1. **Project Overview**

**Title:**

**Supermarket Sales Analysis: Trends, insights, and Recommendations**

**Introduction/Summary:**

Managing a supermarket is no easy task – now imagine running three in different cities. How do you track sales performance across locations, understand customer preferences, identify which products drive the most revenue, and determine the best times to adjust inventory? Are membership customers spending more than regular shoppers? Without data analysis, these crucial insights remain hidden.

This project analyzes supermarket sales data to identify key trends in revenue, customer preferences, and profitability. The dataset contains transaction records from three (3) branches of a supermarket in three different cities, covering various product categories, pricing, sales volume, customer demographics and ratings.

The main objectives of this analysis were to:

* Understand sales performance across the different branches/cities and product line.
* Identify peak sales periods and customer purchasing trends.
* Analyze gross income and profit margins to determine the most and least profitable products.
* Provide actionable insights to improve business operations and customer satisfaction.

A dashboard was created to visualize critical metrics and key performance indicators (KPIs), including:

* **Total Sales Revenue, Gross Profit, Total Orders and Total Products Sold** – Measuring overall business performance.
* **Sales and Profit by Product Line and City** – Identifying the most profitable products and locations.
* **Sales Distribution by Gender, Customer Type and Payment Method** – Understanding customer demographics and payment preferences.
* **Peak Sales Hours and Daily/Monthly Sales Trends** – Highlighting when sales are highest.

**Key Findings:**

* **Best-Selling Product Line:** The **Food and beverages** product line was the best-selling category, generating a total sales revenue of $56,144 and a profit of $2,674.
* **Most Profitable City:** Among all cities, the city **Napyitaw** was themost profitable, generating a total profit of $5,265 from total sales of $110,569.
* **Peak Sales hours:** The peak sales hour was 19:00 (7pm), indicating that the highest number of transactions occurred in the early evening.
* **Sales by Customer Type:** Members contributed 50.85% of total sales, while normal customers accounted for 49.15%.
* **Payment Preference:** Customer payment preferences were almost evenly distributed among the three payment methods. Cash was the most used accounting for 34.74% of transactions, followed closely by E-wallet at 34.06% and credit cards at 31.20%. This indicates that cash-paying customers had higher average order values.
* **Customer Ratings:** The average rating was 7.0, suggesting a generally positive customer experience. However, there is room for improvement to enhance satisfaction and drive higher ratings.

This analysis provides valuable insights into this supermarket sales performance, customer behaviour, and profitability across different product lines and locations. Key findings revealed peak sales hours, top-performing branches and products lines, customer demographic trends, and payment preferences. These insights can help optimize product stocking and availability during peak hours, pricing strategies, and marketing efforts to enhance overall revenue and customer satisfaction.

1. **Data Description**

The dataset used for this analysis was sourced from Kaggle. It contains 1000 rows and 17 columns of supermarket sales transactions from three branches. It provides insights into customer purchases, payment methods, and product performance.

**Source**

The dataset was obtained from Kaggle and represents structured sales data collected from a supermarket’s point-of-sale system.

**Content**

This dataset includes a mix of categorical, numerical, and time-based data, allowing for in-depth analysis.

* **Number of Rows:** 1000
* **Number of Columns:** 17
* **Data Types:**
* **Categorical Data:** Branch, City, Customer Type, Gender, Product Line, Payment Method.
* **Numerical Data:** Unit Price, Quantity, Tax, Total, COGS, Gross Margin Percentage, Gross Income, Rating.
* **Date and Time Data:** Date, Time of Purchase.

**Column Descriptions**

Each column represents key aspects of a transaction:

* **Invoice ID:** Computer generated sales slip invoice identification number.
* **Branch:** Dataset contains three (3) branches of supercenter identified by A, B, and C.
* **City:** Location of supercenters. There are three cities.
* **Customer Type:** There are two customer types – **Members** for customers using member cards, and **Normal** for customers without member card.
* **Gender:** Gender type of customers, either male or female.
* **Product Line:** General item categorization groups – Electronic Accessories, Fashion Accessories, Food and Beverages, Health and Beauty, Home and Lifestyle, Sports and Travel.
* **Unit Price:** Price of each product in $.
* **Quantity:** Number of products purchased by customer.
* **Tax:** 5% tax fee for customer purchase.
* **Total:**  Total price including tax.
* **Date:** Date of purchase (from Jan. 2019 - Mar. 2019).
* **Time:** Purchase time(between 10am – 9pm).
* **Payment:** Payment used by customer for purchase. There are three (3) methods available – cash, credit card, and E-wallet.
* **COGS:** Cost of goods sold.
* **Gross Margin Percentage:** The percentage of profit relative to the sales price.
* **Gross Income:**  The total profit made on the transaction.
* **Rating:** Customer satisfaction rating on their overall shopping experience (on a scale of 1-10).

**Key Features**

The most important features of this dataset include:

* **Sales Trends:** Tracks total sales and revenue across different branches, cities and product lines.
* **Customer Demographics:** Analyzes shopping behaviour based on gender and membership status.
* **Peak Sales Hours:** Identifies the time of day when sales are highest.
* **Payment Preferences:** Examines which payment methods customers prefer.
* **Profitability Metrics:** Highlights gross income, profit margins, and cost of goods sold.
* **Customer Satisfaction:** Usesratings to assess service and overall satisfaction.

This structured dataset serves as the foundation for analyzing sales performance, customer behaviour, and profitability.

1. **Data Preprocessing**

**Initial Exploration**

The dataset was first imported into Microsoft Excel. Since it was in CSV format, it was converted into columns for easier manipulation. A duplicate of the original dataset was created to ensure that the cleaning process could be tracked and reversed if necessary.

Basic exploratory checks were conducted:

* **Missing Values:** TheFilter for Blanks option was used to check for blank cells, and none was found.
* **Duplicate Records:** The dataset was checked for duplicate entries using Conditional Formatting, but none were found.

**Data Cleaning**

To ensure accuracy and consistency in the dataset, the following cleaning steps were performed:

1. **Correcting Errors in Numerical Columns**

* Some numerical columns, such as Unit Price and COGS, had formatting issues where Excel flagged them as errors. These errors were indicated by small green markers in the corners of the cells, which typically signal that the values are stored as texts instead of numbers. To fix this, all values in these columns were converted to numeric format using the formatting tool.

**2. Fixing Inconsistencies in Tax and Cost Calculations**

* The **Tax 5%** column contained incorrectly placed decimal points and commas instead of decimal points. A new column was created to recalculate the tax using:

New Tax 5% = COGS ✕ 0.05

The old column was deleted after verification.

* The **COGS (Cost of Goods Sold)** column was recalculated to ensure accuracy:

COGS = Unit Price ✕ Quantity

The recalculated column replaced the original.

* The **Total Sales** column contained inconsistent values, misplaced decimal points, and comma separators. To correct this, a new column was generated and the total sales was recalculated using:

Total = COGS + Tax 5

The old column was deleted after confirming accuracy.

**3. Date and Time Formatting**

* The **Date** column was initially formatted as M/D/Y, but it needed to be in D/M/Y format. The Text to Columns option was used to adjust the format accordingly.
* To extract additional insights, two new columns were created:
* **Day of the Week:** Extracted using:

=TEXT (Date, “TTT”)

* **Month:** Extracted using:

=TEXT (Date, “MMM”)

* The **Time** column was formatted to 24-hour time format and validated to ensure all times fell between 10:00 and 21:00 using the formula:

=IF (AND (TIME >= TIME (10,0,0), A1 <= TIME (21,0,0)), “valid”, “invalid”)

All records were found to be valid.

* Additionally, an **Hour** column was created by extracting the hour from the Time column using:

=TEXT (TIME, “HH”)

This allowed for further analysis of peak sales hours.

**4. Correcting Gross Income and Gross Margin Percentage**

* The **Gross Income** column contained formatting issues with commas instead of decimal points. A new column was created by recalculating:

Gross Income = Total – COGS

* The **Gross Margin Percentage** column had similar formatting inconsistencies. It was recalculated using:

Gross Margin = (Gross Income/Total) ✕ 100

* The recalculated columns replaced the original ones for accuracy.

**5. Handling the Rating Column**

* The **Rating** column was reviewed and found to be consistent with no missing or incorrect values, so it was left unchanged.

**New Features**

To enhance the analysis, additional columns were created:

* **Sales Hour:** Extracted from the Time column to analyze peak transaction times.
* **Day of the Week:** Extracted from the Date column for sales trend analysis.
* **Month:** Extracted from the Date column to identify monthly trends.

1. **Exploratory Data Analysis (EDA)**

This section involves visualizing key trends, identifying patterns, and deriving insights from the supermarket sales data. The analysis focuses on sales performance, profitability, customer demographics, payment preferences, and time-based trends. An interactive dashboard was also created to allow for dynamic data exploration.

**Key Performance Indicators (KPIs)**

To summarize the overall performance of the supermarkets, the following key metrics were calculated:

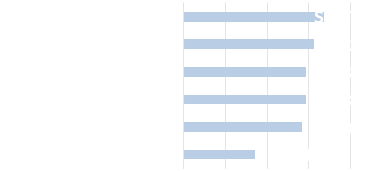
* **Total Sales Revenue:** $322,967
* **Gross Profit:** $15,379
* **Total Orders:** 1000
* **Total Products Sold:** 5510
* **Average Rating:** 7.0/10



**Sales and Profit Analysis**

**Sales and Profit by Product Line**

* **Food and Beverages** emerged as the highest-selling category, generating $56.1k in sales, followed closely by **Sports and Travel** ($55.1k).
* **Health and Beauty** had the lowest total sales contributing $49.2k.
* When analyzing profitability, **Food and Beverages** led again with a total profit of $2,674, followed closely by **Sports and Travel** ($2,625).
* The **Health and Beauty** category recorded the lowest profit at $2,343.



**Profit by Product line**



**Sales by Product line**

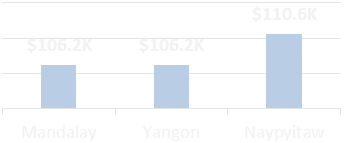
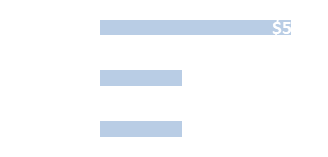
**Insights:**

* The sales and profit distribution suggests that while all product lines perform relatively well, **Food and Beverages** consistently drive both revenue and profit, making it a key focus for future promotions.
* The **Health and Beauty** category, despite having the lowest revenue, still contributes significantly. Targeted marketing could help boost its performance.
* Profitability across product lines is relatively balanced, indicating a well distributed revenue stream across categories.

**Sales and Profit by City**

* **Naypyitaw** recorded the highest sales, generating $110.6k in revenue while **Mandalay** and **Yangon** both had $106.2k.
* Profitability followed a similar trend, with **Naypyitaw** contributing $5,265 in profit, whereas **Mandalay** and **Yangon** each brought in $5,057.

**Profit by City**



**Sales by City**

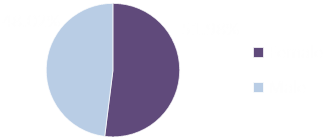
**Insights:**

* **Naypyitaw** outperforms the other cities in both sales and profitability. This suggests higher customer engagement or stronger purchasing power in that location.
* Further research could determine whether specific product preferences or branch-level promotions drive Naypyitaw’s stronger numbers.
* **Mandalay** and **Yangon** show equal sales and profits, indicating consistent business performance across these two cities.

**Customer Demographics**

**Gender-Based Sales Distribution**

* **Female** customers accounted for 51.98% of total sales ($167.9k) while **male** customers contributed 48.02% ($155.1k).



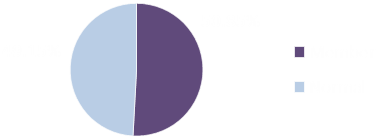
**Percentage of Total Sales by Gender**

**Insights:**

* This suggests a fairly balanced gender distribution, with a slight edge towards female shoppers.

**Sales by Customer Type**

* **Members** accounted for 50.85% of total sales ($164.2k), while **normal** customers contributed 49.15% ($158.7k).



**Percentage of Total Sales by Customer Type**

**Insights:**

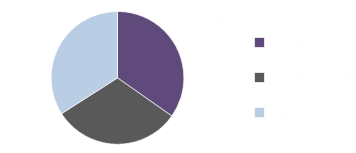
* Members and normal customers contribute almost equally to sales, indicating that membership benefits may need improvement to encourage more exclusive customer engagement.

**Payment Preferences**

The three payment methods were used almost evenly:

* **Cash:** 34.74% of total sales revenue ($112.2k).
* **E-wallet:** 34.06% ($109.9k).
* **Credit Card:** 31.2% ($101.0k).

However, analyzing transaction counts reveals that **E-wallet** was used the most. It was used 345 times, while **cash** was used 344 times, and **credit card** 311 times.



**Percentage of Total Sales by Payment Type**

**Insights:**

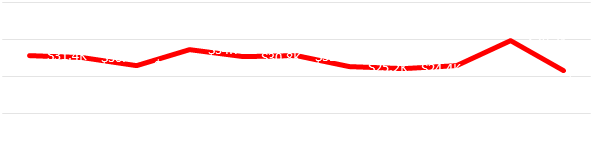
* **E-wallets** and **cash** are the preferred payment methods, each contributing nearly 35% of total sales.
* While **credit card** usage is lower, it still accounts for a significant portion of transactions, making it a viable payment method for future promotions.
* The three payment methods were used almost equally. This suggests that customers enjoy payment flexibility, making it essential to continue supporting all options.

**Sales Trends and Patterns**

**Peak Sales Hour**

* Sales peaked at 19:00, generating $39.7K, while 20:00 had the lowest sales at $23.0K.

**Peak Sales Hours**

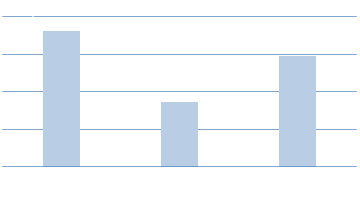


**Insights:**

* This chart shows a high customer influx at 19:00, tapering off later at night.
* The drop in sales at 20:00 suggests that customers may be finishing their shopping by then, making it less optimal for major promotions.

**Monthly Sales Trends**

* **January** had the highest total sales ($116.3K), followed by **March** ($109.5K).
* **February** recorded the lowest sales ($97.2K).



**Monthly Sales**

**Insights:**

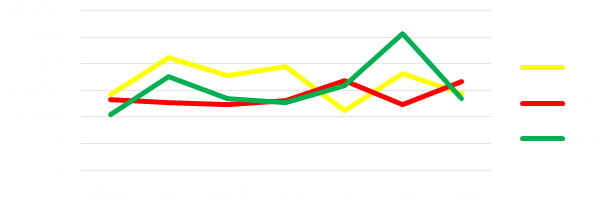
* **February’s** lower performance could be due to fewer shopping days in the month.

**Sales Distribution Across Days of the Week**

Sales varied unpredictably across weekdays in different months:

* **January:** Sales peaked on Tuesdays ($21,205) and were lowest on Fridays ($11,219).
* **February:** Wednesdays had the lowest sales ($12,406), while Fridays recorded the highest sales ($16,802).
* **March:** Mondays had the lowest sales ($10,421), while Saturdays recorded the highest sales ($25,771).

**Sales Distribution across Days of the Week**



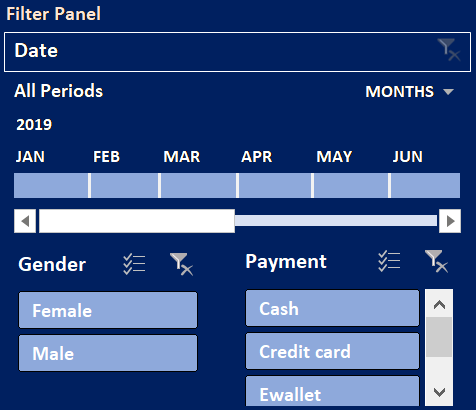
**Insights:**

* Sales patterns fluctuated across different months, with no single day consistently having the highest or lowest sales.
* The lack of a consistent pattern suggests that external factors may influence daily sales trends.

**Interactive Dashboard Features**

To facilitate data exploration, interactive filter panels were incorporated into the dashboard:

* **Gender filter** to analyze sales performance by male vs. female customers.
* **Payment method filter** to compare transaction distributions across Cash, Credit, and E-wallets.
* **A timeline filter** to allow users to assess sales trends over different time periods.



**Key Takeaways**

* **Food and Beverages** is the top-performing category in both sales and profit.
* **Naypyitaw** is the most profitable branch, outperforming Mandalay and Yangon.
* **E-wallet** transactions occur most frequently, but cash generates the highest revenue.
* Sales peak during evening hours (19:00).
* **January** sees the highest sales, while February experiences the lowest.

1. **Conclusion and Recommendations**

The supermarket sales analysis provided valuable insights into sales performance, customer behaviour, and payment trends. This section summarizes key findings, offers data-driven recommendations to enhance business performance, and highlights any limitations of the analysis.

**Key Findings**

1. **Top-Selling and Most Profitable Product Line:**

* **Food and Beverages** generated the highest sales ($56.1K) and the most profit ($2,674).
* **Health and Beauty** had the lowest sales ($49.2K) and the least profit ($2,343).

**Implication:** Supermarkets should ensure a steady stock of high-demand products and explore ways to boost sales of underperforming categories.

1. **Sales and Profit by City:**

* **Naypyitaw** generated the highest revenue ($110.6K) and profit ($5,625), while **Mandalay** and **Yangon** had lower but equal sales ($106.2K) and profits ($5,057).

**Implication:** Naypyitaw may have a larger or more engaged customer base. Further analysis into branch-specific factors (e.g. promotions, customer demographics) can help optimize sales strategies for Mandalay and Yangon.

1. **Customer Demographics:**

* **Female** customers accounted for 51.98% of total sales, while **male** customers contributed 48.02%.
* **Members** contributed slightly more (50.85%) than **normal** customers ($49.15%) to total sales.

**Implication:** The membership program is attracting a balanced mix of customers, but additional loyalty incentives could increase membership sign-ups and retention.

1. **Payment Trends:**

* **Cash** transactions generated the most revenue (34.74%), but **E-wallet** was used most frequently (345 times).

**Implications:** Customers are increasingly using digital payments. Supermarkets could capitalize on this trend by offering discounts for E-wallet transactions to encourage cashless payments.

1. **Time-Based Sales Patterns:**

* Sales peaked at 19:00 ($39.7K) and were lowest at 20:00 ($23.0K).
* Saturday recorded the highest weekly sales ($25,771) while Mondays saw the lowest ($10,421).

**Implication:**  Evening shopping hours are the most profitable, and weekends (especially Saturdays) attract the most customers. Businesses can introduce special promotions or discounts to drive traffic on slower days.

1. **Monthly Trends:**

* **January** had the highest sales ($116.3K), while February had the lowest ($97.2K).

**Implication:** Seasonal variations may affect sales. Investigating external factors (e.g. holidays, promotions) can help plan marketing campaigns to boost sales during slower months.

**Recommendations**

Based on these findings, the following actions are recommended to improve overall sales and profitability:

1. **Optimize Inventory Management:**

* Maintain higher stock levels for food and beverages due to their strong demand.
* Investigate why health and beauty products underperform and consider targeted promotions to boost sales.

1. **Enhance Sales in Mandalay and Yangon:**

* Conduct customer research to understand differences in customer behaviour across branches.
* Implement branch-specific promotions or localized marketing strategies to attract more customers in lower-performing cities.

1. **Leverage Digital Payment Trends:**

* Offer incentives such as cashback or discounts for E-wallet transactions to encourage digital payments.
* Ensure that all checkout points support seamless digital transactions to improve customer experience.

1. **Boost Off-Peak Sales:**

* Implement “Happy Hour” discounts from 18:00 – 20:00 to extend peak-hour sales.
* Introduce weekday or early evening promotions to drive sales during historically low-sales hour.

1. **Strengthen Customer loyalty Programs:**

* Enhance the benefits of the membership program by offering exclusive discounts, early access to sales or personalized offer.
* Encourage more customers to sign up for membership by promoting its benefits at checkout and via digital marketing.

1. **Plan Seasonal Promotions:**

* Since January saw the highest sales, analyze factors contributing to its success (e.g. holiday sales, new year promotions) and replicate them.
* Implement targeted marketing campaigns in February which had the lowest sales, to increase customer engagement during this period.

**Limitations of the Analysis**

While this analysis provided useful insights, certain limitations should be acknowledged:

1. **Limited Dataset Size:**

* The dataset contains only 1,000 transactions, which may not be fully representative of long-term sales trends.

**Recommendation:** Future analyses should incorporate larger and more recent datasets to enhance accuracy.

1. **Lack of External Factors Consideration:**

* The analysis did not account for external influences like economic conditions, holidays, weather patterns, or competitor activity that may affect sales.

**Recommendation:** Integrating external data sources would provide a more comprehensive understanding.

1. **No Customer Segmentation Beyond Gender and Membership Type:**

* The analysis did not explore age groups, income levels, or shopping habits, which could provide deeper insights into customer behaviour.

Recommendations: Collecting and analyzing customer demographics and purchasing patterns would allow for more targeted marketing strategies.

1. **Assumption of Data Accuracy Post-Cleaning:**

* While data cleaning ensured consistency, manual recalculations of columns (e.g. COGS, Gross Income) assume that initial errors were only formatting-related.

**Recommendations:** In future analysis, validating data accuracy with original transaction records would be ideal.

**Summary and Next Steps**

This supermarket sales analysis highlighted key areas of strength and improvement opportunities. Implementing the recommendations outlined above can help:

* Increase revenue by optimizing inventory and sales strategies.
* Improve customer satisfaction with personalized promotions and loyalty incentives.
* Enhance operational efficiency by aligning business strategies with customer behaviour trends.

For future analyses, expanding the dataset, incorporating external factors, and refining customer segmentation would provide even deeper insights into supermarket performance and customer preferences.