Supermarket Data Analysis

Introduction :

Within this dataset lies a treasure trove of insights into the operations of a bustling supermarket. Each entry offers a glimpse into the dynamic world of retail, detailing the intricacies of every transaction. From the unique Invoice ID to the location-specific Branch and City, every aspect of the shopping experience is meticulously recorded. We delve into the diverse demographics of our customers, exploring their preferences and habits through variables such as Customer type and Gender. Product Line unveils the rich tapestry of offerings lining our shelves, while Unit Price and Quantity shed light on purchasing behaviour. Tax and Total Price illuminate the financial aspect, providing transparency into the costs involved. Date and Time capture the ebb and flow of consumer activity, painting a vivid picture of shopping patterns over time. Payment methods showcase the evolving landscape of transactions, from traditional cash exchanges to modern digital payments.

Questionnaires :

1. Which of the given cities having slab by 5% perform better than all the others?
2. Which customer gender ordered most item from all three branches?
3. Compare highest and lowest rating products on the basis of units sold.
4. Analysing unit sold and unit price in data :Answer the following sub questions:

a. What is the degree of freedom?

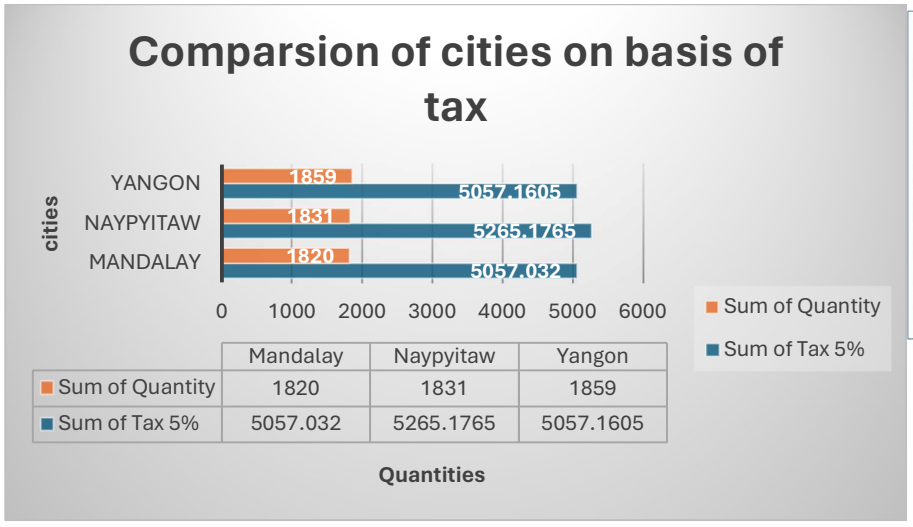
b. Corelation of unit price and revenue generator.

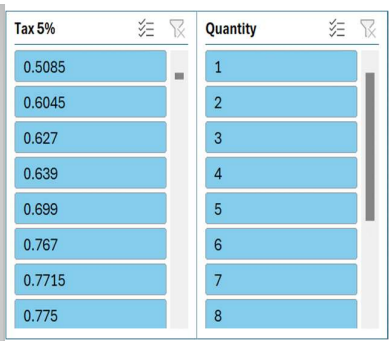
c. What result you can draw from regression of two data(-ve & +ve).

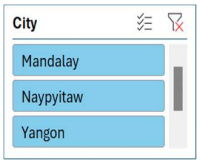
d. What product you will suggest as per the city data analysis to each type of customer?

Analytics :

Q1.Which of the given cities having slab by 5% perform better than all the others?

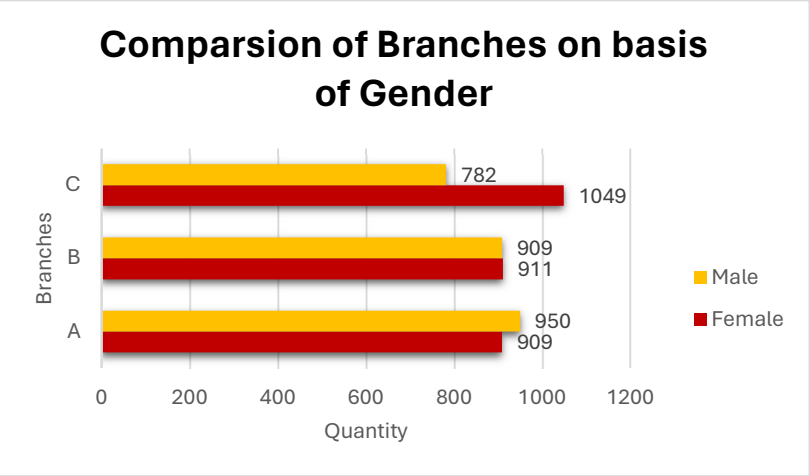






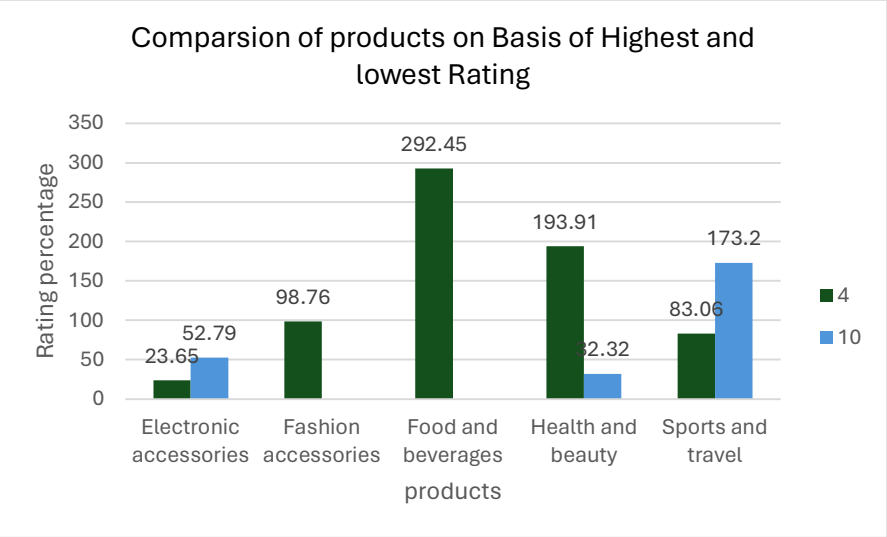
Ans: Among the cities analysed, Naypyitaw stands out as the top performer based on the 5% slab. This conclusion is drawn from a thorough assessment of the total gross income generated by each city. Naypyitaw's superior performance is evident from its ability to yield a higher total gross income compared to the other cities under consideration.

Q2. Which customer gender ordered most item from all three branches?



Ans: The data reveals that female customers have shown a consistent trend of ordering the most items across all three branches. This pattern underscores the significant role of female clientele in driving sales and contributing to overall business activity. Their preference for a wide range of products across different branches suggests a diverse set of interests and needs that the supermarket caters to effectively.

Q3. Compare highest and lowest rating products on the basis of units sold.



Ans: Comparison of highest and lowest rating products on the basis of units sold can be seen from the above chart.

Q4. Analysing unit sold and unit price in data :Answer the following sub questions:

a. What is the degree of freedom?

Regression Statistics	
Multiple R	0.010778
R Square	0.000116
Adjusted R Square	-0.00089
Standard Error	26.50636
Observations	1000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	81.45587	81.45587	0.115937	0.733555
Residual	998	701181.9	702.5871		
Total	999	701263.4			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	55.13394	1.789115	30.81632	4.6E-147	51.62308	58.6448	51.62308	58.6448
Quantity	0.097676	0.286863	0.340495	0.733555	-0.46525	0.660599	-0.46525	0.660599

Ans: The degrees of freedom for this regression analysis are:

df_regression = 1

df_residuals = 998

b. Correlation of unit price and revenue generator.

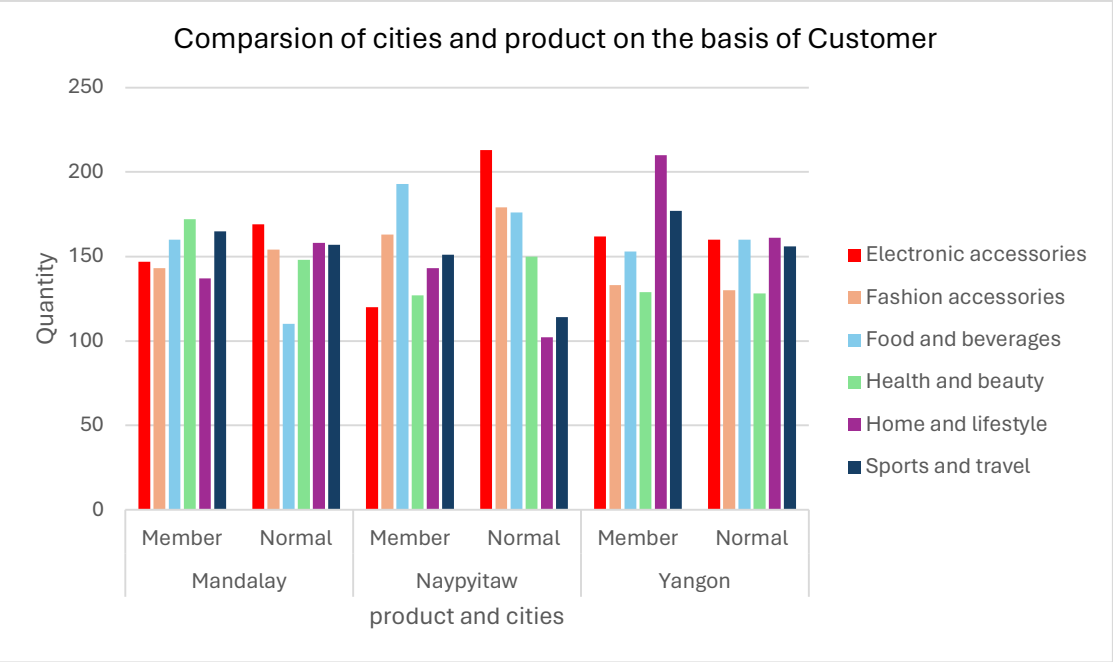
	Unit price	*Total*
Unit price	1	0.633962
Total	0.633962	1

Ans: The correlation coefficient for the relationship between "Unit price" and "Total" is 0.634. This indicates a moderately positive linear relationship between the two variables. There is a moderate positive correlation (r = 0.634) between unit price and total.

c. What result you can draw from regression of two data(-ve & +ve).

Ans: The regression analysis indicates that the Quantity variable does not have a significant impact on the dependent variable. The model fails to adequately explain the variations observed in the dependent variable based on changes in Quantity. In essence, there seems to be no meaningful association between the quantity of something (as represented by Quantity) and the outcome being measured. This suggests that other factors beyond Quantity likely influence the dependent variable, and further investigation or refinement of the model may be necessary to identify relevant predictors.

d. What product you will suggest as per the city data analysis to each type of customer?



Ans: Based on the city data analysis:

1. Mandalay: Focus on electronic accessories, fashion accessories, and sports/travel gear.

2. Naypyitaw: Prioritize food and beverages, electronic accessories, and fashion accessories.

3. Yangon: Emphasize food and beverages, home goods, and fashion accessories.

Tailor promotions for Member and Normal customers based on their buying behaviours and preferences.

Conclusion and Review :

Looking at the supermarket data provides valuable insights into customer behaviour and product performance. It helps the store understand what customers prefer, which products are selling well, and how different branches are performing. By analysing customer ratings, the store can gauge satisfaction levels and identify areas for improvement. Moreover, studying temporal trends allows the store to anticipate peak hours and plan promotions accordingly. Overall, leveraging these insights enables the store to make informed decisions about inventory management, pricing strategies, and customer satisfaction initiatives, ultimately driving business growth and success.