

Supermarket Sales Analysis

SUBMITTED TO:

MD. ERFAN

ASSISTANT PROFESSOR & CHAIRMAN

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

UNIVERSITY OF BARISHAL

mdtanviranitullah@gmail.com

Computer Fundamentals & Office Applications

12/17/24

SUBMITTED BY:

MD TANVIR ANAYETULLAH

DEPARTMENT OF ACCOUNTING & INFORMATION SYSTEMS

UNIVERSITY OF BARISHAL

EDGE ROLL : 01-045-08

Introduction

The ability to analyze supermarket sales data effectively is essential for gaining valuable insights into customer behavior, product performance, and overall business operations. This project focuses on exploring a comprehensive dataset of supermarket sales to answer critical business questions, as outlined in the accompanying project document. By addressing these questions, we aim to uncover actionable trends and patterns to support strategic decision-making.

The scope of this project encompasses several key areas of analysis, including customer demographics, product and pricing performance, sales trends, profitability, and multivariate analysis. Some of the questions we aim to answer include identifying the branch with the highest sales, understanding customer purchasing behavior based on membership status, evaluating the performance of different product lines, and determining the impact of payment methods on revenue.

Using Microsoft Excel as the primary tool for analysis, the project involves data cleaning, visualization, and interpretation to present meaningful findings. This report is structured to systematically address the questions provided, offering detailed insights and recommendations based on the data.

By the end of this project, the analysis will provide stakeholders with a clear understanding of key sales metrics, customer behavior trends, and factors influencing profitability, enabling data-driven strategies to optimize business performance.

Methodology

The methodology for this project is structured into several systematic steps to ensure a thorough analysis of the supermarket sales data. The approach combines data preparation, exploration, and analysis techniques to address the specific business questions outlined in the project document.

1. Data Collection

For this analysis we need a supermarket sales dataset. So kaggle.com is the best way to get dataset. There are hundreds of datasets uploaded here and I can create my own datasets. To collect a proper dataset I first took help from chatgpt to find out which dataset from kaggle.com is good for me. Accounting related and sales related dataset in chatgpt between kaggle.com website chatgpt gives me idea about supermarket sales report. Then I went to the kaggle.com website and searched by typing supermarket sales report and I got this dataset which link is <https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales>. The primary dataset, containing detailed supermarket sales records, was obtained in Excel format.

A supporting document outlining the project questions provided a clear framework for the analysis.

2. Data Preparation

- **Data Cleaning:**
 - The dataset was examined for missing, duplicate, or inconsistent entries.
 - Necessary corrections and formatting were applied to ensure data accuracy and reliability.
- **Data Organization:**
 - Variables such as branch, city, customer type, gender, product line, unit price, quantity, rating, payment methods, and revenue metrics were identified and categorized.
 - Derived variables (e.g., total revenue, average gross income) were calculated where required.

•

3. Exploratory Data Analysis (EDA)

- **Descriptive Statistics:**
 - Summary statistics (e.g., mean, median, mode, and standard deviation) were calculated for key variables to understand general trends.
- **Visualization:**
 - Charts (e.g., bar charts, pie charts, and scatter plots) and graphs were used to identify patterns and trends across branches, product lines, and customer segments.

4. Analytical Techniques

- **Customer and Demographic Analysis:**
 - Sales and revenue trends were segmented by branch, city, customer type, and gender.
 - Comparative analysis was performed to explore differences in purchasing behavior.
- **Product and Pricing Analysis:**
 - Revenue and profitability metrics were analyzed across product lines.
 - Relationships between variables such as unit price and rating were examined using correlation techniques.
- **Sales and Performance Analysis:**
 - Time-based analysis was conducted to determine peak shopping hours and preferred payment methods.
 - Revenue contributions from different payment methods were compared.
- **Profitability Analysis:**
 - Gross income and transaction sizes were assessed to identify high-performing branches and product lines.
- **Customer Behavior and Trends**
- **Multivariate Analysis:**
 - Relationships among multiple variables (e.g., city, customer type, payment method) were evaluated using pivot tables and advanced statistical methods.

5. Tools and Software

- Microsoft Excel:
 - Used for data cleaning, calculations, and visualization.
 - Pivot tables and Excel functions (e.g., VLOOKUP, SUMIF, and COUNTIF) facilitated detailed analysis.
- Statistical Techniques:
 - Correlation and comparative analyses were performed to explore relationships and differences across variables.

6. Interpretation and Reporting

- Key findings were synthesized into actionable insights aligned with the business questions provided.
- Data visualizations and summary tables were included to support the findings and make the results easier to understand.
- Recommendations were developed based on the analysis, with a focus on enhancing sales performance and profitability.

This structured approach ensures a comprehensive and accurate analysis of the supermarket sales data, providing meaningful insights for informed decision-making.

For the analysis I select some questions:

1. Customer and Demographic Analysis

- a. Which branch has the highest sales volume, and how does it vary by city?
- b. How does the "Customer type" (e.g., Member vs. Non-member) influence purchasing patterns (e.g., average total sales or quantity purchased)?
- c. Is there a significant difference in the average rating given by male vs. female customers?
- d. Which gender contributes more to overall sales revenue?
- e. What is the average total revenue per transaction for each branch?
- f. What is the percentage distribution of customer types (Member vs. Normal)?

2. Product and Pricing Analysis

- a. Which "Product line" generates the highest revenue across all branches?
- b. What is the relationship between "Unit price" and "Rating"? Do higher-priced products receive higher ratings?
- c. Which "Product line" has the highest average gross income and the highest average cogs?
- d. Are any product lines associated with consistently high or low ratings?

3. Sales and Performance Analysis

- a. During what time of day (e.g., morning, afternoon, evening, night) do customers spend the most on purchases?
- b. How does the "Payment" method (e.g., Cash, Credit Card) impact the total revenue generated?
- c. What is the average "Tax 5%" collected per transaction for each branch?

d. What is the most frequently used payment method?

4. Profitability Analysis

- Which branch has the highest average "Gross income" per transaction?
- What is the correlation between "Quantity" purchased and "Gross income"?

5. Customer Behavior and Trends

- Do "Members" purchase higher quantities on average compared to "Non-members"?
- What is the average transaction size (based on "Total") for each "Payment" method?

6. Multivariate Analysis

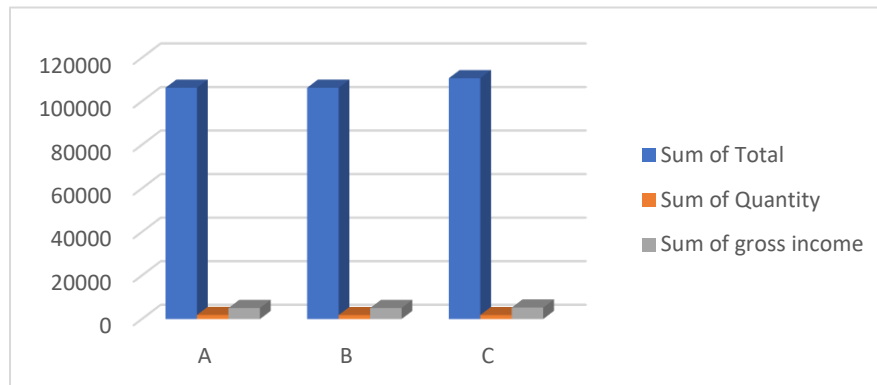
- Are there significant differences in "Gross income" based on "City", "Customer type", and "Payment" method combined?
- How does the "Rating" differ across combinations of "Branch" and "Product line"?
- Is there any significant trend between "Quantity" purchased and "Total" revenue across cities?

Answer the Question

1. Customer and Demographic Analysis

- Which branch has the highest sales volume, and how does it vary by city?

Row Labels	Sum of Total	Sum of Quantity	Sum of gross income
A	106200.3705	1859	5057.1605
B	106197.672	1820	5057.032
C	110568.7065	1831	5265.1765



From the above table and graph we find A branch has the highest quantity and C branch has the highest gross income.

- b. How does the "Customer type" (e.g., Member vs. Non-member) influence purchasing patterns (e.g., average total sales or quantity purchased)?

Customer Type	Average Quantity	of Average of Total
Member	5.558882236	327.7913054
Normal	5.460921844	318.1228557

Here we see that member customers are purchase higher quantity per purchase than normal customer.

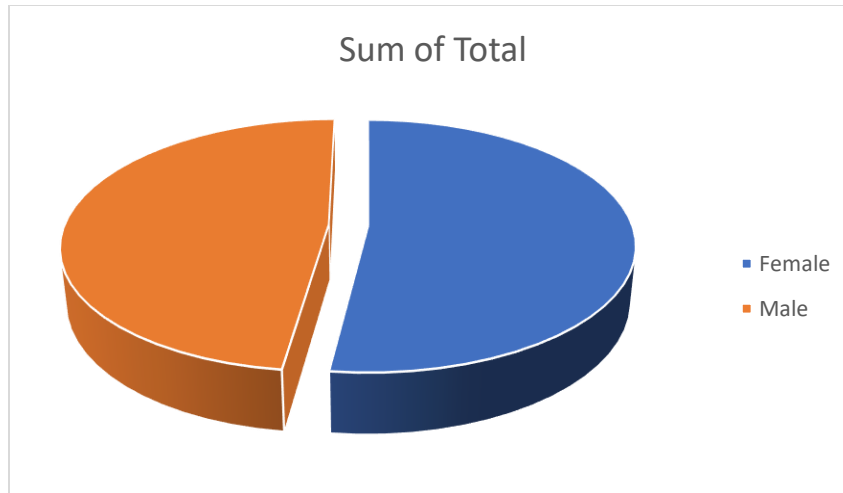
- c. Is there a significant difference in the average rating given by male vs. female customers?

Gender	Average of Rating
Female	6.964471058
Male	6.980961924

There is a slightly difference but no significant difference in average rating given by male and female customers.

- d. Which gender contributes more to overall sales revenue?

Gender	Sum of Total	Sum of gross income
Female	167882.925	7994.425
Male	155083.824	7384.944



From the above analysis we can evaluate that female customers are contribute the more in overall sales revenue.

e. What is the average total revenue per transaction for each branch?

Branch	Average of gross income	Sum of gross income
A	14.87400147	5057.1605
B	15.2320241	5057.032
C	16.05236738	5265.1765

f. What is the percentage distribution of customer types (Member vs. Normal)?

Customer type	Count of Customer type	Percentage
Member	501	50.1%
Normal	499	49.9%
Grand Total	1000	100%

There is no significant difference in the percentage of member and normal customers.

2. Product and Pricing Analysis

a. Which "Product line" generates the highest revenue across all branches?

Row Labels	Sum of gross income
Electronic accessories	2587.5015
A	872.2435
B	811.9735
C	903.2845
Fashion accessories	2585.995
A	777.7385
B	781.5865
C	1026.67
Food and beverages	2673.564
A	817.2905
B	724.5185
C	1131.755
Health and beauty	2342.559
A	599.893
B	951.46
C	791.206
Home and lifestyle	2564.853
A	1067.4855
B	835.6745
C	661.693
Sports and travel	2624.8965
A	922.5095
B	951.819
C	750.568
Grand Total	15379.369

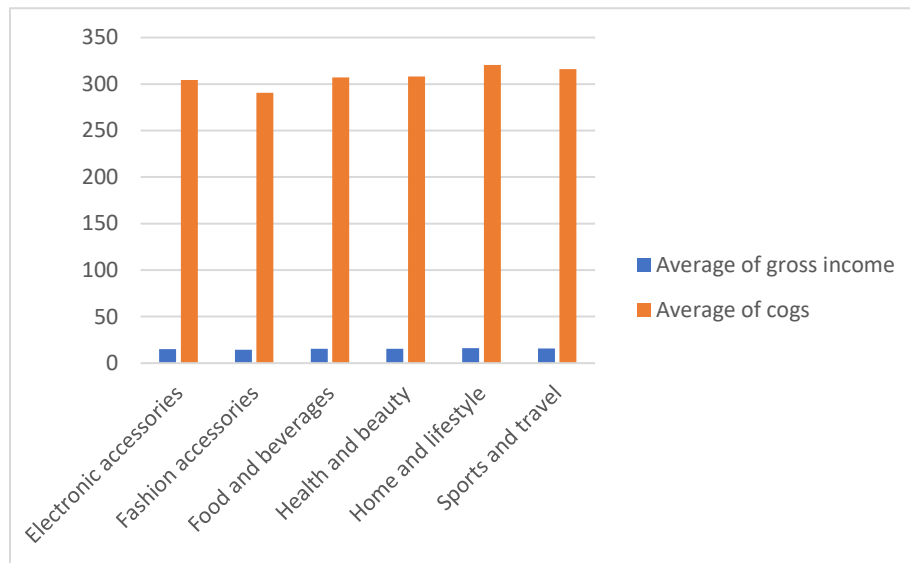
In all the product lines, **Food and beverages** generates the highest revenue across all branches.

b. What is the relationship between "Unit price" and "Rating"? Do higher-priced products receive higher ratings?

-0.008777507 is the relationship between "Unit price" and "Rating". When unit price is increasing rating is decreased by 0.008777507.

- c. Which "Product line" has the highest average gross income and the highest average cogs?

Row Labels	Average of gross income	Average of cogs
Electronic accessories	15.22059706	304.4119412
Fashion accessories	14.5280618	290.561236
Food and beverages	15.36531034	307.3062069
Health and beauty	15.41157237	308.2314474
Home and lifestyle	16.03033125	320.606625
Sports and travel	15.81262952	316.2525904



From the following analysis, Home and lifestyle has the highest average gross income which is 16.03033125 and highest average cost of goods sold which is almost 321.

- d. Are any product lines associated with consistently high or low ratings?

Row Labels	Average of Rating
Electronic accessories	6.924705882
Fashion accessories	7.029213483
Food and beverages	7.113218391
Health and beauty	7.003289474
Home and lifestyle	6.8375
Sports and travel	6.91626506
Grand Total	6.9727

No, there is no product lines associated with consistently high or low ratings.

3. Sales and Performance Analysis

- a. During what time of day (e.g., morning, afternoon, evening, night) do customers spend the most on purchases?

For this problem, Firstly I divide times into 3 catagories. 6am to 12pm is Morning, 12pm to 6pm is Afternoon and 6pm to 12am is Evening. After catagories the time into Morning, Afternoon and Evening we find that Customers are mostly spend or purchase at afternoon.

Row Labels	Sum of Total
Afternoon	172468.5585
Evening	88699.38
Morning	61798.8105
Grand Total	322966.749

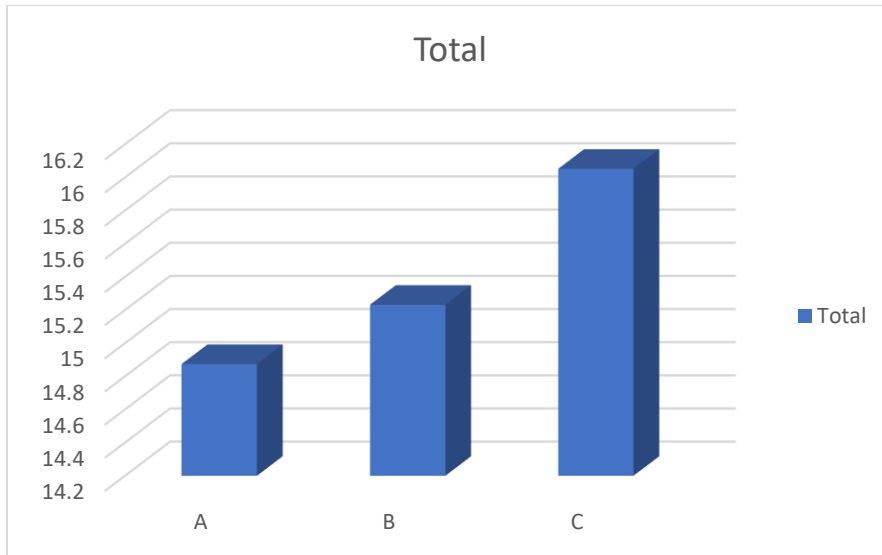
- b. How does the "Payment" method (e.g., Cash, Credit Card) impact the total revenue generated?

Row Labels	Sum of gross income
Cash	5343.17
Credit card	4798.432
Ewallet	5237.767
Grand Total	15379.369

There is no significant impact by payment method on total revenue.

- c. What is the average "Tax 5%" collected per transaction for each branch?

Row Labels	Average of Tax 5%
A	14.87400147
B	15.2320241
C	16.05236738
Grand Total	15.379369



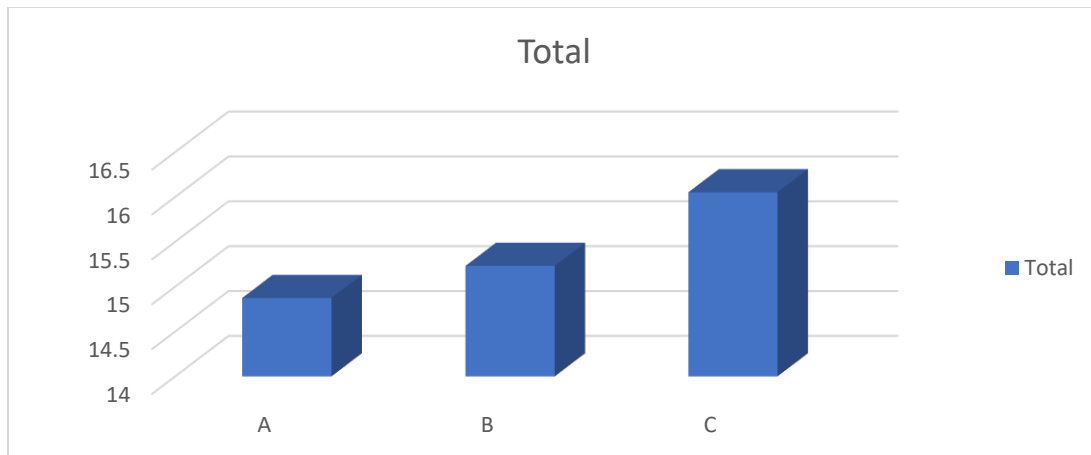
d. What is the most frequently used payment method?

Row Labels	Count of Payment
Cash	344
Credit card	311
Ewallet	345

4. Profitability Analysis

a. Which branch has the highest average "Gross income" per transaction?

Row Labels	Average of gross income
A	14.87400147
B	15.2320241
C	16.05236738
Grand Total	15.379369



b. What is the correlation between "Quantity" purchased and "Gross income"?

0.705510186 is the correlation between "Quantity" purchased and "Gross income".
When "Quantity" increased "gross income" increased.

5. Customer Behavior and Trends

a. Do "Members" purchase higher quantities on average compared to "Non-members"?

Row Labels	Average of Quantity
Member	5.558882236
Normal	5.460921844
Grand Total	5.51

Yes members purchase higher quantities.

b. What is the average transaction size (based on "Total") for each "Payment" method?

Row Labels	Average of Total
Cash	326.1818895
Credit card	324.0098778
Ewallet	318.8206
Grand Total	322.966749

6. Multivariate Analysis

- a. Are there significant differences in "Gross income" based on "City", "Customer type", and "Payment" method combined?

Row Labels	Sum of gross income
Mandalay	5057.032
Member	2557.366
Cash	805.4515
Credit card	1076.57
Ewallet	675.3445
Normal	2499.666
Cash	877.38
Credit card	701.7565
Ewallet	920.5295
Naypyitaw	5265.1765
Member	2708.6325
Cash	955.421
Credit card	950.5845
Ewallet	802.627
Normal	2556.544
Cash	1096.2865
Credit card	493.5805
Ewallet	966.677
Yangon	5057.1605
Member	2554.1655
Cash	842.033
Credit card	723.868
Ewallet	988.2645
Normal	2502.995
Cash	766.598
Credit card	852.0725
Ewallet	884.3245
Grand Total	15379.369

There is no significant difference in gross income based on city. But a significant difference are shown in Member and Normal customer in the different cities.

b. How does the "Rating" differ across combinations of "Branch" and "Product line"?

Row Labels	Average of Rating
A	7.027058824
Electronic accessories	6.911666667
Fashion accessories	6.878431373
Food and beverages	7.253448276
Health and beauty	6.9
Home and lifestyle	6.930769231
Sports and travel	7.257627119
B	6.818072289
Electronic accessories	7.116363636
Fashion accessories	6.722580645
Food and beverages	6.994
Health and beauty	7.1
Home and lifestyle	6.516
Sports and travel	6.509677419
C	7.072865854
Electronic accessories	6.747272727
Fashion accessories	7.44
Food and beverages	7.08030303
Health and beauty	6.998076923
Home and lifestyle	7.06
Sports and travel	7.028888889
Grand Total	6.9727

There is a significant rating difference between different Branches. From the following analysis Branch A has the higher rating compare to other branches.

c. Is there any significant trend between "Quantity" purchased and "Total" revenue across cities?

Row Labels	Sum of Quantity	Sum of gross income
Mandalay	1820	5057.032
Naypyitaw	1831	5265.1765
Yangon	1859	5057.1605
Grand Total	5510	15379.369

No, there is no significant trend between "Quantity" purchased and "Total revenue" across the cities.