

# Supermarket Sales Dataset Analysis

## Introduction:

Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of supermarket sales. From fundamental transactional details such as Invoice ID, Date, Time, and Payment Method to more nuanced factors like Branch Location, Customer Type, Gender Demographics, Product Line, and Product Ratings, every facet has been meticulously documented.

## Questionnaire:

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

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

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

Q4. Analyzing units sold and unit price data answer the following sub questions

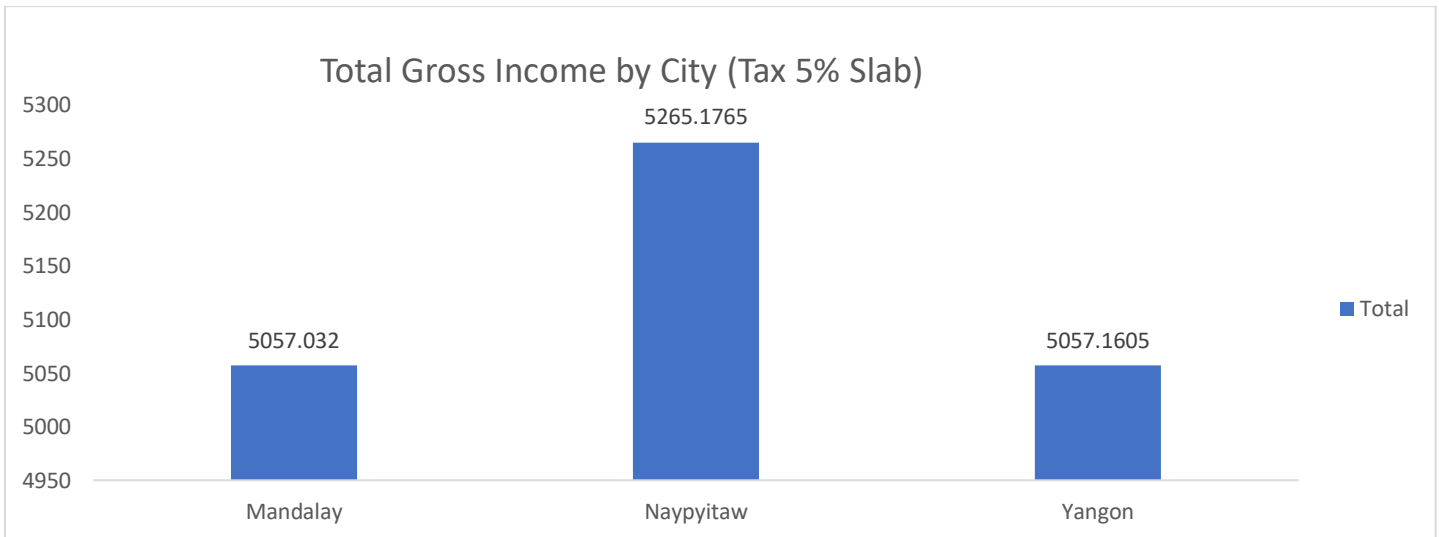
- a) What is the degree of freedom?
- b) Co-relation of Unit price and revenue generated
- c) What result you can draw from regression of the two data

Q5. What product will you suggest as per the city data analysis to each type of customer?

## Analytics:

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

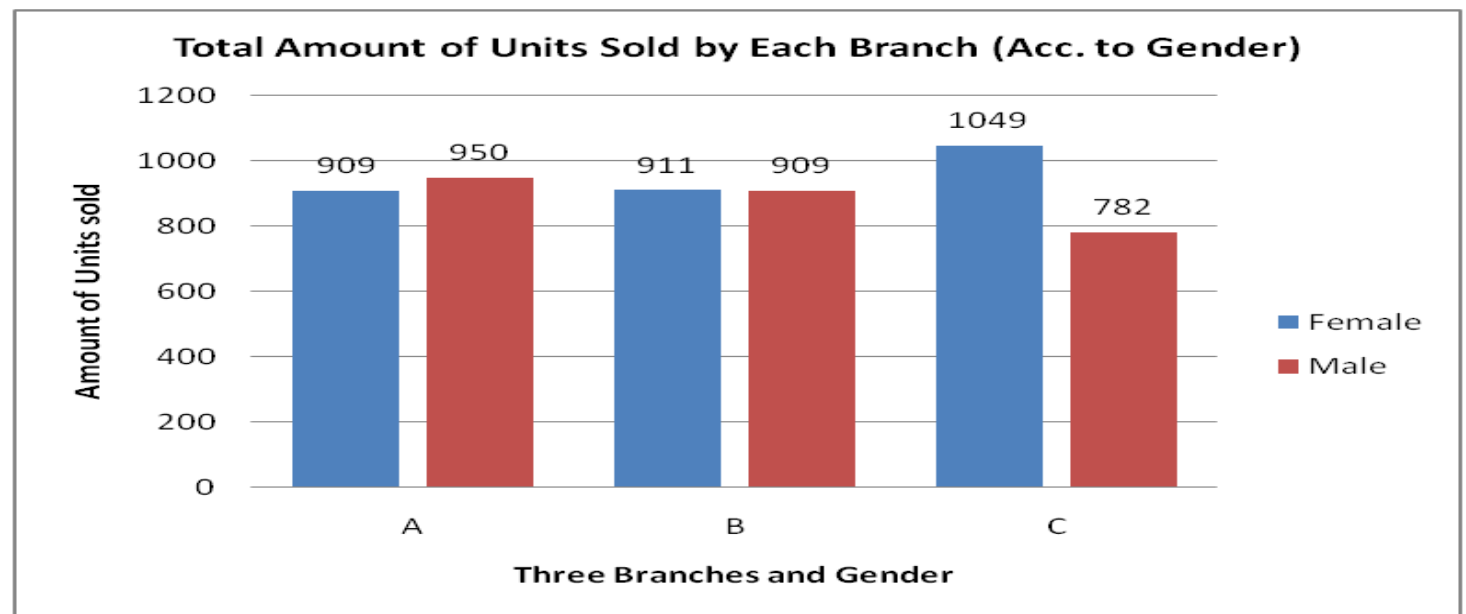
Ans



Based on the data analyzed, the city that outperformed all is **Mandalay**. This conclusion is drawn from superior performance in total sales/revenue generation compared to the other cities in the same tax slab of 5%.

**Q2. Which customer gender ordered most items from all the three branches?**

Ans.



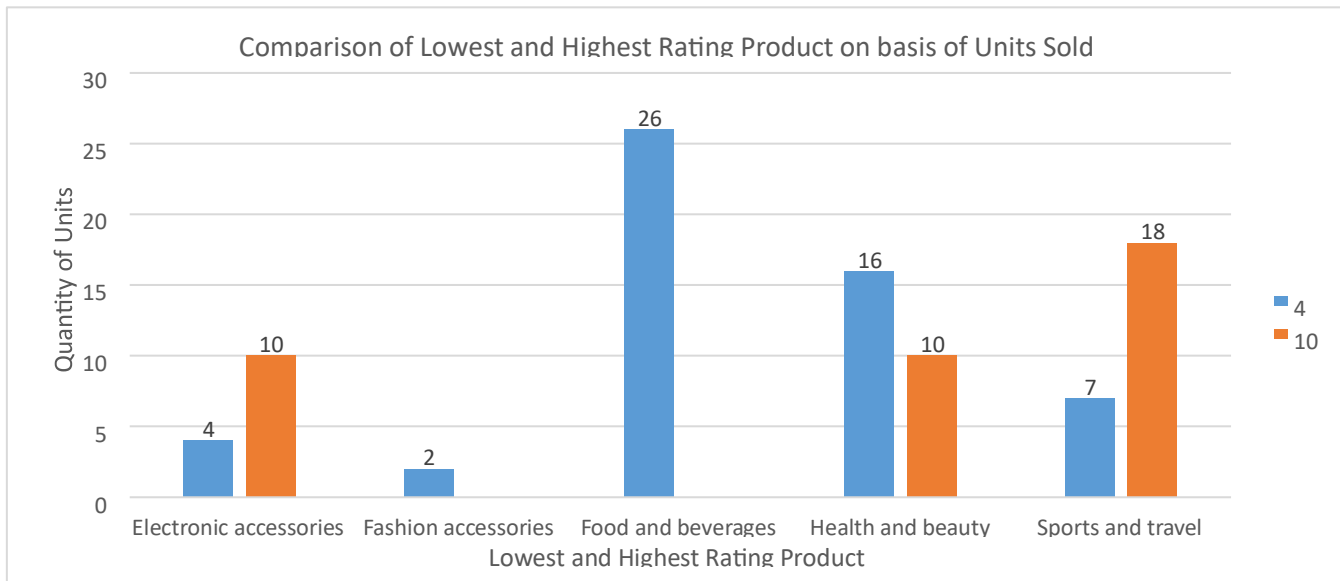
Our analysis of the Supermarket Sales Data revealed the following:

- At Branch A, females placed the highest number of orders.
- Branch B saw higher number of orders placed by Females
- Meanwhile, at Branch C, males placed the most orders.

| Quantity | Gender | Branch |
|----------|--------|--------|
| 1        | Female | A      |
| 2        | Male   | B      |
| 3        |        | C      |
| 4        |        |        |
| 5        |        |        |
| 6        |        |        |
| 7        |        |        |
| 8        |        |        |

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

**Ans.**



Upon analyzing the Supermarket Sales Data, we discovered that product ratings ranged from a minimum of 4 to a maximum of 10.

- Electronic Accessories with higher ratings garnered more customer purchases, indicating a preference for quality in this category.
- Fashion accessories and food and beverages mainly comprised lower-rated products in customer purchases.
- Health and beauty products also leaned towards lower-rated items in customer preferences.
- However, in the Sports and Travel category, customers showed a tendency to purchase higher-rated products.

**Q4. Analyzing units sold and unit price data answer the following sub questions**

- What is the degree of freedom?
- Co-relation of Unit price and revenue generated
- What result you can draw from regression of the two data

**Ans.**

|                              |              |  |  |  |  |  |
|------------------------------|--------------|--|--|--|--|--|
| SUMMARY OUTPUT               |              |  |  |  |  |  |
|                              |              |  |  |  |  |  |
| <i>Regression Statistics</i> |              |  |  |  |  |  |
| Multiple R                   | 0.010777564  |  |  |  |  |  |
| R Square                     | 0.000116156  |  |  |  |  |  |
| Adjusted R Square            | -0.000885732 |  |  |  |  |  |
| Standard Error               | 2.924724997  |  |  |  |  |  |

|              |                     |                       |               |                |                       |                  |
|--------------|---------------------|-----------------------|---------------|----------------|-----------------------|------------------|
| Observations | 1000                |                       |               |                |                       |                  |
|              |                     |                       |               |                |                       |                  |
| ANOVA        |                     |                       |               |                |                       |                  |
|              | <i>df</i>           | <i>SS</i>             | <i>MS</i>     | <i>F</i>       | <i>Significance F</i> |                  |
| Regression   | 1                   | 0.9917274             | 0.991727      | 0.115937       | 0.733555221           |                  |
| Residual     | 998                 | 8536.908273           | 8.554016      |                |                       |                  |
| Total        | 999                 | 8537.9                |               |                |                       |                  |
|              |                     |                       |               |                |                       |                  |
|              | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i>      | <i>Upper 95%</i> |
| Intercept    | 5.443794599         | 0.215314544           | 25.28299      | 2.1E-109       | 5.021273429           | 5.86631577       |
| Unit price   | 0.001189202         | 0.003492565           | 0.340495      | 0.733555       | -0.005664411          | 0.008042815      |

- The degree of freedom of the analyzed data is 1.
- The correlation between unit price and generated revenue was found to be 0.63392, indicating a moderate positive relationship. The analysis focused on the columns of unit price and total revenue, employing the CORREL function.
- Upon examining the regression results, we aimed to discern the relationship between quantity and unit price, exploring how customers' purchasing quantity correlates with the unit price of a product.

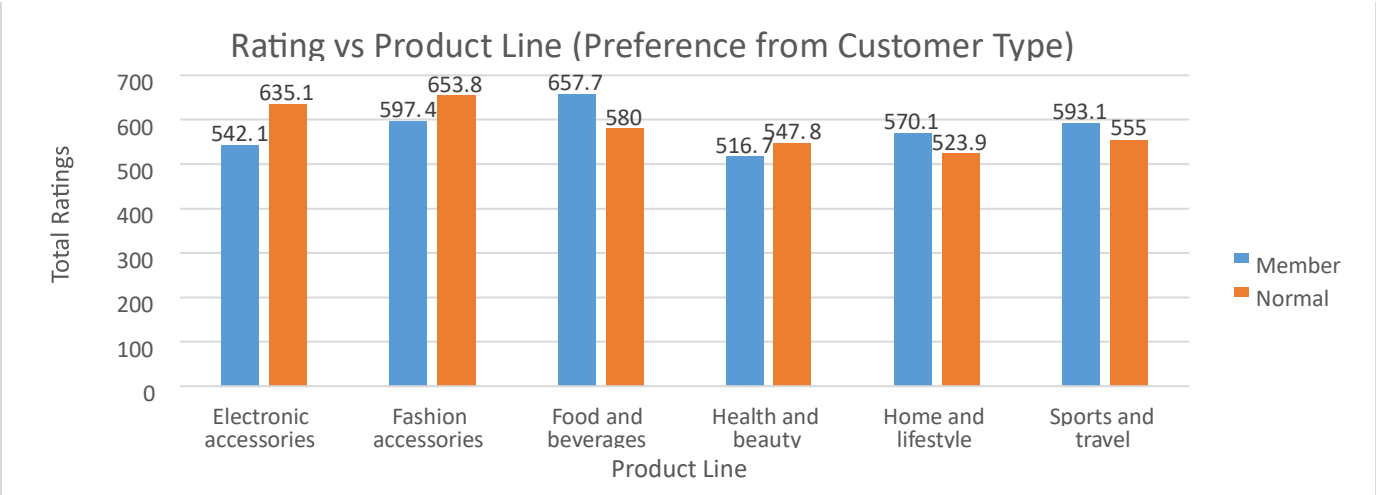
However, from the regression analysis, it's evident that the observed trend lacks consistency. The expected outcomes derived from the trend deviate significantly from the actual outcomes.

With a degree of freedom of 1, the trendline equation stands as

Quantity = 0.0012x + 5.4438. Despite this equation, the coefficient of determination (R<sup>2</sup>) is merely 0.0001, highlighting the inconsistency in customer buying patterns solely based on unit price.

#### Q5.What product will you suggest as per the city data analysis to each type of customer

Ans. As per the city Data Analysis, **Food and Beverages** will be a good option for **Member** type customer and **Fashion Accessories** for **Normal** type of customers.



| Rating | Customer type | Product line           |
|--------|---------------|------------------------|
| 4      | Member        | Electronic accessories |
| 4.1    | Normal        | Fashion accessories    |
| 4.2    |               | Food and beverages     |
| 4.3    |               | Health and beauty      |
| 4.4    |               | Home and lifestyle     |
| 4.5    |               | Sports and travel      |
| 4.6    |               |                        |
| 4.7    |               |                        |

## Conclusion and Reviews

In summary, the analysis of supermarket sales dynamics reveals valuable insights into consumer behavior and operational trends. Key findings include Mandalay's strong performance, gender-specific ordering patterns, and product recommendations based on city data. Further exploration is recommended on the relationship between product ratings and sales volume, as well as unit price correlation. Clear visuals can enhance understanding, and the report provides actionable recommendations for targeted marketing and strategic investment

