Supermarket sales on Power Bi

19 March 2024   23:06 PM

Python(Jupyter notebook) -will use for data cleaning and feature engineering
Power Bi - Use for Visualization

Although we can plot some graph in python as well but I will plot all graph in power bi

Data Description :

I have obtained this data from Kaggle for a Supermarket Sales. I will analyze the data and draw my conclusions.

The growth of supermarkets in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data. Predictive data analytics methods are easy to apply with this dataset.

Below are the link
https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales

Below are the dataset column with description:

1. **Invoice ID**: A unique identifier for each transaction or purchase invoice.
2. **Branch**: Represents the branch of the supermarket where the transaction took place. It could be Branch A, Branch B, etc.
3. **City**: The city where the supermarket branch is located.
4. **Customer type**: Indicates the type of customer, such as Member or Normal (non-member).
5. **Gender**: Gender of the customer making the purchase.
6. **Product line**: The category or type of product being sold, such as Electronics, Fashion Accessories, Food and Beverages, etc.
7. **Unit price**: The price of a single unit of the product.
8. **Quantity**: The number of units of the product purchased in the transaction.
9. **Tax 5%**: The amount of tax (5%) applied to the total purchase.
10. **Total**: The total amount paid by the customer for the purchase, including tax.
11. **Date**: The date of the transaction.
12. **Time**: The time of the transaction.
13. **Payment**: The method of payment used for the transaction, such as Cash, Credit Card, etc.
14. **cogs**: Cost of Goods Sold, represents the total cost of the products sold before adding tax.
15. **gross margin percentage**: The gross margin percentage, which is a measure of profitability calculated as (Total - cogs) / Total.
16. **gross income**: The gross income generated from the sale, calculated as Total - cogs.
17. **Rating**: Customer satisfaction rating given for the overall shopping experience, often on a scale of 1 to 10.

I will conduct some univariate analysis, where I will examine each column individually.

Objective:
To analyze which city is generating good sales and identify the issues faced by another store that is not performing well in terms of sales

In **Branch** Columns we have total 3 Brach (A,B,C) below are the Branch column values count and contribution :
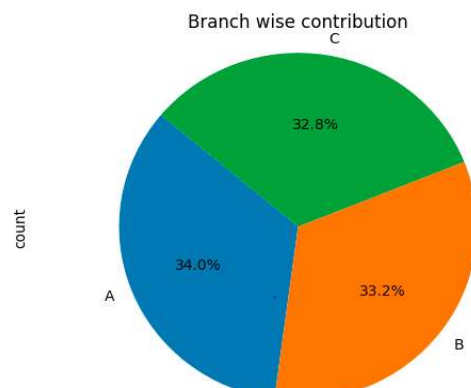
If i examine the contribution per branch, we find that Branch A has the highest sales count, but in terms of total sales amount, Branch C is leading.
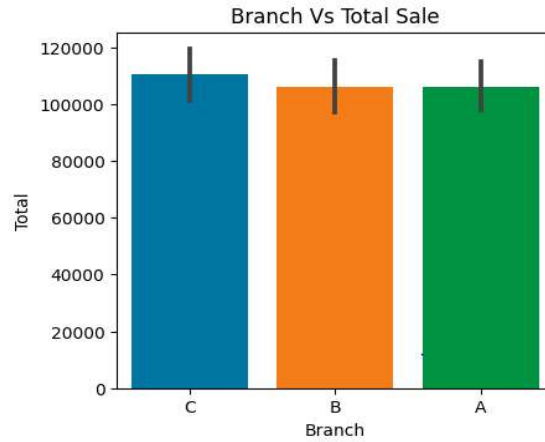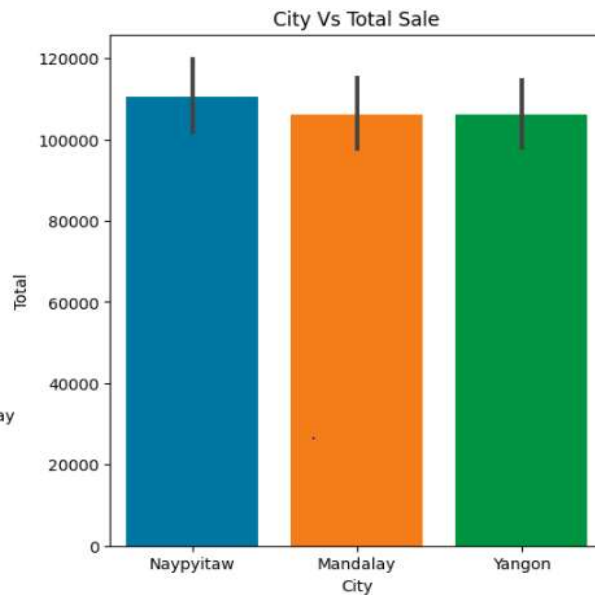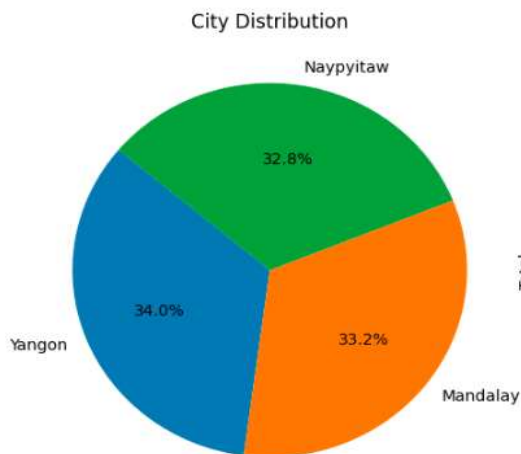
Branch , Total counts
A   340
B   332
C   328

Branch wise contribution

Need to plot Branch wise total sale:

City:

In the City column, I can see that **Yangon** has the maximum contribution compared to other cities. However, when I plot the city versus total sales, it becomes clear that **Naypyitaw** has higher sales than other cities

I can create a contingency table for these categorical columns and analyze the total sales by city and branch:
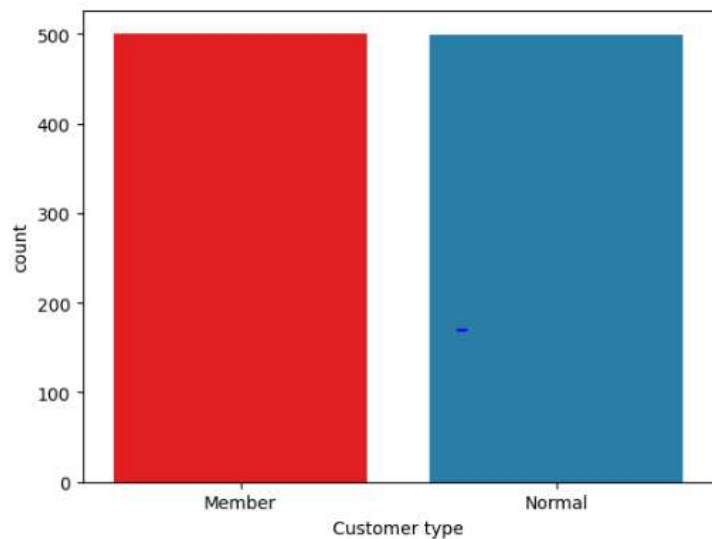
- Upon analysis, I found that Branch **'C'** and City **'Naypyitaw'** are doing maximum sales compared to other cities

City	Mandalay	Naypyitaw	Yangon
Branch			
A	NaN	NaN	32.882757
B	32.881921	NaN	NaN
C	NaN	34.335333	NaN

|				
A	NaN		NaN	32.882757
B	32.881921	NaN	·	NaN
C	NaN	34.235322		NaN

Customer Type:

We observed the customer type plot and did not find a significant variance in customer types
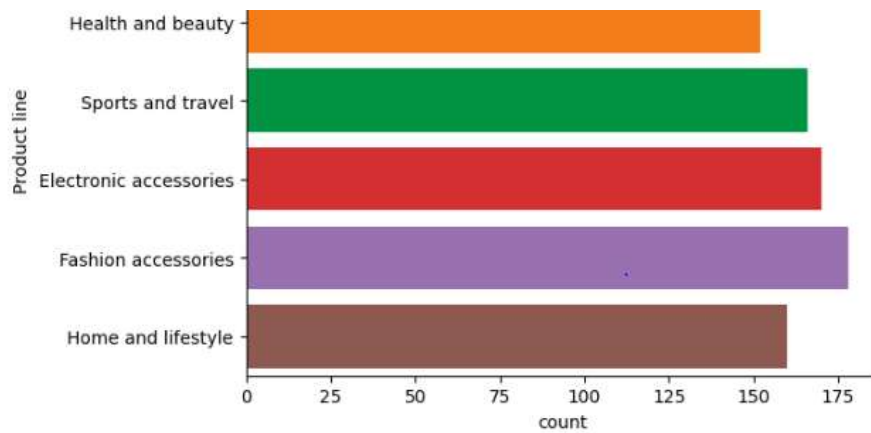

Need to plot customer type vs gander :

- In the member segments, we have a higher number of female customers compared to male customers, but in the Normal segment, we have a higher number of male customers compared to female customers
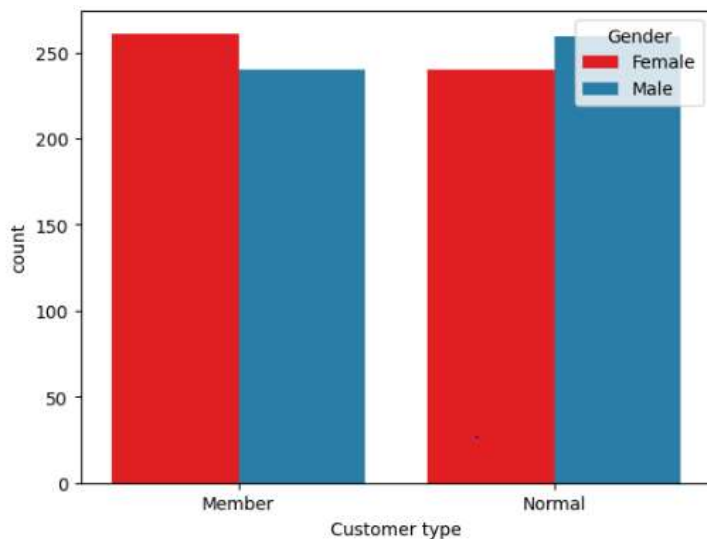
Product Line:

Fashion and food products are bought most frequently.

Unit price:

Below are some summary statistics for the **'Unit_price'** column:
The minimum unit price is 10, the mean is 55, and the maximum is 99.

| index	Unit price
count	1000.000000
mean	55.672130
std	26.494628
min	10.080000
25%	32.875000
50%	55.230000
75%	77.935000
max	99.960000

Below is a plot showing Kernel Density Estimation (KDE), where the x-axis represents 'unit_price' and the y-axis represents probability.
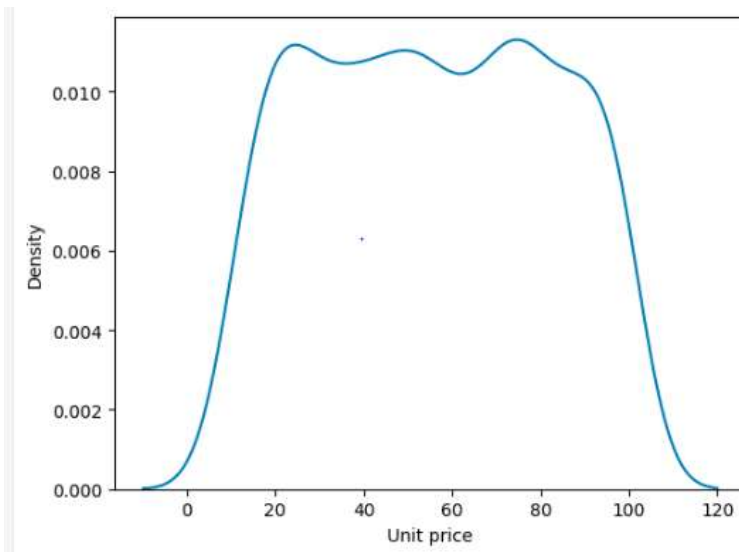
Quantity:

We can clearly see that the most frequently sold product quantity is 10. Most customers purchase 10 units most of the time
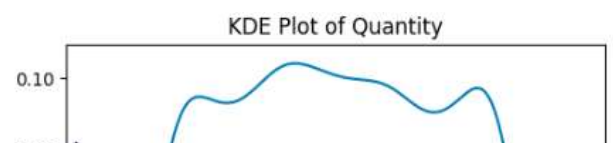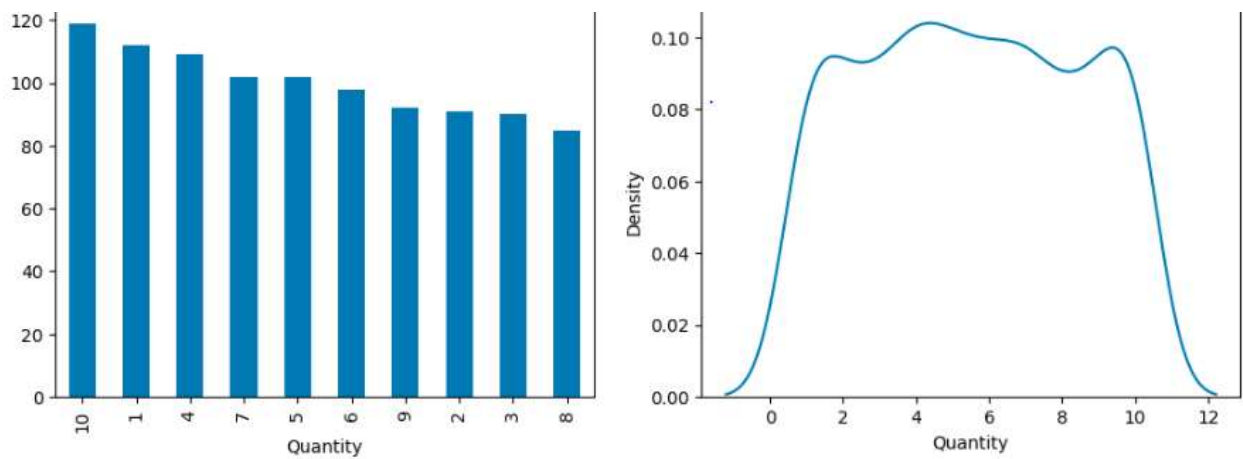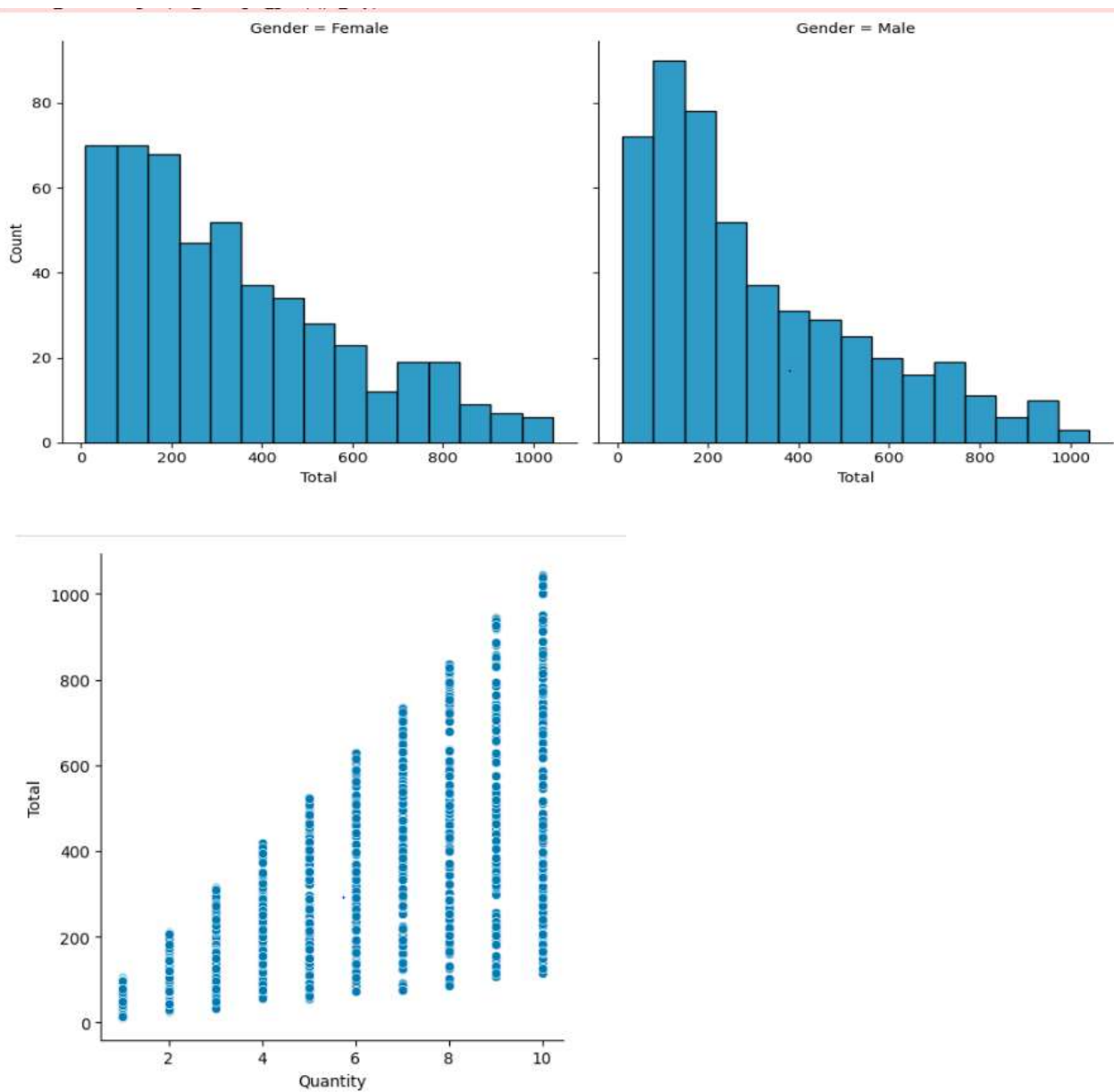
Total Amount:

We can see from the distribution graph below that when the amount is smaller, the buying quantity is high, and when the amount is high, the buying quantity is low. We can see the if the **quantity** is increasing then is **Total Amount** is also increasing . See the below **Scatterplot**
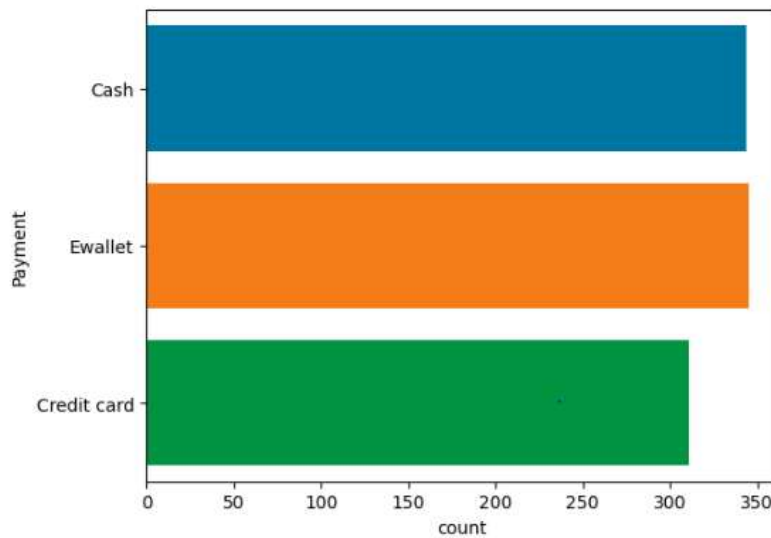
Total Amount Vs Gender

Payments:

Payments:

Most of customer are doing transaction from Cash and E- Wallet.

New col analysis:

66]:	Product line	Quantity	Total	cogs	gross income
Product line					
Electronic accessories	170	971	54337.5315	51750.03	2587.5015
Fashion accessories	178	902	54305.8950	51719.90	2585.9950
Food and beverages	174	952	56144.8440	53471.28	2673.5640
Health and beauty	152	854	49193.7390	46851.18	2342.5590
Home and lifestyle	160	911	53861.9130	51297.06	2564.8530
Sports and travel	166	920	55122.8265	52497.93	2624.8965

I have some analysis with all the columns I will write down some question then I will find some solution :

1- Is there a relationship between the time of day (morning, afternoon, evening) and the average quantity of products purchased per transaction?
2- What is the impact of the gross margin percentage on customer satisfaction rating, and is it consistent across different branches?
3- Are there any significant differences in the average unit price of products sold between different cities, considering both Member and Normal customer types?
4- Will see the MoM Total Sales graph

1st question:

There is a slight variation in the average quantity of products purchased across different time slabs, with the highest average observed in the afternoon.
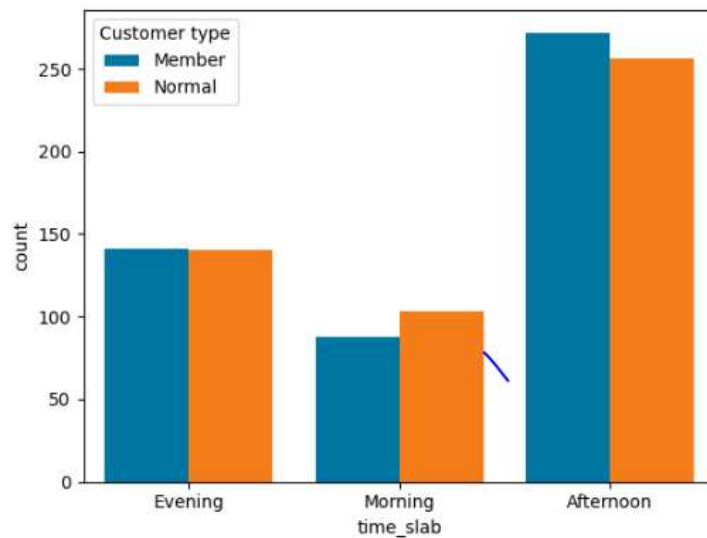
```
      time_slab  Quantity
0   Afternoon    5.579545
1     Evening    5.430605
2     Morning    5.434555
```


Average Quantity of Products Purchased per Transaction by Time of Day

We plotted a line graph where I can observe that during the **afternoon**, a higher number of customers purchase products compared to morning and evening.
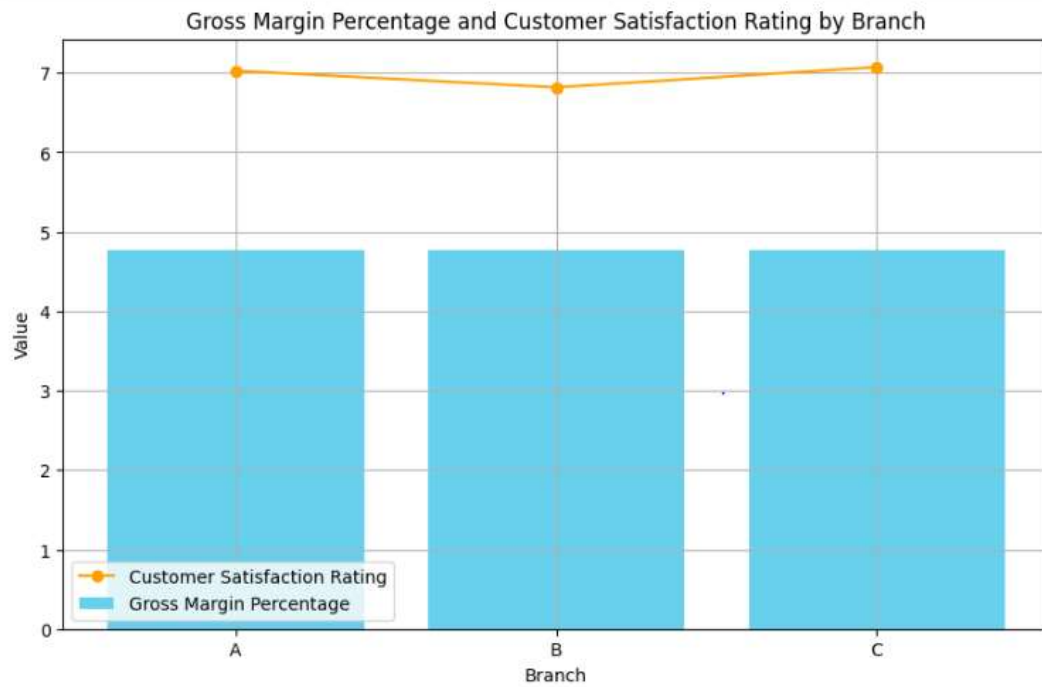

2nd question:

Observations:
- The gross margin percentage appears to be consistent across all branches, with each branch having an average of approximately 4.76%.
- There are slight differences in the average customer satisfaction ratings between branches, with Branch C having the highest rating (7.07) and Branch B having the lowest rating (6.82). However, these differences are relatively small.

	Branch	gross margin percentage	Rating
0	A	4.761905	7.027059
1	B	4.761905	6.818072
2	C	4.761905	7.072866

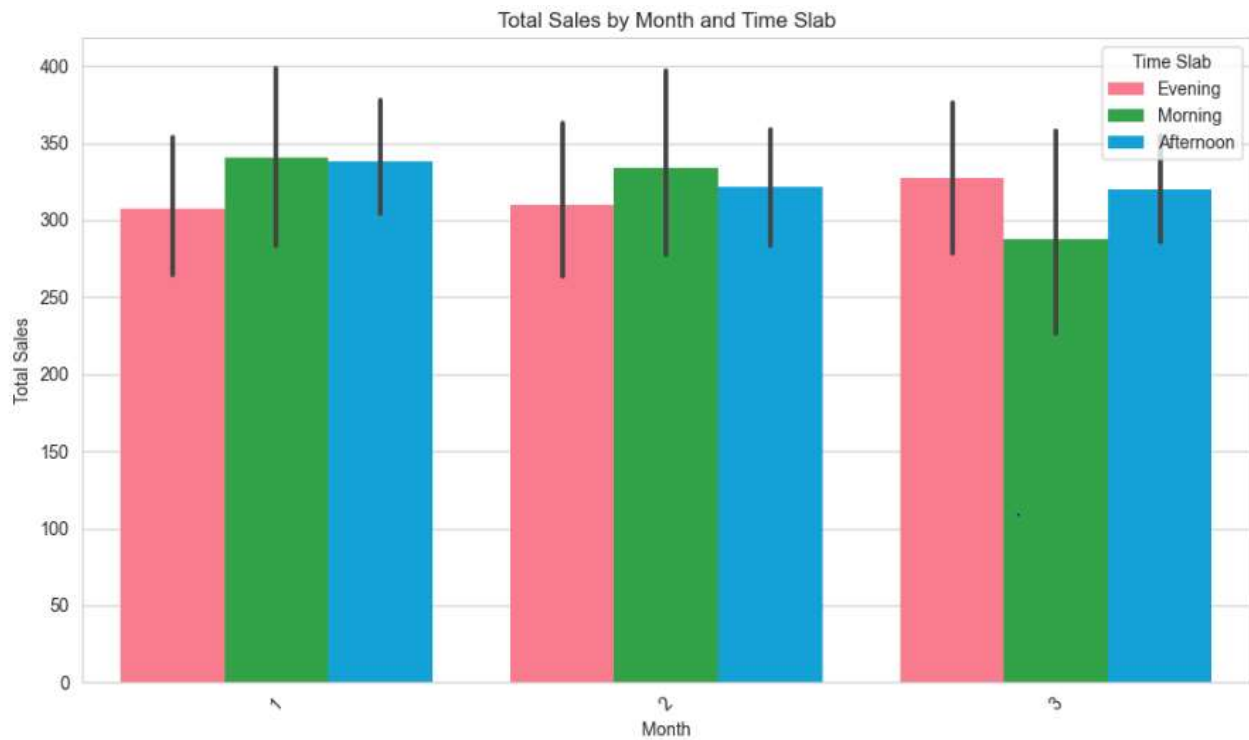## Gross Margin Percentage and Customer Satisfaction Rating by Branch

3rd Question:

- The average unit price varies across cities and customer types.
- Naypyitaw stands out with a higher average unit price for members compared to other cities, indicating potential pricing strategies or market dynamics specific to that city.
- Mandalay and Yangon show relatively consistent average unit prices between member and normal customers

Customer type	City	Member	Normal
0	Mandalay	55.349515	55.965329
1	Naypyitaw	58.849349	54.227799
2	Yangon	54.380180	55.167630

4th Question:

Total Sales by Month and Time Slab

Conclusion:

Positive side:

- Branch A has high sales count, indicating good customer traffic.
- Naypyitaw exhibits higher total sales, suggesting a potential growth market.
- Consistent gross margin percentage across branches indicates stable profitability.
- Afternoon sees high product purchase, reflecting active business hours.
- Fashion and food products sell the most, indicating popular items.

Negative side:

- Gender distribution variance in membership segments might suggest a need for targeted marketing.
- Evening and morning sales might need improvement compared to the afternoon.