Optimizing Supermarket Sales With Data-Driven Analysis And Dashboard Creation

An In-depth Study Using Historical Sales Data

Prepared by:

Sachith Yamannage

Contact Information:

Email: Sachithnimesh999@gmail.com

Date of Completion:

June 2024

Executive Summary

This project focuses on optimizing supermarket sales through data-driven analysis and the creation of interactive dashboards. The primary objectives were to analyze historical sales data spanning three branches over three months, identify key sales trends, and provide actionable insights to enhance business operations.

Methods employed included data cleaning, SQL-based queries for sales analysis by product line, customer segment, and payment method, and visualization of findings using charts and graphs. Key findings revealed the top-selling products across branches, customer preferences by gender and membership status, peak sales times, and customer satisfaction levels.

Recommendations include leveraging insights to tailor marketing strategies, optimize inventory management, and improve customer service experiences. Implementing these strategies is expected to enhance overall sales performance and customer satisfaction, positioning the supermarket chain competitively in the market.

Introduction

In today's competitive retail landscape, supermarkets face significant challenges in optimizing sales and enhancing customer satisfaction. The advent of data analytics has revolutionized decision-making processes, enabling businesses to extract valuable insights from vast datasets. This project focuses on leveraging data analytics techniques to analyze historical sales data from a supermarket chain. By examining sales trends, customer purchasing behaviors, and operational efficiencies across three branches over a three-month period, the project aims to uncover actionable insights.

The objectives of this study include identifying best-selling products, understanding customer preferences based on gender and membership status, analyzing sales performance by payment method, and evaluating customer satisfaction levels. Through the use of SQL queries and interactive dashboard visualizations, this analysis provides a comprehensive overview of the supermarket's performance metrics. The findings from this study will inform strategic decisions aimed at improving inventory management, optimizing marketing campaigns, and ultimately enhancing the overall shopping experience for customers.

By adopting a data-driven approach, this project seeks to empower supermarket management with the tools and insights needed to make informed decisions that drive growth and competitiveness in the dynamic retail environment. The following sections will detail the methodology, data analysis techniques, key findings, and actionable recommendations derived from this study.

Dataset Description

The dataset used in this project comprises historical sales records from a supermarket chain operating across three branches in various cities. It spans a period in 2019, capturing detailed transactional information for each sale.

Metadata - https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales

Attributes included in the dataset:

- Invoice ID: Unique identifier for each sales transaction.

- Branch: Identifier for the supermarket branch.

- City: Location of the supermarket branch.

- Customer type: Classification of customers as either 'Member' (using a membership card) or 'Normal' (non-member).

- Gender: Gender of the customer making the purchase.

- Product line: Categorization of products into various groups such as Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel.

- Unit price: Price of each product in dollars ($).

- Quantity: Number of units purchased by the customer.

- Tax 5%:Tax amount applied (5% of the total price).

- Total: Total amount paid by the customer, including tax.

- Date: Date of the transaction.

- Time:Time of the transaction (recorded in a 24-hour format).

- Payment: Payment method used by the customer (Cash, Credit card, Ewallet).

- COGS (Cost of Goods Sold): Cost of goods sold.

- Gross margin percentage: Gross margin percentage.

- Gross income: Gross income generated from the sale.

- Rating: Customer rating of their shopping experience on a scale from 1 to 10.

Purpose of the Dataset:

The dataset is invaluable for predictive analytics and business intelligence purposes within the supermarket industry. It allows for the analysis of sales performance across different product categories, customer segments, and branches, providing insights into consumer behavior, operational efficiency, and market trends. Insights derived from this dataset can guide strategic decision-making processes aimed at improving profitability, enhancing customer satisfaction, and optimizing business operations.

Database Design

This section outlines the database design for managing the supermarket sales data using SQL tables. The design includes six tables: Invoice, Customer, Product, Sales, Branch, and City.

Explanation of Database Design:

- **Invoice Table:** Stores details of each sales transaction including invoice ID, branch ID, customer ID, product ID, date, time, payment method, and customer rating. Foreign keys reference the Branch, Customer, and Product tables.
- **Customer Table:** Contains information about customers including customer ID, customer type (Member or Normal), and gender.
- **Product Table:** Holds data on products including product ID (auto-incremented for uniqueness), product line, and unit price.
- **Sales Table:** Records specific details of each sales transaction such as sales ID (auto-incremented), invoice ID, quantity purchased, tax applied, total amount paid, cost of goods sold (COGS), gross margin percentage, and gross income. It references the Invoice table via the invoice ID.
- **Branch Table:** Identifies each branch with a unique branch ID (e.g., A, B, C) and links it to a specific city ID in the City table. It references the City table.
- **City Table:** Stores details of cities where branches are located, with a unique city ID and city name.
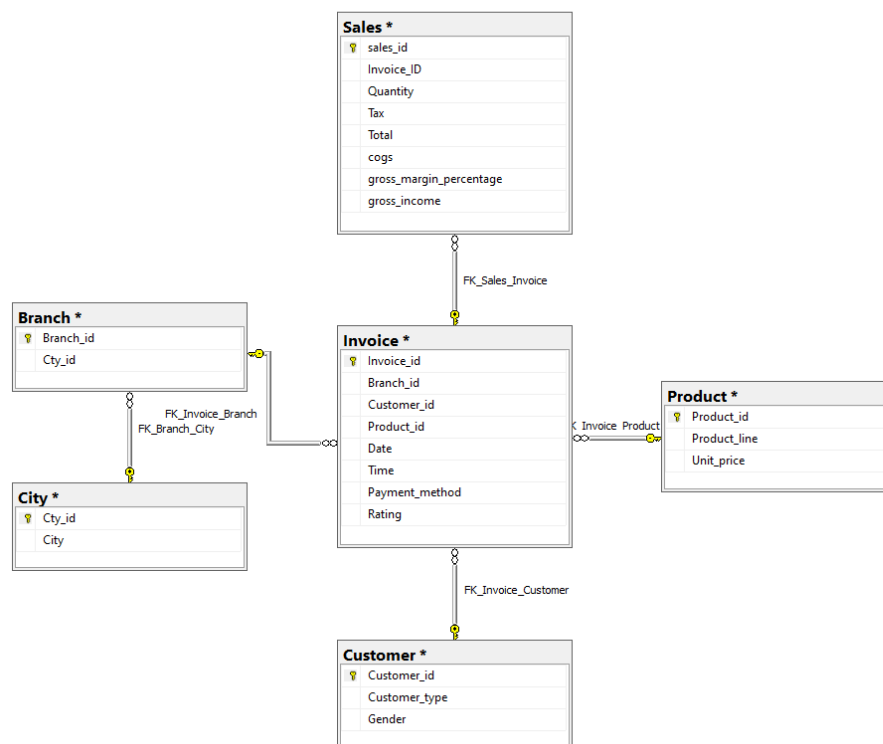
Business Problems and Analytical Approach

In this section of the report, we identify key business problems encountered by the supermarket chain and outline the analytical approach used to address these challenges through data-driven analysis and dashboard creation.

Identified Business Problems:

1. **Optimizing Product Performance Across cities:**
   - **Problem:** Determine which products are the best-sellers in each city to optimize inventory management and marketing strategies.
2. **Understanding Customer Purchase Behavior:**
   - **Problem:** Analyze customer purchasing patterns by gender and customer type (Member vs. Normal) to tailor promotions and improve customer satisfaction.
3. **Effectiveness of Payment Methods:**
   - **Problem:** Evaluate the popularity and performance of different payment methods (Cash, Credit Card, Ewallet) to streamline payment processes and enhance customer experience.
4. **Peak Sales Times and Operational Efficiency:**
   - **Problem:** Identify peak sales hours to optimize staffing and operational efficiency, ensuring adequate resources during high-demand periods.
5. **Customer Satisfaction Analysis:**
   - **Problem:** Analyze customer ratings and feedback to identify areas for improvement in service delivery and product offerings.

Analytical Approach:

1. **Optimizing Product Performance Across cities:**
   - **Approach:** Utilize SQL queries to calculate total sales and quantity sold for each product across cities.

```sql
SELECT
    Product.Product_line,
    City.City,
    SUM(Sales.Quantity) AS TotalQuantitySold,
    SUM(Sales.Total) AS TotalSales
FROM
    [Sales analysis].[dbo].[Sales] AS Sales
JOIN
    [Sales analysis].[dbo].[Invoice] AS Invoice ON Sales.[Invoice_ID] = Invoice.Invoice_id
JOIN
    [Sales analysis].[dbo].[Product] AS Product ON Invoice.Product_id = Product.Product_id
JOIN
    [Sales analysis].[dbo].[Branch] AS Branch ON Invoice.Branch_id = Branch.Branch_id
JOIN
    [Sales analysis].[dbo].[City] AS City ON Branch.Cty_id = City.Cty_id
GROUP BY
    Product.Product_line,
    City.City
ORDER BY
    Product_     column City(nvarchar, not null)
    City.City;
```

Results | Messages

Product_line	City	TotalQuantitySold	TotalSales
Electronic accessories	Mandalay	316	17051.4434318542
Electronic accessories	Naypyitaw	333	18968.9744625092
Electronic accessories	Yangon	322	18317.1135635376
Fashion accessories	Mandalay	297	16413.3165130615
Fashion accessories	Naypyitaw	342	21560.0701627731
Fashion accessories	Yangon	263	16332.5085086823
Food and beverages	Mandalay	270	15214.888507843
Food and beverages	Naypyitaw	369	23766.8549594879
Food and beverages	Yangon	313	17163.1004638672
Health and beauty	Mandalay	320	19980.6600799561
Health and beauty	Naypyitaw	277	16615.3259010315
Health and beauty	Yangon	257	12597.7529392242
Home and lifestyle	Mandalay	295	17549.1645393372
Home and lifestyle	Naypyitaw	245	13895.5530767441
Home and lifestyle	Yangon	371	22417.195514679
Sports and travel	Mandalay	322	19988.1991157532
Sports and travel	Naypyitaw	265	15761.9280557632
Sports and travel	Yangon	333	19372.6994094849

2. Understanding Customer Purchase Behavior:

- o **Approach:** Segment sales data by gender and customer type using SQL queries

```sql
SELECT
    [Sales analysis].[dbo].[Customer].[Gender],
    [Sales analysis].[dbo].[Customer].[Customer_type],
    SUM([Sales analysis].[dbo].[Sales].[Quantity]) AS TotalQuantity,
    SUM([Sales analysis].[dbo].[Sales].[Total]) AS TotalSales
FROM
    [Sales analysis].[dbo].[Sales]
JOIN
    [Sales analysis].[dbo].[Invoice] ON [Sales analysis].[dbo].[Sales].[Invoice_ID] = [Sales analysis].[dbo].[Invoice].[Invoice_id]
JOIN
    [Sales analysis].[dbo].[Customer] ON [Sales analysis].[dbo].[Invoice].[Customer_id] = [Sales analysis].[dbo].[Customer].[Customer_id]
GROUP BY
    [Sales analysis].[dbo].[Customer].[Gender],
    [Sales analysis].[dbo].[Customer].[Customer_type]
ORDER BY
    [Sales analysis].[dbo].[Customer].[Gender],
    [Sales analysis].[dbo].[Customer].[Customer_type];
```

00 % ▾ ◂

▦ Results ▦ Messages

	Gender	Customer_type	TotalQuantity	TotalSales
1	Female	Member	1492	88146.9434356689
2	Female	Normal	1377	79735.9816303253
3	Male	Member	1293	76076.500623703
4	Male	Normal	1348	79007.323515892

3. Effectiveness of Payment Methods:
   a. **Approach:** Analyze sales distribution by payment method using SQL queries.

```sql
SELECT
    [Sales analysis].[dbo].[Invoice].[Payment_method],
    COUNT(*) AS TotalTransactions,
    SUM([Sales analysis].[dbo].[Sales].[Total]) AS TotalSales
FROM
    [Sales analysis].[dbo].[Sales]
JOIN
    [Sales analysis].[dbo].[Invoice] ON [Sales analysis].[dbo].[Sales].[Invoice_ID] = [Sales analysis].[dbo].[Invoice].[Invoice_id]
GROUP BY
    [Sales analysis].[dbo].[Invoice].[Payment_method]
ORDER BY
    TotalSales DESC;
```

) % ▾ ◂

▦ Results ▦ Messages

Payment_method	TotalTransactions	TotalSales
Cash	344	112206.569979668
Ewallet	345	109993.106870651
Credit card	311	100767.07235527

4. Peak Sales Times and Operational Efficiency:
   a. **Approach:** Aggregate sales data by hour using SQL queries.

```sql
SELECT
    DATEPART(HOUR, [Sales analysis].[dbo].[Invoice].[Time]) AS SalesHour,
    COUNT(*) AS TotalTransactions,
    SUM([Sales analysis].[dbo].[Sales].[Total]) AS TotalSales
FROM
    [Sales analysis].[dbo].[Sales]
JOIN
    [Sales analysis].[dbo].[Invoice] ON [Sales analysis].[dbo].[Sales].[Invoice_ID] = [Sales analysis].[dbo].[Invoice].[Invoice_id]
GROUP BY
    DATEPART(HOUR, [Sales analysis].[dbo].[Invoice].[Time])
ORDER BY
    SalesHour;
```

.00 %

⊞ Results ▤ Messages

	SalesHour	TotalTransactions	TotalSales
1	10	101	31421.4810028076
2	11	90	30377.3296012878
3	12	89	26065.8824462891
4	13	103	34723.2270307541
5	14	83	30828.3990039825
6	15	102	31179.50847435
7	16	77	25226.3235359192
8	17	74	24445.2179508209
9	18	93	26030.3400936127
10	19	113	39699.5129795074
11	20	75	22969.5270862579

5. Customer Satisfaction Analysis:
   a. **Approach:** Calculate average customer ratings by city using SQL queries.

```sql
SELECT
    City.City,
    AVG(Invoice.Rating) AS AverageRating
FROM
    [Sales analysis].[dbo].[Invoice] AS Invoice
JOIN
    [Sales analysis].[dbo].[Branch] AS Branch ON Invoice.Branch_id = Branch.Branch_id
JOIN
    [Sales analysis].[dbo].[City] AS City ON Branch.Cty_id = City.Cty_id
GROUP BY
    City.City
ORDER BY
    AverageRating DESC;
```

00 %

⊞ Results ▤ Messages

	City	AverageRating
1	Naypyitaw	7.07286585394929
2	Yangon	7.02705882156596
3	Mandalay	6.81807227450681

By addressing these business problems with the outlined analytical approach, this project aims to provide actionable insights that will support strategic decision-making and operational improvements within the supermarket chain. The next sections will delve into the detailed analysis and findings derived from these methodologies.

Sales Analysis and Insights

1. Optimizing Product Performance Across Cities

Approach: Calculate total sales and quantity sold for each product category across cities.

- **Mandalay**: Electronic Accessories
- **Naypyitaw**: Food and Beverages
- **Yangon**: Home and Lifestyle

This analysis allows us to understand the overall performance of product categories across different cities. For instance, Mandalay shows strong sales in Electronic Accessories, while Naypyitaw focuses more on Food and Beverages. Yangon, on the other hand, sees higher sales in Home and Lifestyle products.

2. Segment Sales Data by Gender and Customer Type

Total Sales: Female customers generated slightly more total sales (88,146.94) compared to male customers (76,076.50).

Customer Type:

- Overall, "Normal" customer types contributed more to total sales than "Member" customers across both genders.

This segmentation helps in understanding purchasing behavior based on gender and customer type, providing insights into customer preferences and spending patterns.

3. Analysis of Sales Distribution by Payment Method

Cash:

- Least popular by transaction volume (344 transactions).
- Second-highest total sales (112,206.57), indicating higher average transaction value.

Ewallet:

- Most popular by transaction volume (345 transactions).
- Slightly lower total sales (109,993.11) compared to cash, suggesting smaller average transaction values.

Credit Card:

- Moderate transaction volume (311 transactions).
- Middle ground in total sales (100,767.07), reflecting average transaction values.

This analysis highlights the distribution of sales across different payment methods, providing insights into customer payment preferences and transaction behaviors.

4. Analysis of Sales Data by Hour

Highest Sales Hours:

- 19 sales hours: Highest total sales (39,699.51).
- Other peak hours include 14 and 13 sales hours as high-sales periods.

Understanding peak sales hours helps optimize staffing and promotional strategies during these times to capitalize on higher customer traffic and sales opportunities.

5. Average Customer Ratings by City

- **Naypyitaw**: Average rating of 7.07
- **Mandalay**: Average rating of 6.82
- **Yangon**: Average rating of 7.03

Analyzing customer satisfaction ratings by city provides insights into customer experiences and perceptions, aiding in identifying areas for improvement and maintaining high levels of customer satisfaction.

These insights derived from the analysis help in optimizing product offerings, understanding customer behavior, improving operational efficiency, and enhancing overall customer satisfaction across different dimensions of sales data.

Dashboard Creation

Comprehensive Sales Performance and Customer Insights Dashboard for Supermarket Chain
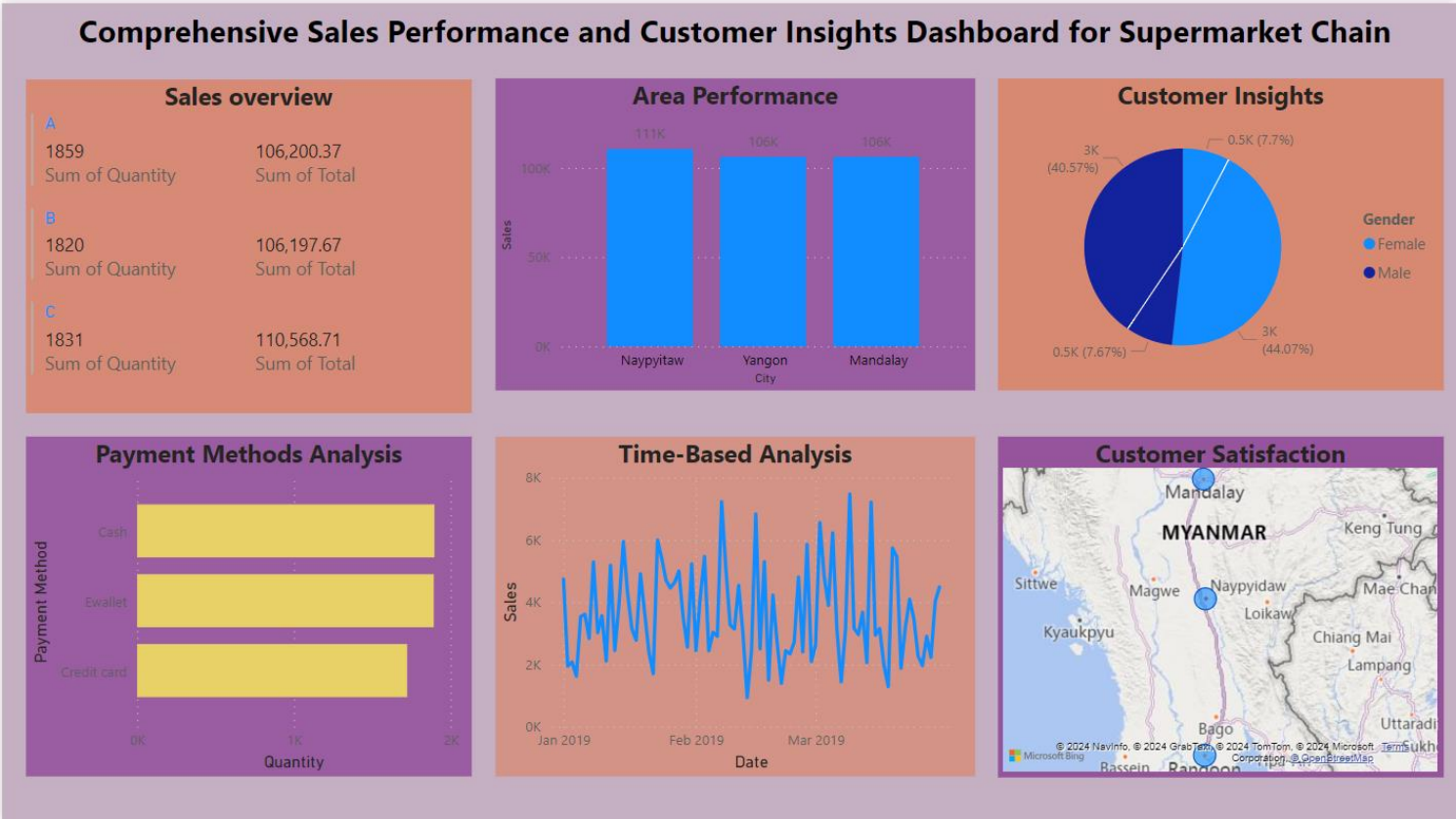
Conclusion

In conclusion, the comprehensive sales performance and customer insights dashboard for the supermarket chain has provided valuable data-driven insights into various aspects of the business. By analyzing sales data across different branches and cities, we identified the top-performing product categories and highlighted regional preferences. The detailed examination of customer purchase behavior, segmented by gender and membership status, revealed significant trends that can inform targeted marketing strategies. Additionally, the analysis of payment methods showcased the varying transaction patterns and preferences among customers, aiding in the optimization of payment processes.

The time-based analysis illuminated peak sales hours, which can guide staffing and operational efficiency, while the assessment of customer satisfaction offered a deeper understanding of customer experiences across different branches. These insights, visualized through intuitive charts and graphs, empower the supermarket chain to make informed decisions, enhance customer engagement, and drive overall business growth.

Ultimately, this dashboard serves as a robust tool for continuous monitoring and improvement, ensuring that the supermarket chain remains competitive in a dynamic market landscape. By leveraging these insights, the supermarket can enhance its product offerings, optimize operations, and elevate the customer shopping experience, thereby achieving sustained success and growth.