WALMART SALES DATA ANALYSIS

BY

SUDEEP KUMAR DAS

DT- 17TH APRIL 2024

Table of Contents

List of Figures

Introduction

Walmart, a global retail juggernaut with 10,586 stores in 24 countries, reigns supreme as the world's largest company by revenue. In February 2023, it reported a staggering $611.3 billion in total revenue for FY2023, solidifying its economic dominance. Yet, beyond its financial might, Walmart boasts the distinction of being the largest private employer worldwide, with a workforce of 2.2 million.

We have been provided with sales data of 3 new branches. The sales team wants the data to be evaluated as well as find out useful insights that will help them in driving sales and improving Branch performance. As a Sales Analyst, it is your job to analyze the data and generate actionable insights that will help the sales team to make informed decisions.

Problem Statement

The sales team wants to know the performance of 3 of our Branches that were recently opened and have been operational for the last 3 months. The sales team has provided us with the sales data for the last 3 months. As a sales analyst, we are expected to analyze the data and generate actionable insights that will help the sales team in making informed decisions and improve sales.

Technology Used

For the analysis, we will be using Python programming language for visualization and conducting univariate and multivariate analysis to draw insights based on the graphs.

Let's Start!

Let's look at what the data looks like. Using the 'head()' function we will wind out the first few rows of the dataset. This is what the data looks like.

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05	13:08:00	Ewallet	522.83	4.761905	26.1415	9.1
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.2200	2019-03-08	10:29:00	Cash	76.40	4.761905	3.8200	9.6
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03	13:23:00	Credit card	324.31	4.761905	16.2155	7.4
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.0480	2019-01-27	20:33:00	Ewallet	465.76	4.761905	23.2880	8.4
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08	10:37:00	Ewallet	604.17	4.761905	30.2085	5.3

Figure 1: First Few Rows of the dataset

The above figure shows that we have successfully loaded the dataset and are ready to conduct our analysis. From the first few rows, we see that the dataset contains the sales details of the store and it consists of some generic information about the stores, some customer information, and some product information. Let's find out some more information about the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Invoice ID              1000 non-null   object
 1   Branch                  1000 non-null   object
 2   City                    1000 non-null   object
 3   Customer type           1000 non-null   object
 4   Gender                  1000 non-null   object
 5   Product line            1000 non-null   object
 6   Unit price              1000 non-null   float64
 7   Quantity                1000 non-null   int64
 8   Tax 5%                  1000 non-null   float64
 9   Total                   1000 non-null   float64
 10  Date                    1000 non-null   object
 11  Time                    1000 non-null   object
 12  Payment                 1000 non-null   object
 13  cogs                    1000 non-null   float64
 14  gross margin percentage 1000 non-null   float64
 15  gross income            1000 non-null   float64
 16  Rating                  1000 non-null   float64
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```

Figure 2: Info of the Dataset

The above figure shows us the info of the dataset. The info tells us a bit more about the dataset. The dataset consists of 1000 rows or records of sales transactions that took place in the stores and 17 columns that consist of some generic information about the stores, some customer information, and some product information. The info also tells us that there are no missing data in the dataset which is a good indication of clean data. We can also tell about the datatype of each of the columns. From an initial look, we can say that the datatypes are accurate. However, we need to change some of the datatype specifically that of the Date and Time column which is of object datatype into Datetime datatype for our ease of analysis. This step will be done at a later stage when we will conduct our feature engineering.

Another step that we will conduct at this point is that we will drop the 'Invoice ID' column as it is non-essential for our analysis.

In the nest step we will conduct a preliminary analysis of the dataset to find out some more information about our dataset. This will be done using the 'describe' function which will give us some more information. The following is the output.

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Invoice ID	1000	1000	602-80-9671	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Branch	1000	3	A	340	NaN	NaN	NaN	NaN	NaN	NaN	NaN
City	1000	3	Yangon	340	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Customer type	1000	2	Member	501	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Gender	1000	2	Female	501	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Product line	1000	6	Fashion accessories	178	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Unit price	1000	NaN	NaN	NaN	55.6721	26.4946	10.08	32.875	55.23	77.935	99.96
Quantity	1000	NaN	NaN	NaN	5.51	2.92343	1	3	5	8	10
Tax 5%	1000	NaN	NaN	NaN	15.3794	11.7088	0.5085	5.92488	12.088	22.4452	49.65
Total	1000	NaN	NaN	NaN	322.967	245.885	10.6785	124.422	253.848	471.35	1042.65
Date	1000	89	2019-02-07	20	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Time	1000	506	14:42:00	7	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Payment	1000	3	Ewallet	345	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cogs	1000	NaN	NaN	NaN	307.587	234.177	10.17	118.497	241.76	448.905	993
gross margin percentage	1000	NaN	NaN	NaN	4.7619	6.22036e-14	4.7619	4.7619	4.7619	4.7619	4.7619
gross income	1000	NaN	NaN	NaN	15.3794	11.7088	0.5085	5.92488	12.088	22.4452	49.65
Rating	1000	NaN	NaN	NaN	6.9727	1.71858	4	5.5	7	8.5	10

Figure 3: Description of the Dataset

After performing this step, we have found out a lot about the dataset. The following are the points that can be concluded at this point.

Initial Insights

1. The dataset consists of 1000 rows and 17 columns. The dataset is arranged concerning the invoice ID of the sales.
2. Since the column of invoice ID is not necessary for our analysis, it is dropped.
3. All the columns appear to be of the correct data type except the date and time column which is in object datatype and moving forward feature engineering needs to be applied to obtain additional insights.
4. There are no null values or duplicate values in the dataset so there is no need for null value treatment.
5. The highest invoice-generating branch is Branch A which generated 340 invoices followed by Branch B generating 332 invoices and then Branch C which generated 328 invoices.
6. Those customers who are enrolled in any form of loyalty program with the store have generated more invoices i.e. 501 and non-members have generated 499 invoices. This needs to be further analyzed in terms of value.

7. The product line which was most bought was Fashion Accessories followed by Food and Beverages and Electronic accessories.
8. The most preferred mode of payment is Ewallet followed by Cash and Credit card.
9. The average unit price of the products is 55.6 with the minimum price being 26.49 and the maximum price being 99.96.
10. The average quantity bought by customers is 5.51 with the minimum being 1 and the maximum quantity bought being 10.

Feature Engineering

Feature engineering is necessary for our dataset especially for the column of Date and Time to extract the month name, day name, and time of day to conduct a more in-depth analysis. We would like to know whether these features in any way affect the sales. Hence, we will convert these columns from object datatype to datetime datatype and then extract the month name and day name. Also, we will convert the time column and bifurcate it into Morning, Afternoon, and Evening, If a sale was made between 5 AM to 12 AM then it is Morning, If a sale was made between 12.01 PM to 04 PM then it is Afternoon and the rest is Evening. Following is the result of feature engineering.

Year	Month	Day	Month_name	Day_name	Time_of_day
2019	1	5	January	Saturday	Afternoon
2019	3	8	March	Friday	Day
2019	3	3	March	Sunday	Afternoon
2019	1	27	January	Sunday	Evening
2019	2	8	February	Friday	Day

Figure 4: Result of Feature Engineering

After performing the feature engineering, we will drop the Month and Day column which shows us the month number and day number as we won't be needing those in our analysis. The rest of the data is useful and ready for our visualization and analysis.

Exploratory Data Analysis (Data Visualization)

1. Branch Performance Insights:

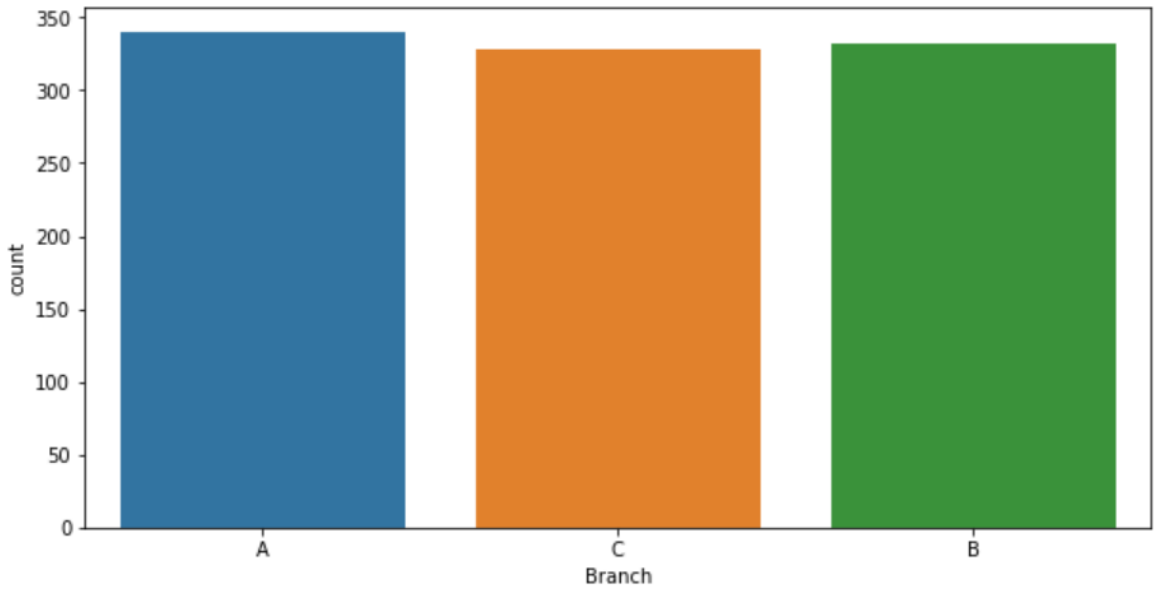

Figure 5: Distribution of Branch

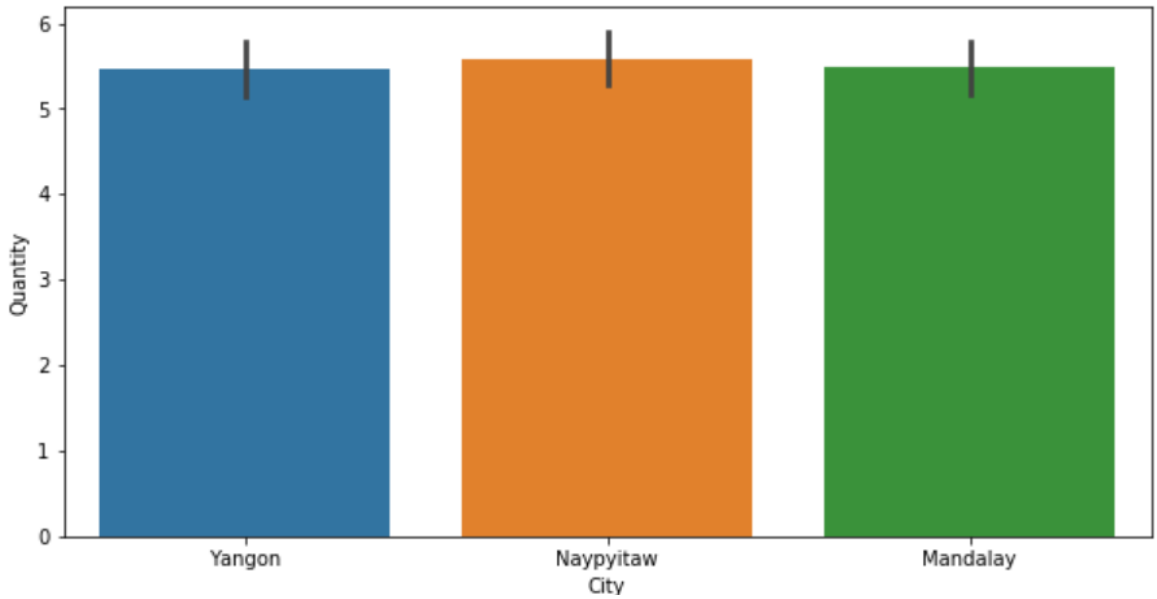

Figure 6: Quantity Sold by each Branch

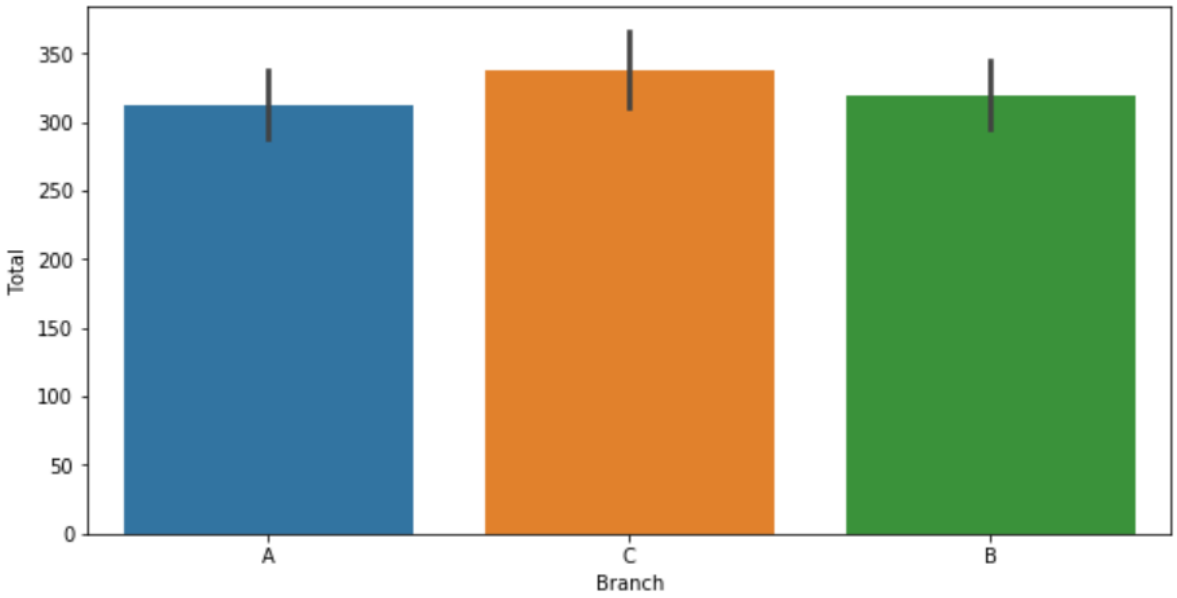

Figure 7: Total sale of each Branch

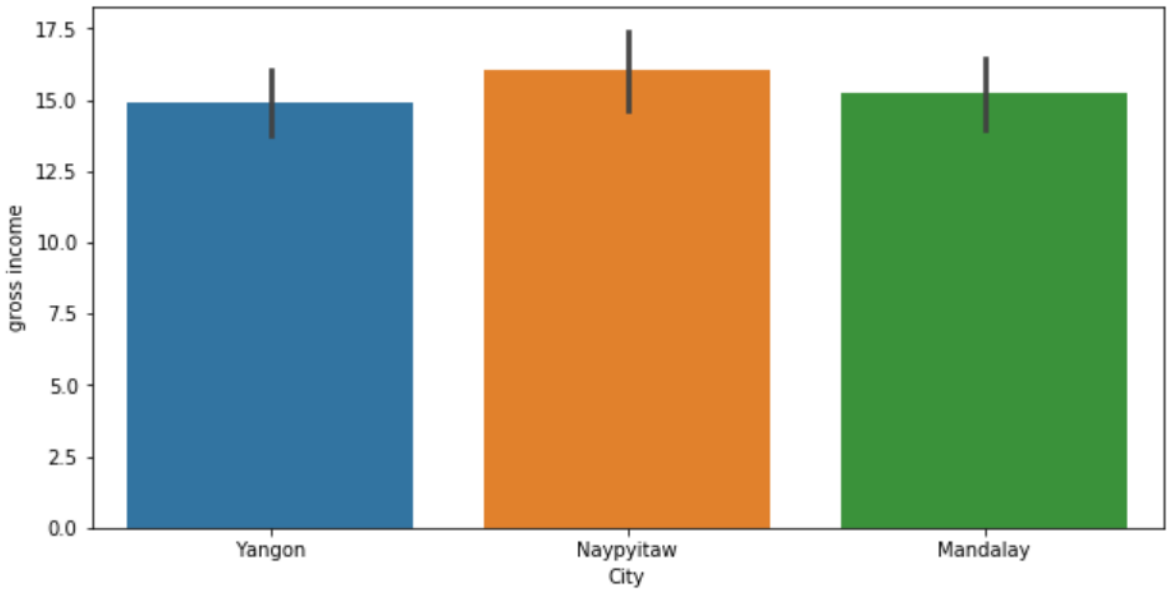

Figure 8: Gross income of each Branch

The above figure shows that Branch A located in Yangon City has generated the most number of Invoices over 3 months followed by Branch B located in Mandalay City and Branch C located in Naypyitaw City. Branch A generated a total of 340 invoices, Branch B generated 332 invoices and Branch C generated 328 invoices.

Insight of the Analysis

The analysis highlights distinct customer trends across the branches. Branch A leads in invoice generation, indicating a high volume of transactions. Meanwhile, Branch C, despite generating fewer invoices, showcases a unique purchasing pattern—customers tend to buy

larger quantities per invoice. This behavior translates into Branch C recording the highest total sales and gross income among all branches. This insightful observation underscores the significance of understanding customer preferences and behaviors. It suggests that while some branches prioritize frequent transactions, others capitalize on larger, less frequent purchases to drive revenue. These findings underscore the importance of understanding nuanced customer preferences and behaviors, offering valuable insights for tailored marketing strategies and inventory management decisions across the branches.

2. Gender Demographics and Purchasing Behavior:

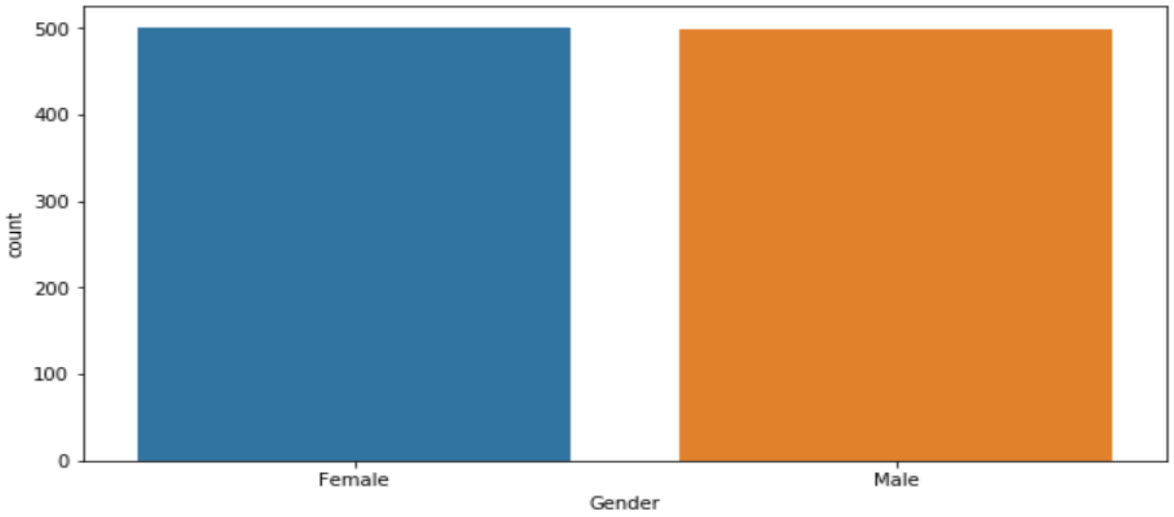

Figure 9: Gender Distribution

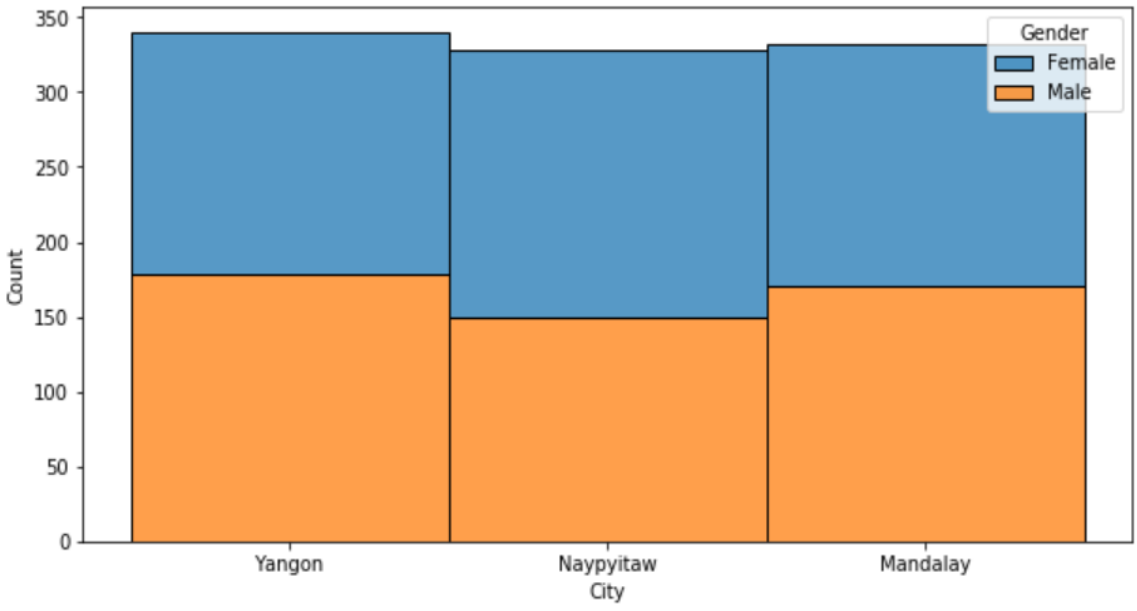

Figure 10: City-wise Gender Distribution

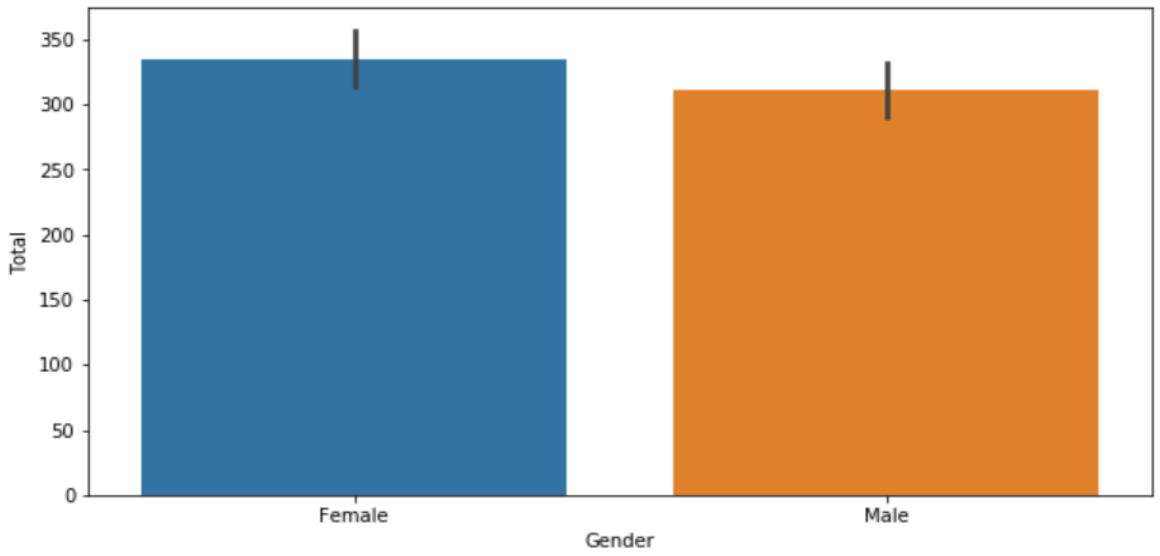

Figure 11: Total Sales generated Gender wise

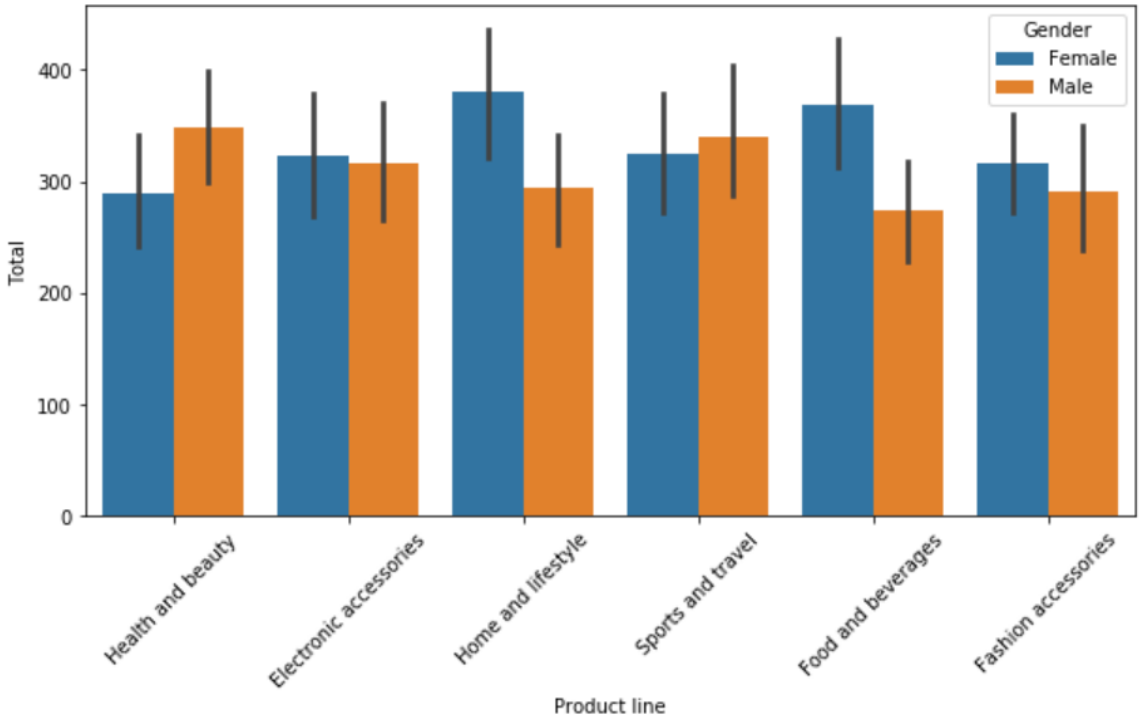

Figure 12: Product Purchase Gender-Wise

The customer base has the majority of the female population. There are 501 female customers and 499 male customers. Although the difference is not much, the analysis reveals some interesting insights in terms of the shopping behavior of the customers. The insights of the analysis are as follows.

Insight of the Analysis

The analysis reveals intriguing dynamics between gender demographics and purchasing behavior. Both male and female populations contribute almost equally to invoice generation, suggesting a balanced engagement with the retail offerings. City-wise distribution further underscores this parity in invoice contribution between genders.

However, a fascinating trend emerges when examining total sales figures. Despite generating comparable invoices, the female population significantly outpaces males in total sales and gross income contribution. This indicates that while both genders may generate similar transaction volumes, females wield greater purchasing power, driving higher revenue and profitability.

A notable deviation from conventional expectations arises in the Health & Beauty Products category. While traditionally associated with female consumers, male shoppers surprisingly dominate this sector. This intriguing insight challenges stereotypes and highlights the nuanced nature of consumer preferences.

Overall, these findings emphasize the importance of understanding gender dynamics in consumer behavior. Tailoring marketing strategies and product offerings to cater to diverse gender preferences can enhance sales performance and foster greater customer satisfaction and loyalty.

3. Loyalty Program Insights:

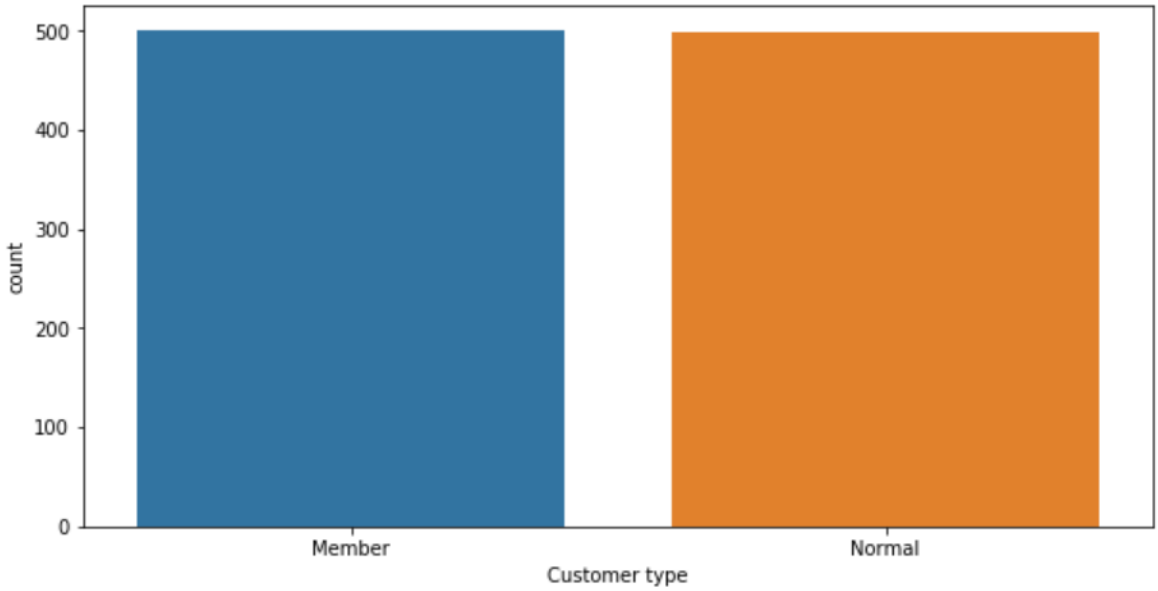

Figure 13: Customer Type Distribution

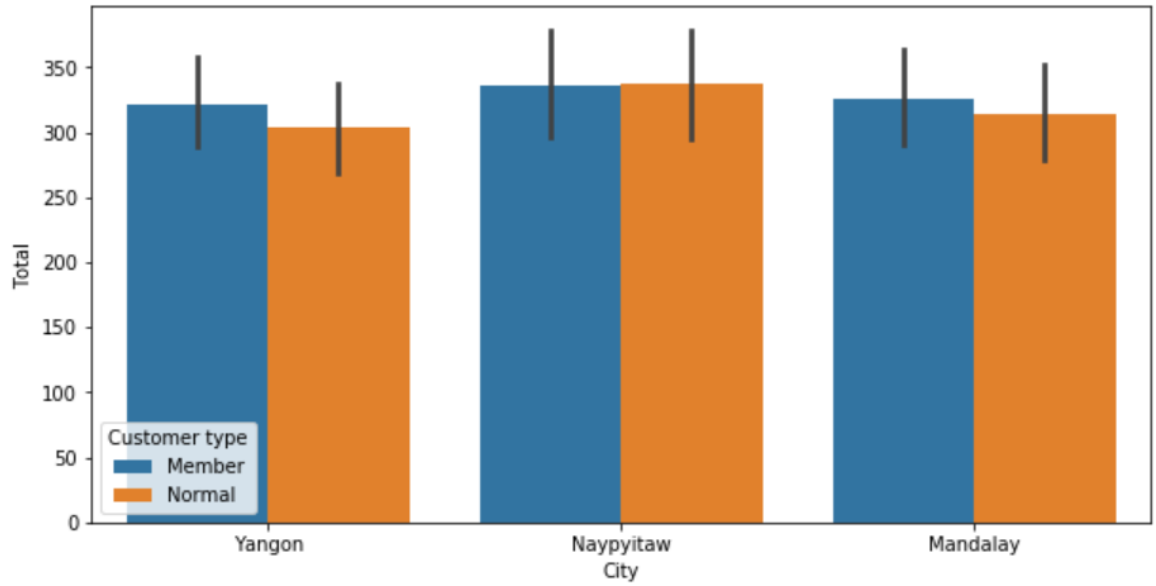

Figure 14: City-wise Customer type Distribution

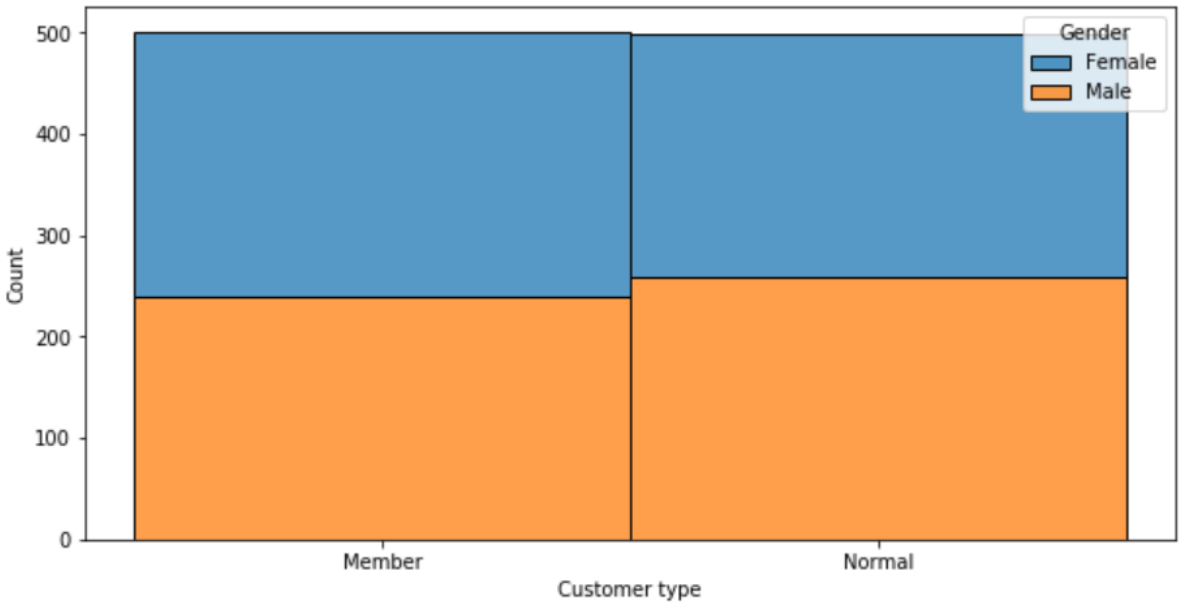

Figure 15: Gender-wise Customer Type Distribution

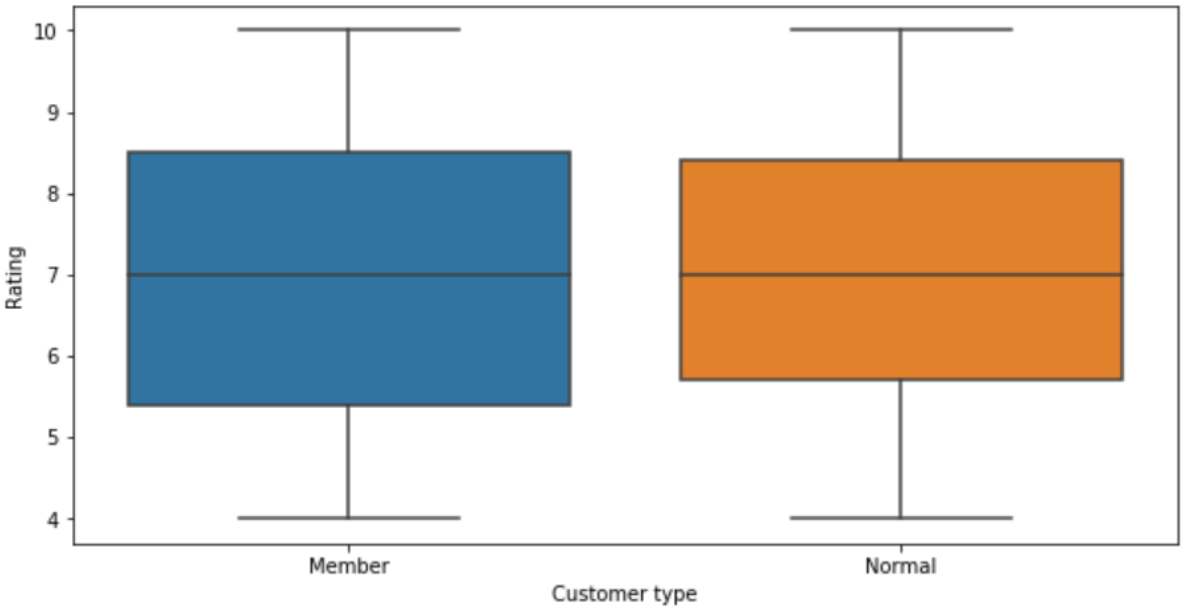

Figure 16: Rating given by Customer type

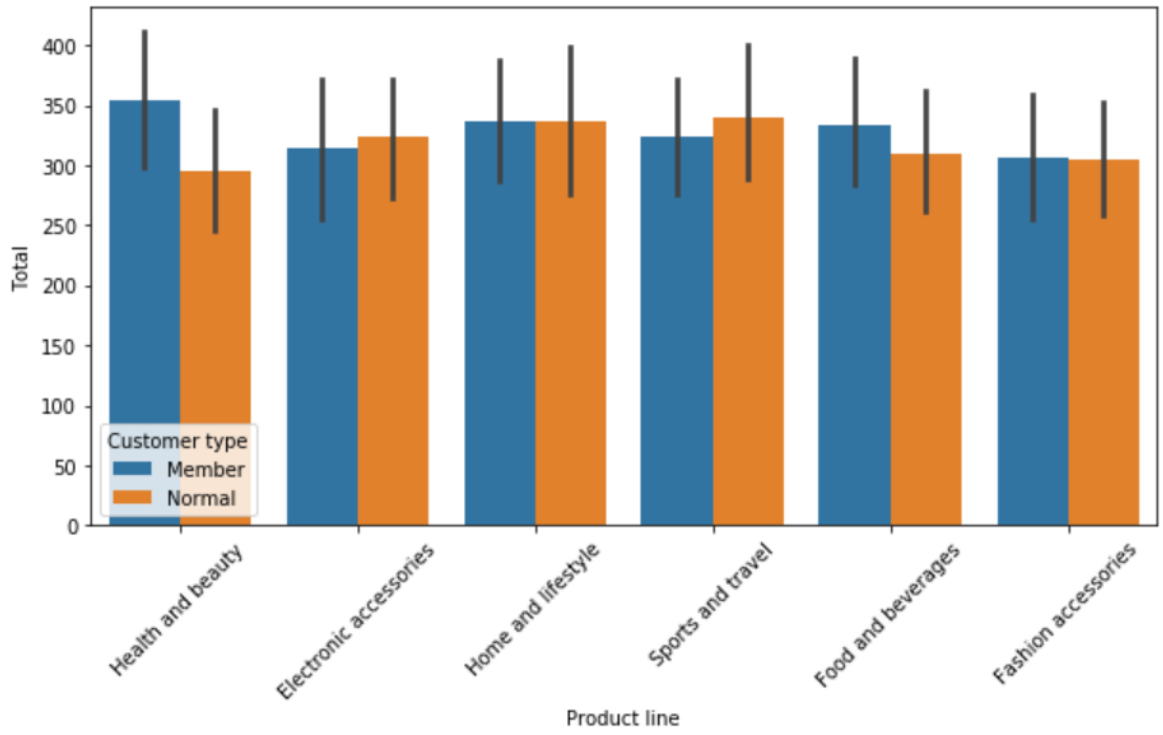

Figure 17: Product Line Purchase of Customer Type

The customer type contains two variables, Members who are enrolled in any form of loyalty program run by the store and Normal who are not part of any loyalty program. There are 501 members and 499. Although there's not much difference in numbers, the purchasing patterns of the customer type reveal some noticeable insights that are as follows.

Insights of the Analysis

Loyalty programs stand as stalwart tools in customer retention strategies, offering benefits to regular patrons while fortifying a steadfast customer base for retail establishments. In our case, with 501 out of 1000 customers enrolled in our loyalty program, and 499 yet to partake, a significant opportunity presents itself for our store's growth. Initiating a concerted effort to enroll more customers into our loyalty program is imperative.

Examining the distribution across cities reveals a compelling narrative: while Yangon City sees a dominance of program members, Naypyitaw City and Mandalay City showcase a noteworthy trend where non-members contribute substantially more to total sales. This underscores the considerable purchasing power of non-members, suggesting that a modest incentive, such as enrollment in our loyalty program, can secure their allegiance, thereby fostering a reliable customer base for the store.

Moreover, an intriguing observation emerges from customer ratings: non-members consistently rate their experiences higher than members. This favorable sentiment among non-members not only facilitates their potential enrollment in our loyalty program but also underscores an urgency to investigate the underlying causes of lower ratings from our member population. A deeper analysis is warranted to pinpoint the root of this disparity and implement remedial measures accordingly.

Delving into product line purchase patterns reveals a nuanced landscape: apart from Health and Beauty products and Food and Beverages, non-members outpace members in other categories, indicating their heightened purchasing propensity across various product segments.

In conclusion, leveraging the insights gleaned from city-wise distributions, customer ratings, and product line analyses presents a strategic opportunity to bolster our loyalty program enrollment, address member satisfaction concerns, and capitalize on the robust purchasing behaviors of our non-member clientele. Through diligent effort and targeted initiatives, we can cement our position as the preferred retail destination while nurturing enduring relationships with our valued customers.

4. Product Line Insights:

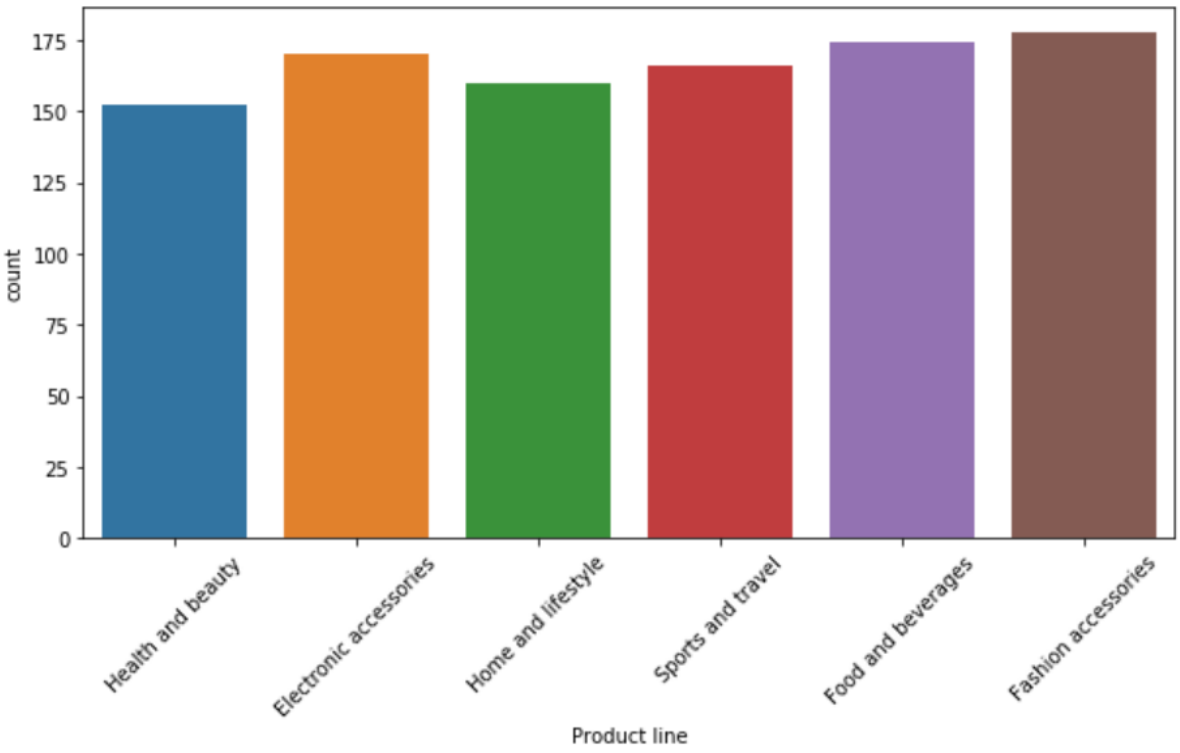

Figure 18: Product Line Distribution

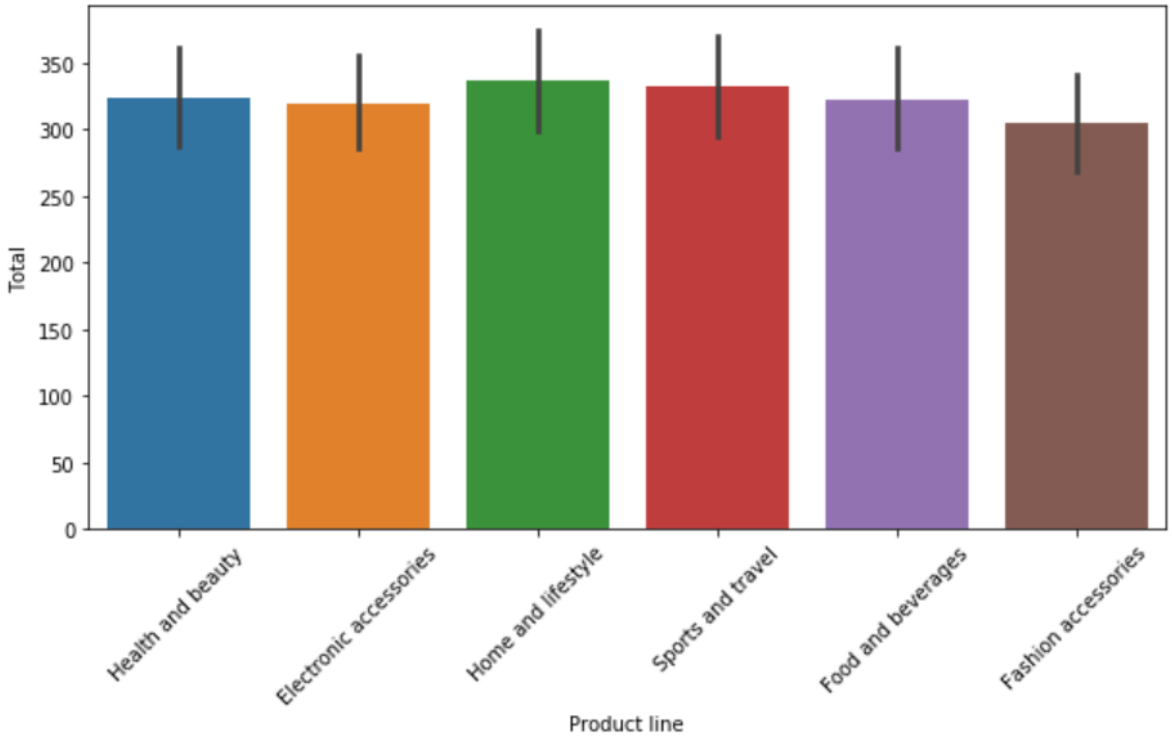

Figure 19: Product Line Contribution in Total Sales

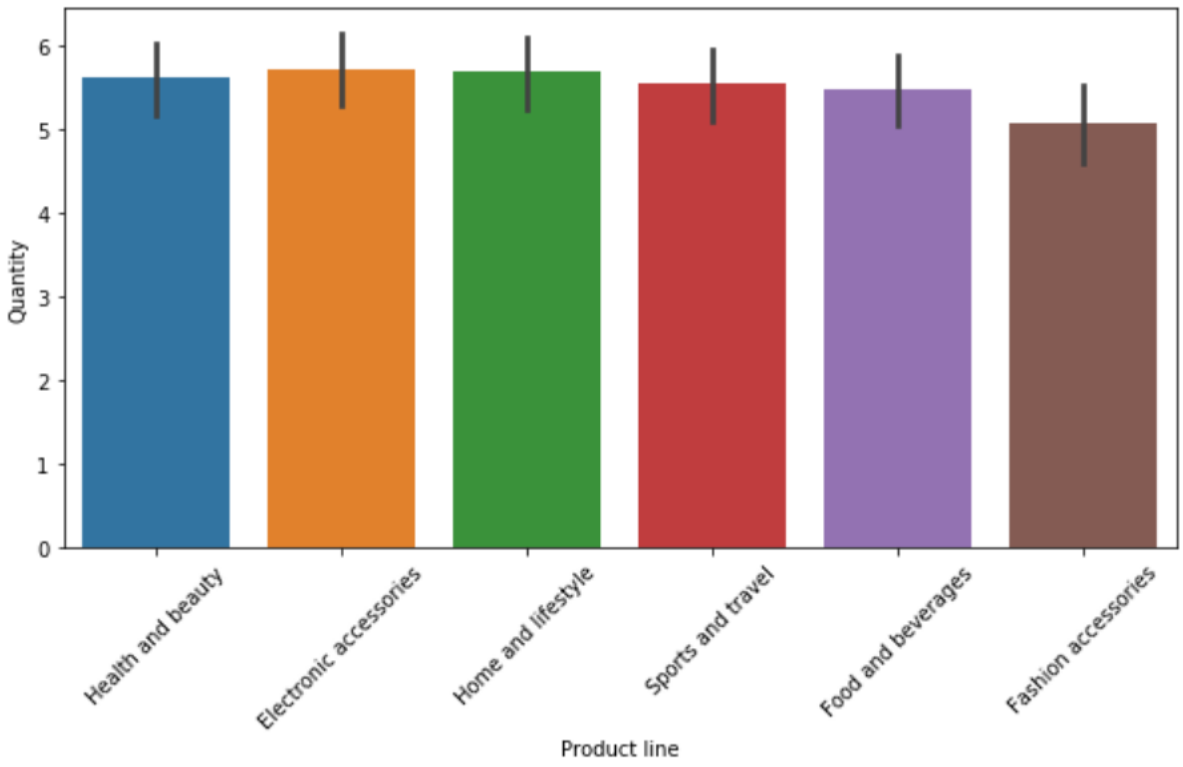

Figure 20: Product Line Contribution in Quantity

Insights of the Analysis

Among the six major product lines, Fashion Accessories emerged as the top-billed category, totaling 178 transactions, closely trailed by Food and Beverages at 174 transactions, Electronic Accessories at 170 transactions, Sports and Travel at 166 transactions, Home and Lifestyle at 160 transactions, and Health and Beauty at 152 transactions.

Interestingly, despite Fashion Accessories topping the billing charts, it contributes the least to the total sales amount and total quantities sold. This disparity underscores the category's popularity and resonance with customers, indicating a robust demand for its offerings. Given its evident appeal, exploring avenues to introduce high-value options within this category presents an opportunity to elevate total sales and enhance customer satisfaction.

5. Payment Preferences:

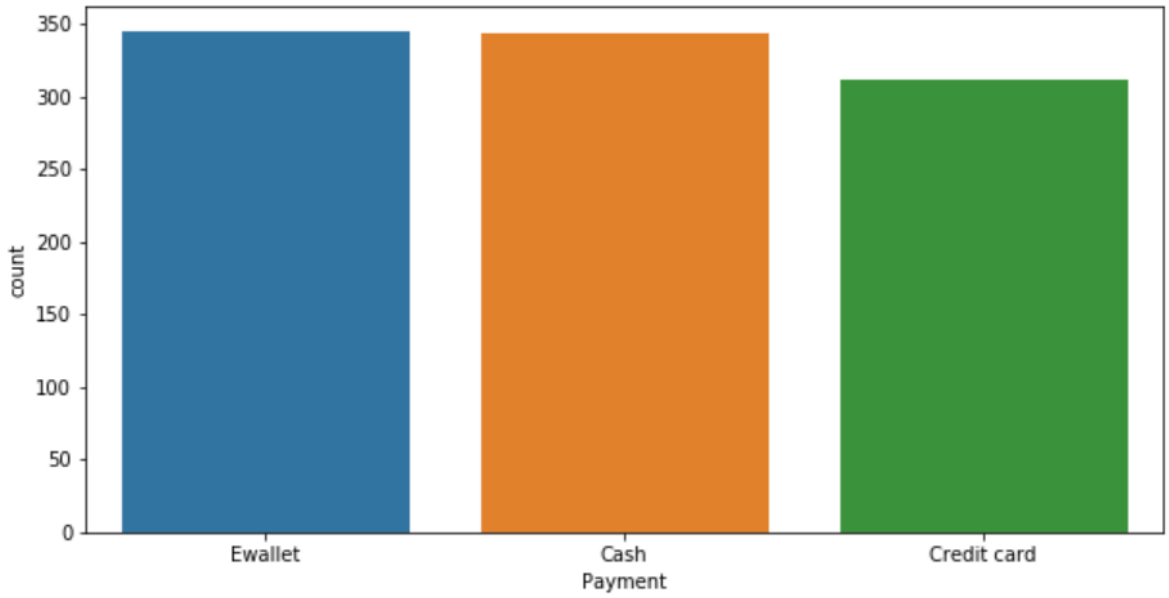

Figure 21: Distribution of Payment Method

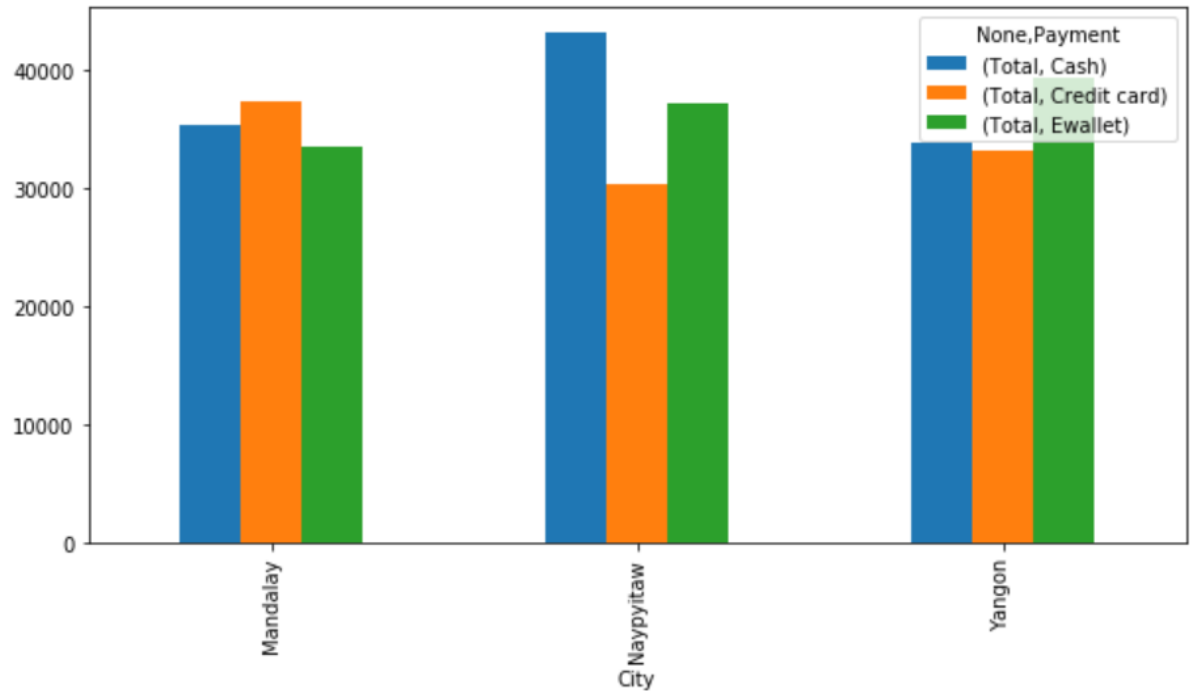

Figure 22: Total Payment City-Wise

Insights Of the Analysis

The favored mode of payment among customers emerges as E-wallet, with 345 transactions recorded, closely trailed by Cash at 344 transactions, and Credit Card at 311 transactions.

A nuanced analysis of payment preferences across cities reveals interesting insights. In Mandalay City, customers exhibit a balanced preference for payment methods. Conversely, in Naypyitaw, a clear inclination towards Cash payments is observed, followed by E-wallet, with Credit Card being the least favored option. Conversely, in Yangon City, E-wallet stands out as the most preferred payment method among customers.

These findings underscore the importance of tailoring payment options to local preferences and highlight opportunities for targeted marketing efforts to promote specific payment methods in different regions.

6. Customer Satisfaction and Ratings:

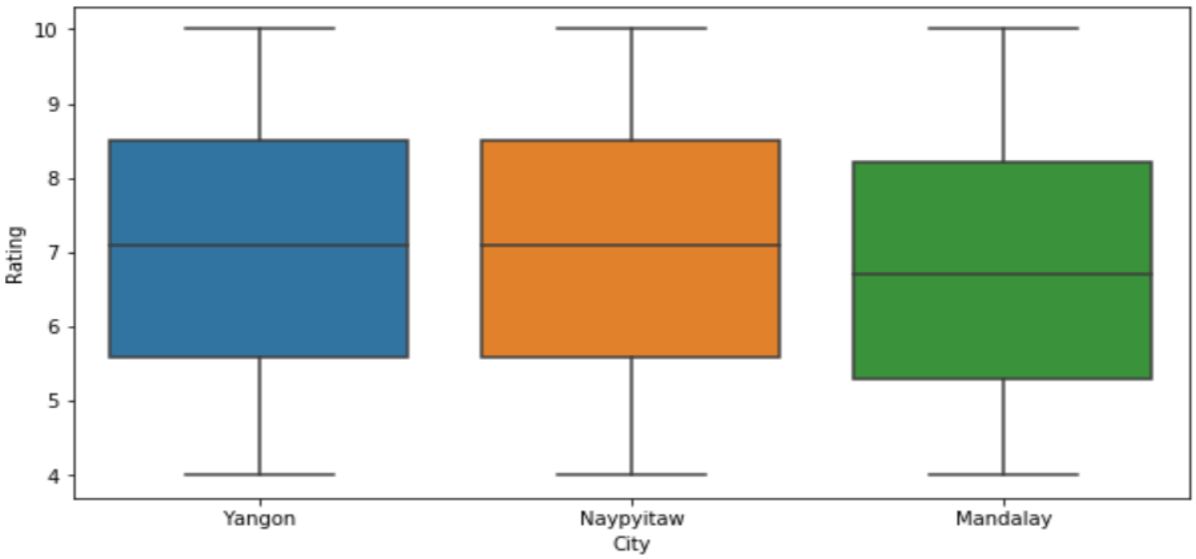

Figure 23: Rating from Cities

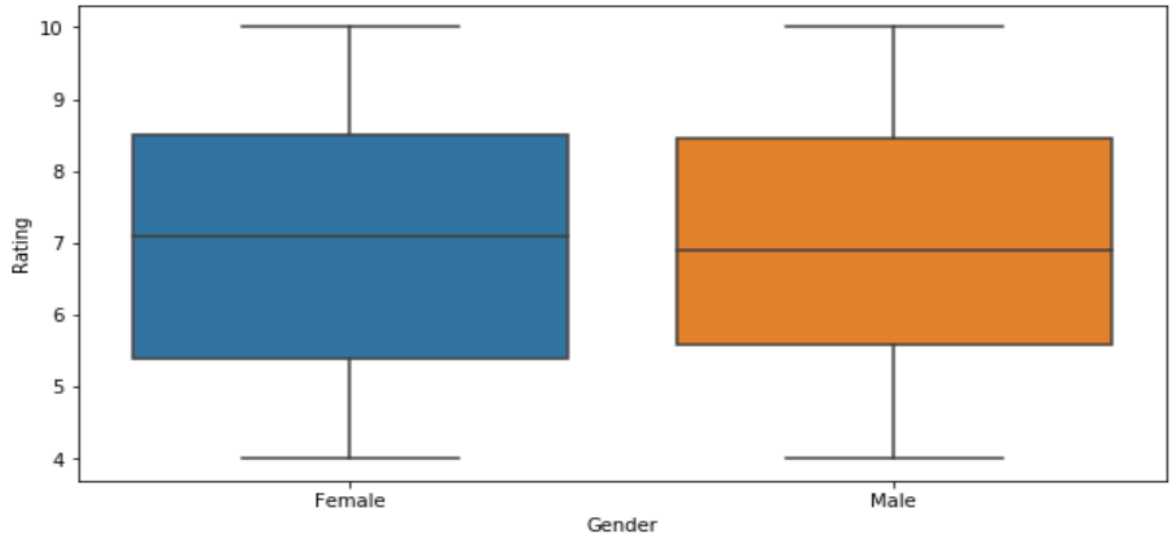

Figure 24: Rating From Gender

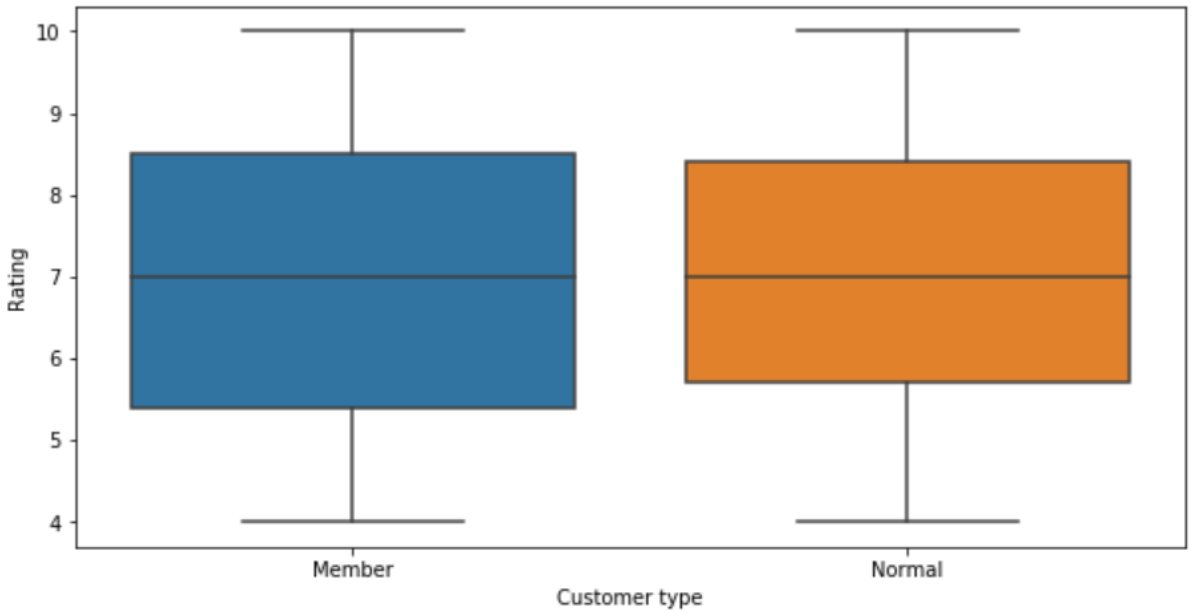

Figure 25: Customer-type Ratings

Insights of the Analysis

Ratings serve as crucial barometers for assessing store performance and uncovering actionable insights to enhance the shopping experience.

Notably, stores in Yangon City and Naypyitaw City exhibit commendable customer satisfaction, boasting average ratings above 7. However, concerns arise for Mandalay City, where average ratings dip below 7, signaling dissatisfaction among customers. Further investigation is warranted to pinpoint underlying issues and implement corrective measures to bolster customer satisfaction and retention.

Interestingly, gender-based analysis reveals a disparity in satisfaction levels, with female customers consistently awarding higher ratings compared to their male counterparts. Moreover, an intriguing finding emerges: both member and non-member customers offer similar ratings, indicating overall satisfaction among non-members. This presents an opportunity to extend loyalty program benefits to non-members, thereby solidifying their allegiance and fostering long-term customer relationships.

These insights underscore the importance of leveraging ratings data to tailor strategies aimed at enhancing customer satisfaction, fostering loyalty, and driving sustained business growth.

7. Sales Insights by Date and Time:

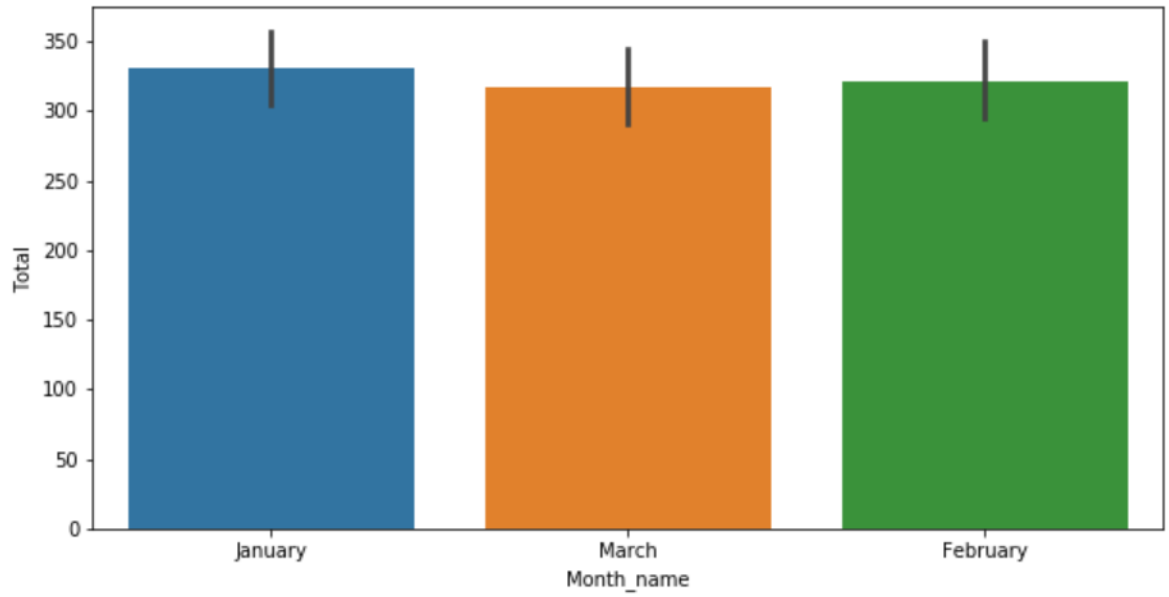

Figure 26: Month-Wise Total Sales

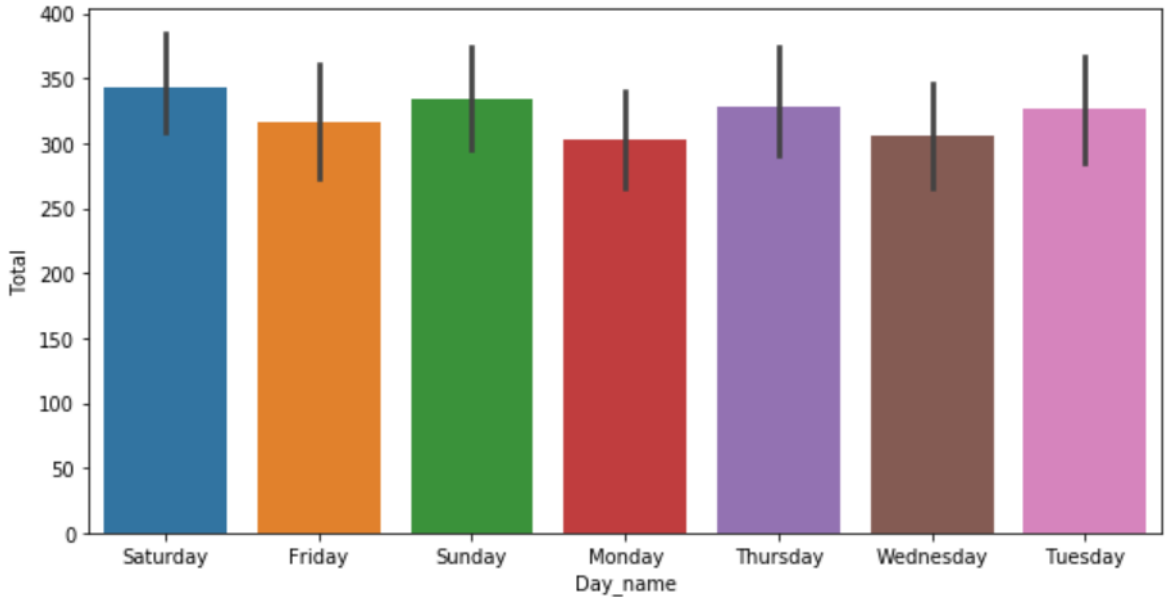

Figure 27: Day-Wise Total Sales

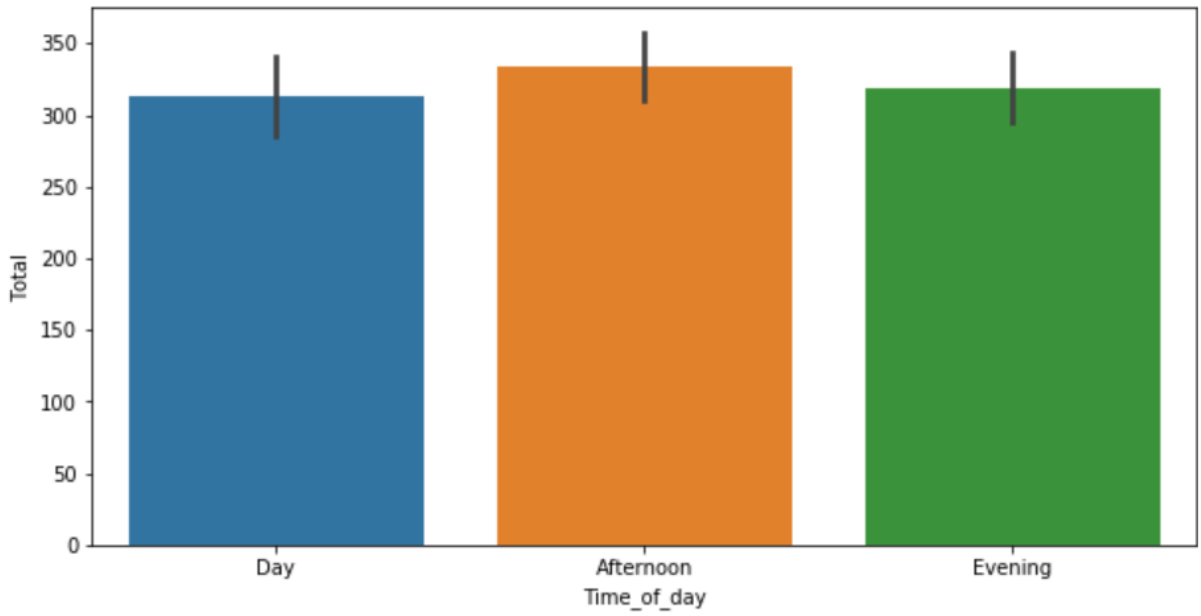

Figure 28: Time Of Day-wise Total Sales

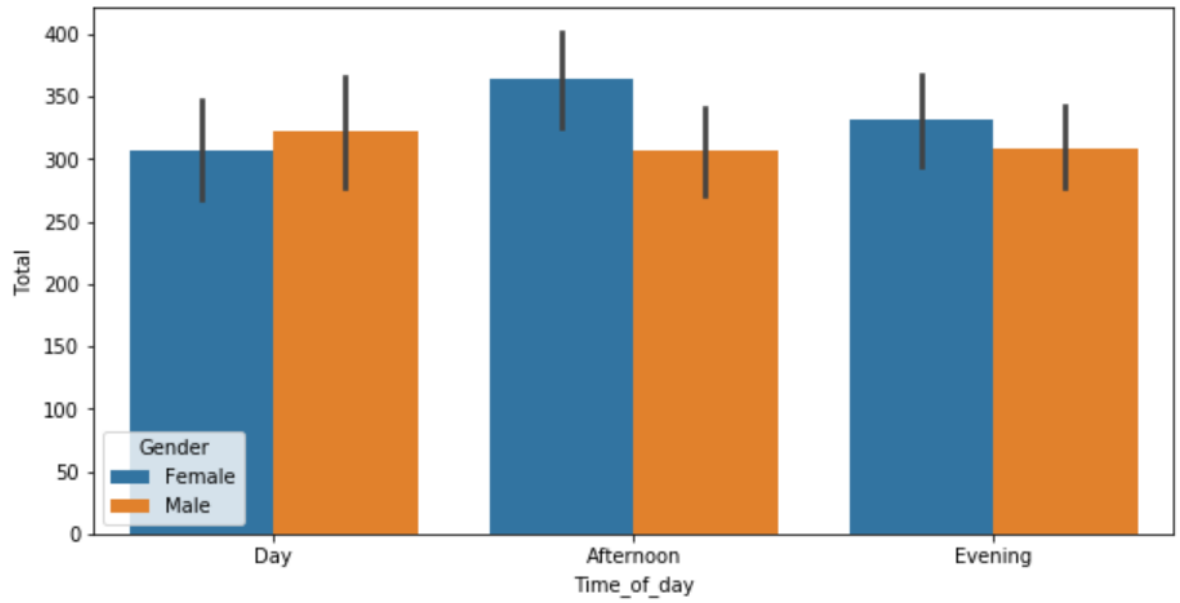

Figure 29: Time of Day Total Sales of Gender

Insights of the Analysis

A meticulous analysis of sales data, segmented by date and time, unveils compelling insights pivotal for crafting targeted promotional campaigns to drive sales.

Notably, sales surge in January, taper off in February and hit their nadir in March. This trend likely reflects heightened consumer spending during the New Year festivities, suggesting an opportune window for implementing sales-oriented promotions to revitalize sales momentum in February and March.

Furthermore, weekends emerge as prime shopping days, witnessing elevated sales presumably due to increased leisure time available to customers. Conversely, Mondays register the lowest sales volume among weekdays, signaling a potential opportunity to introduce weekday-specific promotions aimed at bolstering sales during these slower periods.

An intriguing observation unfolds when examining sales patterns by time of day: sales peak during the afternoon hours, particularly from 12 PM to 4 PM. Notably, during this timeframe, female shoppers exhibit heightened activity compared to their male counterparts. Moreover, female engagement remains pronounced during the evening hours, while mornings witness a surge in male shoppers. Leveraging these nuanced insights can inform targeted marketing strategies tailored to different demographic segments and optimize sales performance throughout the day.

By capitalizing on these insights, retailers can strategically deploy promotions, optimize resource allocation, and tailor marketing efforts to maximize sales potential and enhance overall customer satisfaction.

Conclusion

From the analysis provided, several key conclusions can be drawn:

1. **Branch Performance Insights**:
    - Branch A leads in invoice generation, indicating high transaction volumes, whereas Branch C showcases a unique purchasing pattern with larger quantities per invoice, resulting in the highest total sales and gross income among all branches.
    - Understanding customer preferences and behaviors is essential, as different branches may prioritize either frequent transactions or larger, less frequent purchases to drive revenue.
2. **Gender Demographics and Purchasing Behavior**:
    - Both male and female populations contribute equally to invoice generation across branches.
    - Despite comparable transaction volumes, females wield greater purchasing power, driving higher total sales and gross income contribution.
    - Notably, male shoppers dominate the Health & Beauty Products category, challenging conventional gender stereotypes and highlighting the nuanced nature of consumer preferences.
3. **Loyalty Program Insights**:
    - While a significant portion of customers is enrolled in the loyalty program, there is still untapped potential for growth.

- Non-members contribute substantially to total sales and have provided higher ratings, indicating their satisfaction and the need to incentivize enrollment to secure their allegiance and foster long-term customer relationships.
- The disparity in customer ratings between members and non-members warrants further investigation to enhance member satisfaction and retention.

4. **Payment Preferences**:
    - E-wallet emerges as the preferred mode of payment, with varying preferences across different cities.
    - Tailoring payment options to local preferences is crucial, presenting opportunities for targeted marketing efforts to promote specific payment methods in different regions.
    - Naypyitaw City and Mandalay City showcase a notable trend where non-members contribute substantially more to total sales, highlighting the importance of incentivizing enrollment in loyalty programs.
5. **Product Line Insights**:
    - Fashion Accessories emerge as the top-billed category, followed by Food and Beverages, Electronic Accessories, Sports and Travel, Home and Lifestyle, and Health and Beauty.
    - Despite Fashion Accessories topping the billing charts, it contributes the least to the total sales amount and total quantities sold, indicating an opportunity to introduce high-value options within this category.
    - The notable dominance of male shoppers in the Health & Beauty Products category challenges traditional gender stereotypes and underscores the importance of understanding nuanced consumer preferences.
6. **Customer Satisfaction and Ratings**:
    - Stores in Yangon City and Naypyitaw City exhibit commendable customer satisfaction, while Mandalay City faces challenges with lower average ratings.
    - Female customers consistently award higher ratings compared to males, and both member and non-member customers show similar satisfaction levels, highlighting opportunities to extend loyalty program benefits.
    - Investigating the root causes of lower ratings from member populations is crucial for enhancing overall customer satisfaction and loyalty.

7. **Sales Insights by Date and Time**:

    - Sales peak in January, decline in February, and hit their lowest in March, suggesting opportunities for targeted promotions to revitalize sales momentum.
    - Weekends witness elevated sales, while Mondays register the lowest sales volume among weekdays, indicating potential for weekday-specific promotions.
    - Afternoon hours, particularly from 12 PM to 4 PM, see the highest sales, with female shoppers exhibiting heightened activity during this time.