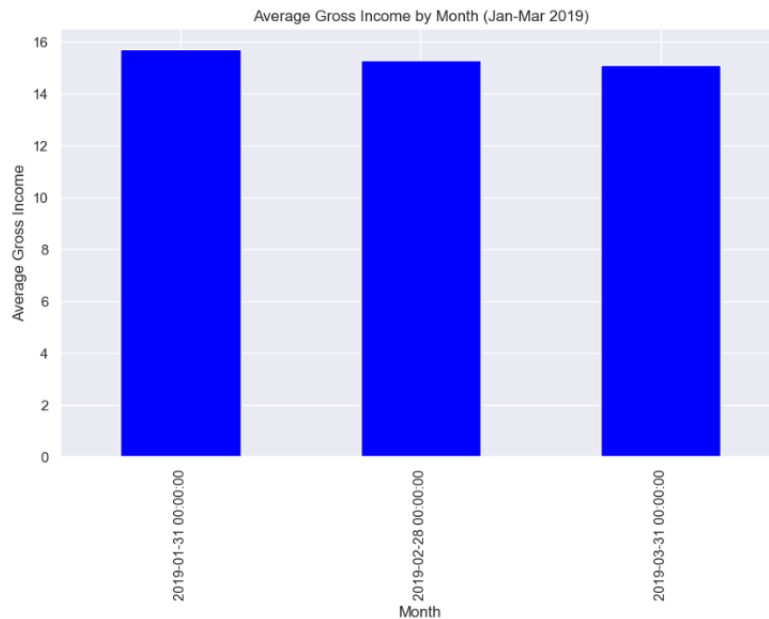
The average gross income for each month between January and March 2019 is displayed in a bar chart. The three-month average gross income shows some minor variations, but overall, it stays quite constant, indicating steady sales during this time.


Average Gross Income by Day (Jan-Mar 2019)
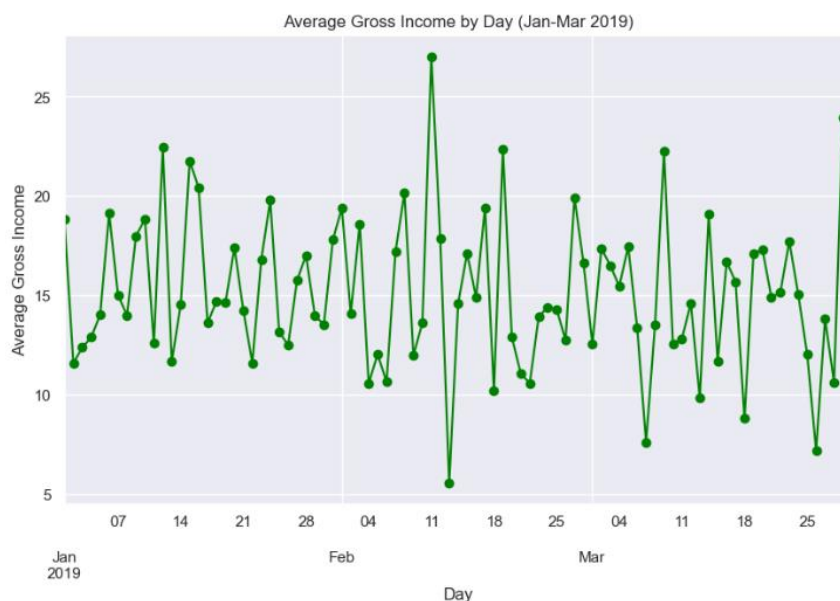
From January to March 2019, the average gross income fluctuated daily, as seen by the line chart, which shows notable daily variability with peaks and troughs throughout the course of the time. Although there is a small upward trend near the end of March, the pattern remains inconsistent overall. This implies that changes in daily income were probably caused by variables like the day of the week, promotions, holidays, weather, or unanticipated events.

Most Profitable Product Line
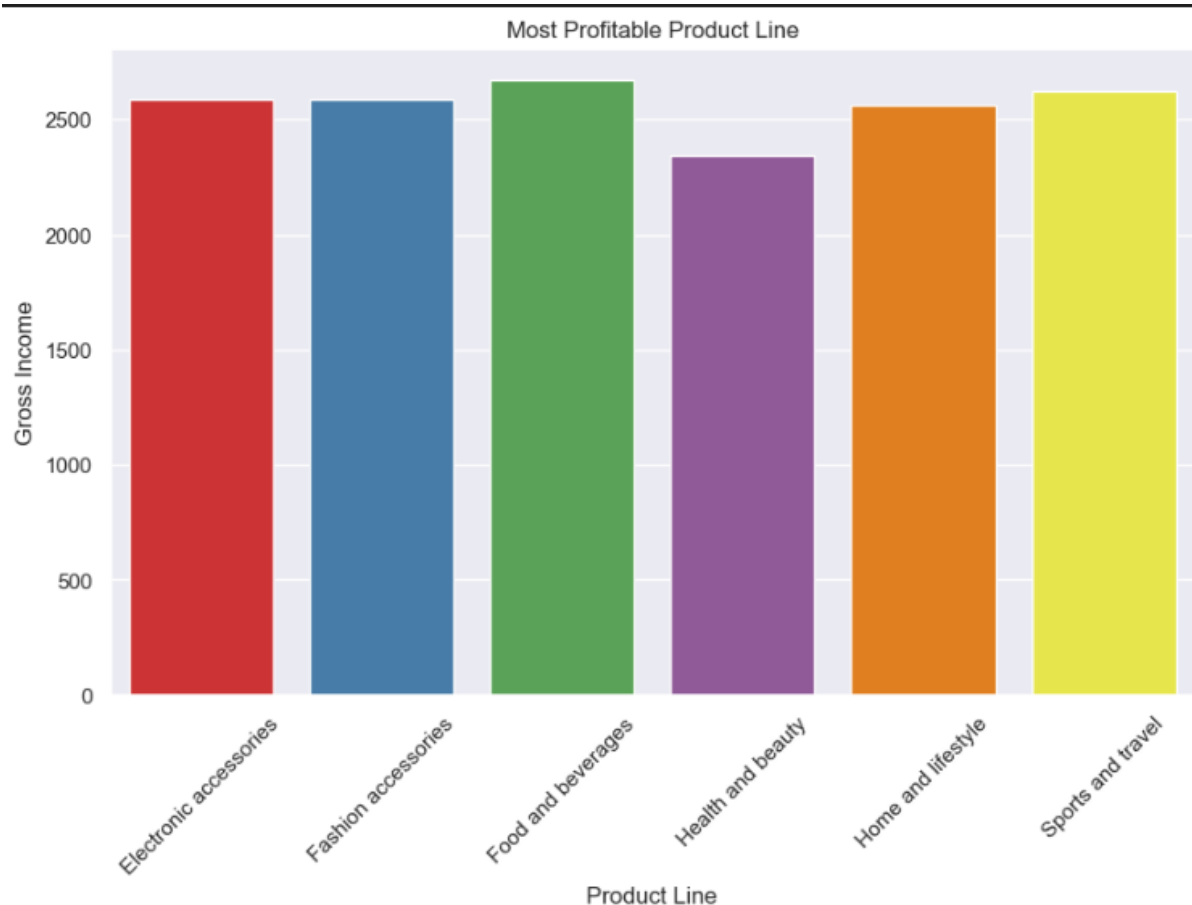
The gross income produced by the various product lines at the supermarket is depicted in the bar chart. The most lucrative category is found to be electronic accessories, closely followed by fashion accessories. The total revenue is also greatly influenced by food and drink. Products in the health and beauty industry show a moderate level of profitability. Products for the home and lifestyle have a similar pattern, whereas the least profitable categories include sports and travel gear. These insights provide important information for maximizing product assortment and marketing tactics by highlighting the disparate revenue contributions of various product categories.

Spending Pattern by Gender
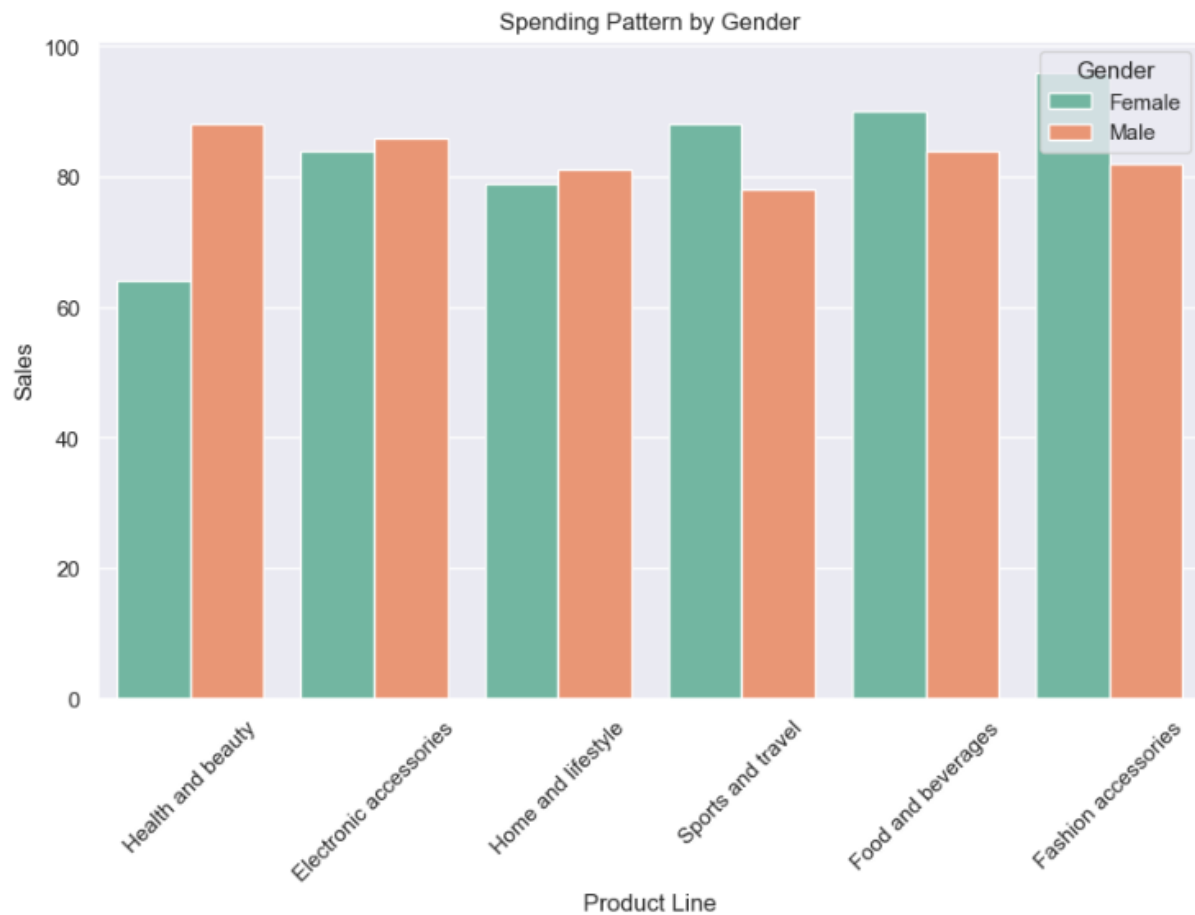
The bar chart illustrates distinct spending patterns between male and female customers across various product lines. Women exhibit higher spending on Fashion accessories, while men tend to spend more on health and beauty products. Both genders demonstrate comparable spending on electronic accessories, home and lifestyle products. Sports and travel accessories appear to be less popular among men as compared to women. This visualization underscores the importance of understanding gender-based preferences for effective targeted marketing and product assortment strategies.

Number of Sales by Day of the Week
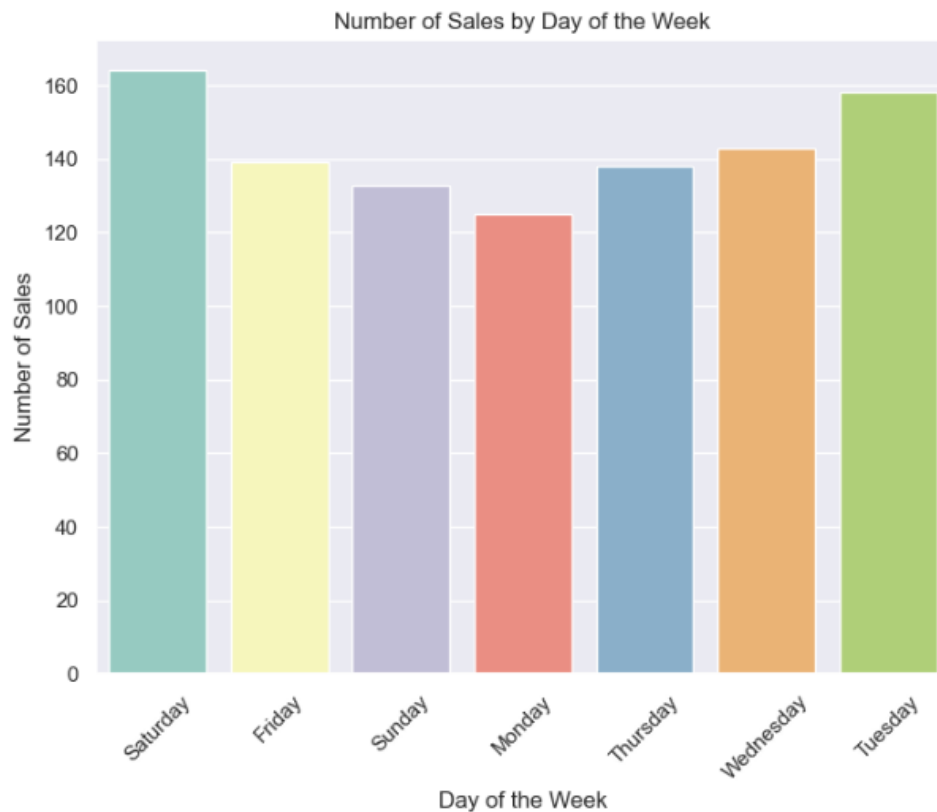
The number of sales for each day of the week is displayed in a bar chart. Sales peak on Saturday, with Tuesday coming in close second. Wednesday and Friday show a sizable volume of transactions as well. On Sunday and Monday, sales are at their lowest. This pattern indicates that client purchases are more active on weekends and during the middle of the week than they are at the start of the week. Staffing plans, inventory control, and marketing tactics can all benefit from these data to improve customer satisfaction and store operations.

Number of Sales by Hour of the Day
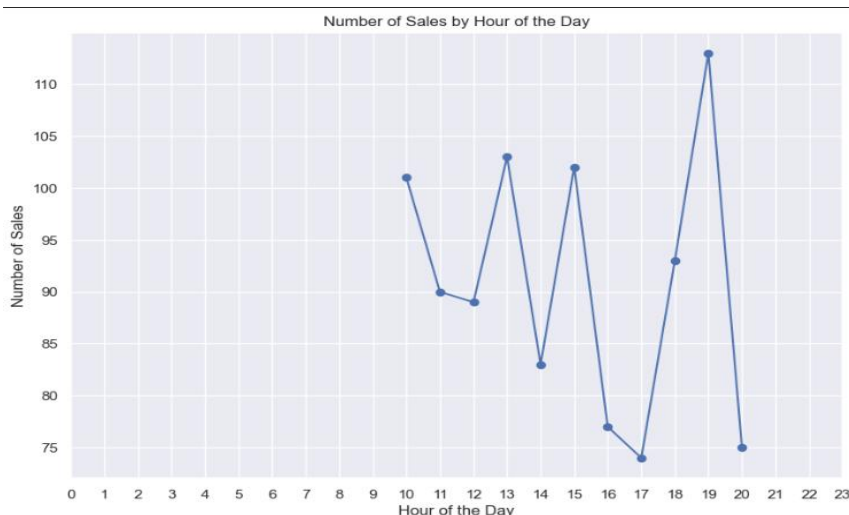
The line graph shows notable variations in the quantity of sales at various times of the day. Peak sales take place between 6 PM and 8 PM, with a significant uptick at 7 PM. There is hardly much sales activity in the early morning and late at night. This pattern indicates that there may be a secondary peak in the evening, with client traffic concentrating during particular daytime periods. In order to maximize customer service and overall retail efficiency, staffing plans, store operations, and promotional methods can all be informed by an understanding of these hourly trends.
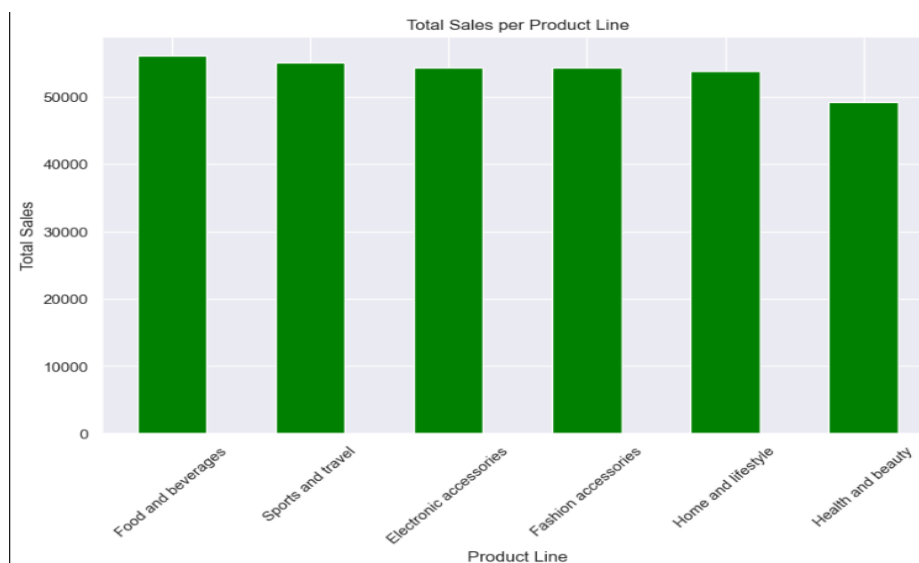
The total sales produced by each of the supermarket's product lines are displayed in a bar chart. Although there are some variances, the general sales numbers, which fall between 50,000 and 60,000 units, seem to be quite stable across various product lines. This indicates that all product categories are performing fairly, with no line having a sales monopoly or performing noticeably worse than the others.

Total Income by City and Product Line
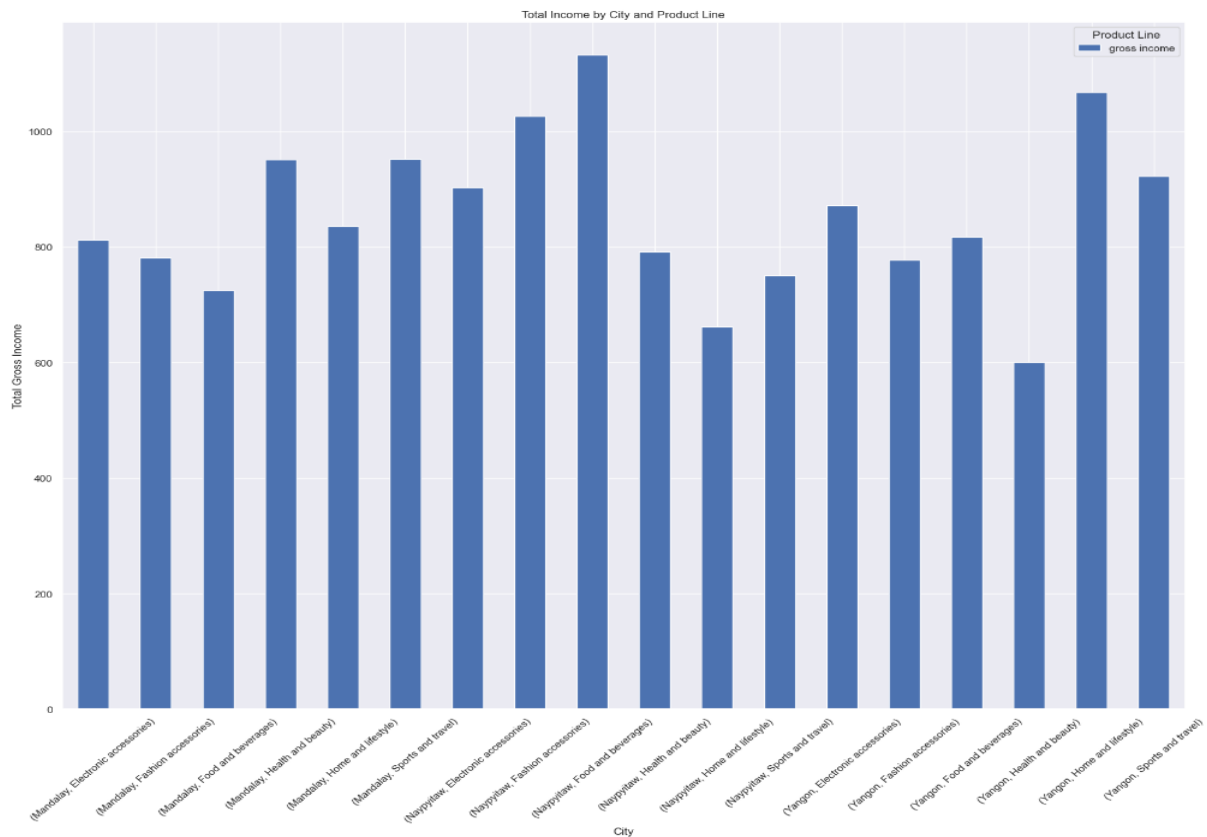
The bar plot makes it abundantly evident that Naypyitaw routinely outperforms Mandalay and Yangon as the major source of revenue and sales across a range of product lines. Sales of health and beauty products are highest in Mandalay, while those of home and lifestyle products are more in demand in Yangon. A strategic expansion plan is advised based on these observations, with a concentration on Naypyitaw and Yangon because of their significant contributions to overall income. Fashion accessories and health and beauty products are in high demand in Naypyitaw, making them wise investment locations. Given Mandalay's impressive achievements in the fields of health and beauty, it could be advantageous to address the declining demand for electronic accessories. Focusing on fashion accessories and home and lifestyle goods is in line with Yangon's sales trends. In line with market expectations, fashion accessories also show promise as an investment for Naypyitaw and Yangon. Food and beverage products, sports goods, and travel products are in great demand in Naypyitaw, which suggests more room for tailored advertising and product offerings. All things considered, business success is likely to be achieved by implementing a well-rounded and diverse plan that is customized to the particular market demands of each city.

CONCLUSION:

The examination of the 1,000-customer supermarket sales dataset yields a number of significant findings that can guide tactical choices. Men and women make up the same portion of the client base, and both genders contribute equally to the sales numbers. With 34.5% of users, e-wallets are the most popular payment method. Credit cards (31.1%) and cash (34.4%) are next in line. When it comes to e-wallets, male customers like them more than female customers (16%) by a margin of 18.5%. With 17.8% of sales, fashion accessories are the most popular product category. Food and drink come in second with 17.4%, and technological accessories comes in third with 17%. Products related to sports and travel, health and lifestyle, and health and beauty come next, with the last category selling the fewest—at 15.2%. Yangon's supermarket has the largest gross income, followed by Mandalay and Naypyitaw. This is because Mandalay is a major regional hub for trade and logistics, and Yangon is a major hub for trade.

A significant correlation has been observed between the kind of client and gross income; members contribute marginally more (50.1%) than regular customers (49.9%). Users of credit cards and e-wallets, whether regular or members, make a substantial contribution to the overall revenue, even more so than do cash users. There are slight disparities in the accessories for home and lifestyle and electronics, but male customers are more likely to purchase health and beauty products than female customers. Female customers are more likely to purchase food and beverages, sports goods, and fashion accessories. It's interesting to see that Branch B has the highest rating of 10.0, while Branch A has the lowest average rating of 4.0 while having the largest sales. This suggests that greater customer satisfaction is not always correlated with larger sales volume. To improve overall business performance and customer pleasure, these insights can direct tactics for targeted marketing, inventory management, and customer service.