**Instructor:** Adeeb Alshakhs

**Subject:** Business Intelligence

**Group**: 13

**Date:** 3/12/2024


# Data Mining in Walmart

**Prepared by**


| | |
|---|---|
| Shahad Albannay | 221416390 |
| Shaima Alhajri | 221419858 |
| Marwoh Alshuaibi | 221419123 |
| Sukaina Altaroti | 221426510 |
| Sadeem Alharthi | 221429375 |

**Table of Contents**

## Introduction

Walmart is one of the world's largest traditional retailers, and one of the fastest growing and most dynamic e-commerce organization ،and the Walmart opened its first stores in 1962 and has continued to look for ways to bring technology to retail to provide a seamless shopping experience for customers ever since (Walmart, n.d.).

Walmart has more than 20,000 stores in 28 countries and with the rapid growth in the size of operations and the increasing complexity of the market, data has become one of the most important assets that companies rely on to achieve success, so Walmart has moved towards adopting business intelligence systems, as Walmart has used these systems for a long time, and Walmart has recently enhanced its technological infrastructure by creating the largest private cloud in the world, which has revolutionized the decision-making process, and has reduced complex processes from weeks to minutes (Yadav, 2024).

Walmart is one of the most noticeable companies that have one of the best examples in the field of business intelligence because of the vast quantity of data it owns, as Walmart makes use of business analytics in inventory management by analyzing historical customer data, sales data, current market trends, and in addition to customer preferences and seasonal changes. Furthermore, data analysis allows for the projection of the need to stock the warehouse with the appropriate amount and, therefore, the avoidance of the situation of having too many stock overs or not having enough stock (Yadav, 2024).

The supply chain of Walmart is indeed a complicated one, since it contains multiple distribution centers, stores, and suppliers. The inefficiencies in supply chain management that could have arisen due to this can imply shipping delays and higher costs, which might cause customer dissatisfaction. In such a case, Walmart would be making use of B.I. to identify the weaknesses and thus, can enhance and see the whole picture of the supply chain by conducting analyses of performance data for suppliers, transportation methods, and inventory levels. Using this data, Walmart can identify associations between products and design marketing campaigns that target specific customer segments (Yadav, 2024).

Understanding customer preferences and behavior is essential for Walmart to deliver personalized experiences and increase sales through the use of advanced analytics

techniques, which give Walmart insights into shopping patterns, customer preferences, and demographic trends (Yadav, 2024).

With this data, Walmart can identify relationships between specific products and design marketing campaigns targeted to specific customer segments (Yadav, 2024).

In this research, we will use a dataset containing detailed information about Walmart's transactions, which includes several important areas that allow for a detailed analysis of business processes and trends using techniques such as BI and Data Mining, with the aim of exploring patterns and trends that have contributed to improving Walmart's operations in how to manage inventory more accurately, improve resource allocation, and achieve more effective marketing strategies, We will also show how data can contribute to strategic decisions that lead to increased customer satisfaction and

## The combined impact of data mining and Business Intelligence (BI) applications

Walmart started using data mining and Business Intelligence (BI) applications and this was transformative, significantly in shifting their operational efficiency, customer experience, and competitive edge.

- **Revolutionized Supply Chain Management**

BI Applications: The ability of tools like the Retail Link system that help suppliers to access real-time sales data and makes them responsible for their management.

Data Mining: Enable pattern recognition in sales to determine restocking needs.

Impact: it helped in inventory management, avoiding stockouts or overstocking, reducing logistics costs, and faster replenishment cycles. These practices led to ensuring product availability and gaining insight into sales. (Costa, 2024)

- **Enhanced Customer Insights**

BI Applications: Transform raw data into actionable information by providing meaningful dashboards and visualizations to monitor and understand customer behavior.

Data Mining: Detect customer's behavior and purchasing patterns to focus on what customers prefer, and benefit by using it for advertising purposes.

Impact: it helped in improving product placement, marketing strategies, and assets in a personalized shopping experience for each customer and used targeted promotions. These practices led to an increase in customer satisfaction and loyalty. (Costa, 2024)

- **Data-Driven Decision-Making**

BI Applications: Executives and managers depend on it as a supporting point, as it provides actionable insights through reporting tools and predictive analytics that will help in supporting their ideas or plans, and to reduce uncertainty.

Data Mining: Assets in discovering hidden trends and anomalies in massive datasets.

Impact: it helped in making faster and more accurate strategic decisions, gaining adaptability to market trends and customer preferences. These practices have led to a reduction in reliance on intuition or guesswork. (Costa, 2024)

## BI Implementation at Walmart

Walmart's participation in Business intelligence applications was documented for the first time in the early 1990s. This strategy was meant to enable the company to gain a competitive advantage through the application of data analytics in its activities. (Wailgum, 2007)

1. Early 1990s: Before the suppliers came on board, Walmart had initiated a Retail Link system that gave them real-time access to sales records of the chain. This innovation contributed towards lowering costs and enhancing supply chain management. (Projectpro, 2023)

2. 2000s: In this decade, Walmart makes its BI development comprehensive by integrating data warehousing and contemporary analytics tools for the study of customer behavior, inventory, and demand forecasting. (Projectpro, 2023)

3. 2010s: Walmart in this period appeared to have largely embraced big data and even integrated machine learning and AI-based analyses in its BI processes. It was also reported that Walmart used Hadoop, a big data platform, to manage and analyze its vast data. Currently, Walmart is recognized for combining, integrating, synchronizing, coordinating, and optimizing critical and complex business processes with advanced Business intelligence software applications, providing great service to its customers and dominating the rivalry in retail business. (Musani, 2023)

## Data Collection and Exploration

The attached database contains important and detailed data about the products, sales and performance of the company, which helps in providing an opportunity to analyze the factors that affect its success. Analyzing the data of Walmart in the e-commerce and retail sector can provide valuable insights to enhance the marketing strategy, improve operations and manage the supply chain effectively.

## The main benefits of analyzing this data include

• **Operations and inventory management:** Improve inventory management and reduce shortages and surpluses by discovering and evaluating areas related to quantities and prices.

• **Financial performance analysis:** Understand trends and improve the accuracy of profit management by studying gross income and return on sales.

• **Customers experience improvement:** Improve marketing strategies, pricing and after-sales services by analyzing and understanding customer classifications and preferences.

• **Trend analysis:** Studying times and days of purchase helps in identifying time patterns.

### Data Description

The following database contains several columns representing various information as follows:

• **Invoice ID:** Invoice number (categorical data).

• **Branch:** Sales branch (categorical data).

• **City:** The city in which the branch is located (categorical data).

• **Customer Type:** Customer type (categorical data).

• **Gender:** Customer gender (categorical data).

• **Product Line:** Product category (categorical data).

• **Unit Price:** Unit price in dollars (quantitative data).

• **Quantity:** Quantity sold (quantitative data).

• **VAT:** Added tax (quantitative data).

• **Total:** Total amount paid in dollars (quantitative data).

• **Date:** Transaction date (categorical data).

• **Time:** Transaction time (categorical data).

• **Payment Method:** Payment method (categorical data).

• **COGS:** Cost of goods sold (quantitative data).

• **Gross Margin Percentage:** Gross profit margin percentage (quantitative data).

• **Gross Income:** Gross income in dollars (quantitative data).

• **Rating:** Customer's rating of the service (quantitative data).

• **Time of Day:** The time of the transaction (e.g. morning, afternoon and evening) (categorical data).

• **Day Name:** Name of the day (categorical data).

• **Month Name:** Name of the month (categorical data).

To determine the appropriate type of analysis for each column, the data was classified into categorical and quantitative.

• **Categorical data:** Invoice ID, Branch, City, Customer Type, Gender, Product Line, Date, Time, Payment Method, Time of Day, Day Name, Month Name.

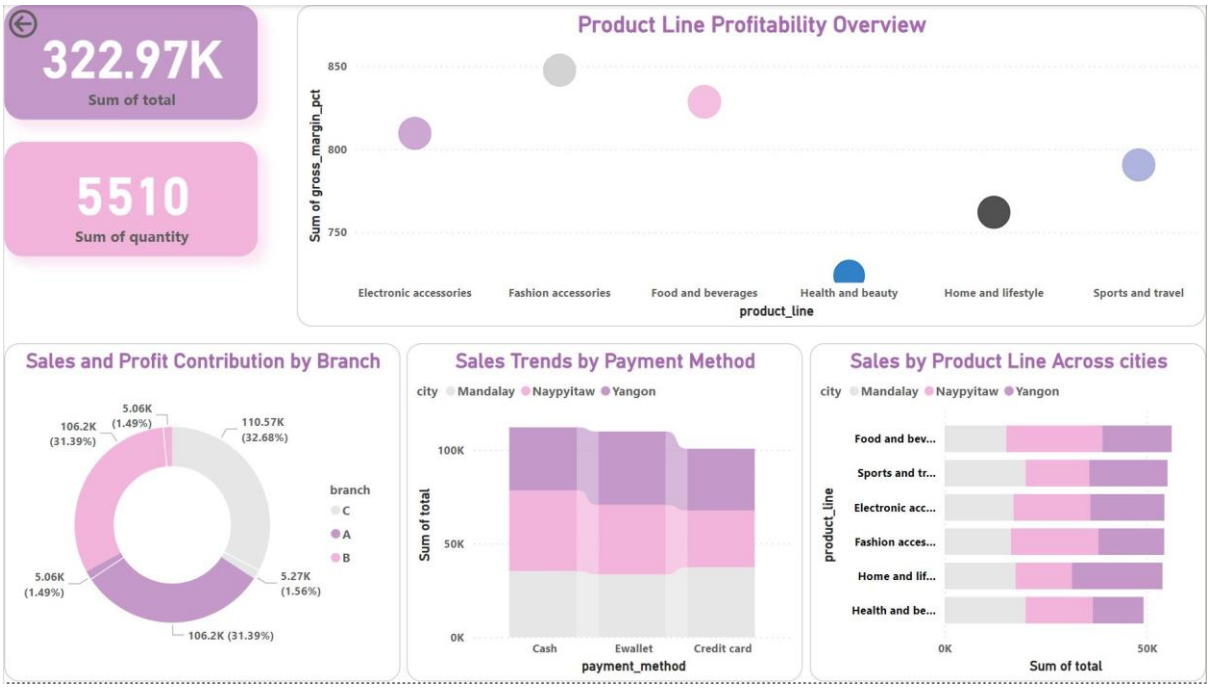
• **Quantitative data:** Unit Price, Quantity, VAT, Total, COGS, Gross Margin Percentage, Gross Income, Rating.

## Data Cleaning

The data cleaning process revealed no missing values or outliers has been found in the dataset, ensuring the dataset's integrity. The dataset also shows that there are no incompatible data types, indicating that the dataset is in a good shape already.

Therefore, the dataset is ready for analysis and visualization, and there will be no issue in the upcoming calculations or visualization. Overall, a clean dataset will provide a precise insight in the end of the analysis process.

## Data Visualization



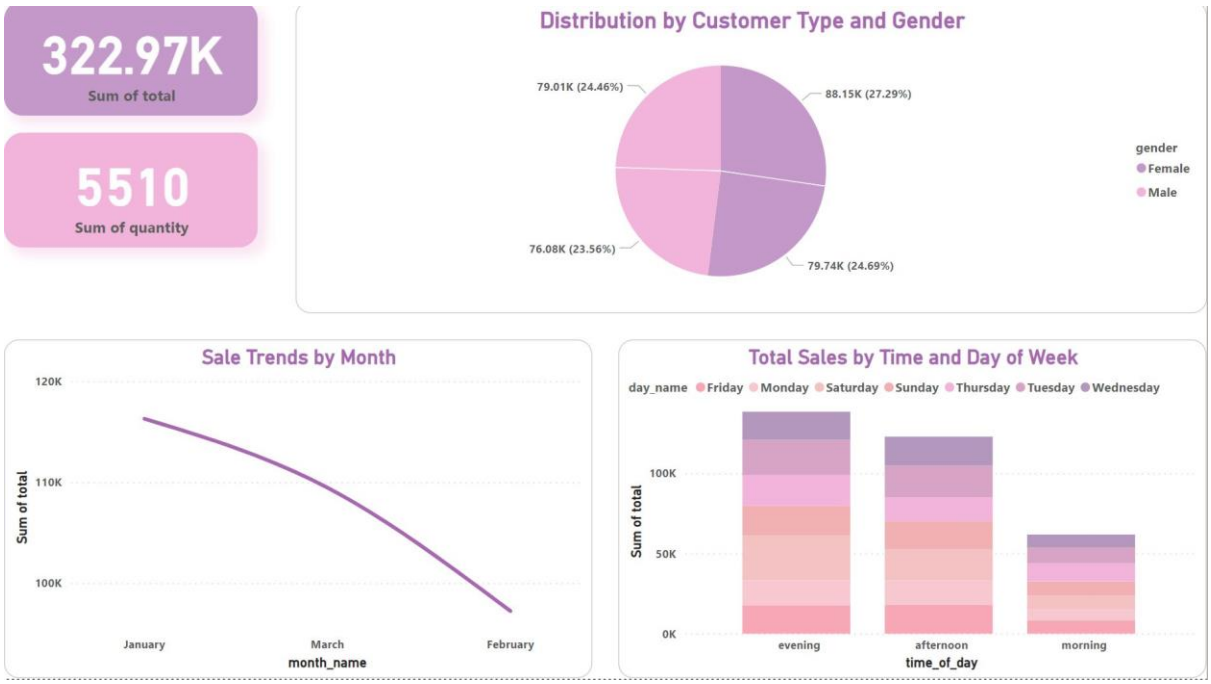

Data was visualized using Microsoft Power BI. The total sales volume is 322.97K, and the quantity sold is 5510 units across all product lines and cities.

On the scatter plot diagram, we have six product lines in this dataset, which are electronic accessories, fashion accessories, food and beverages, health and beauty, home and lifestyle, and sports and travel. On the x-axis, there are all the product lines, and on the y-axis, there is the sum of the gross margin percentage. In this chart, the higher the bubble, the more profitable the category is. The most profitable three categories are electronic accessories, fashion accessories, and food and beverages; they are relatively close in profitability; however, fashion accessories come in first place as most product lines that contributed significantly to the profit with gross income of USD 2586.00 and 902 quantity sold, which makes a gross margin percentage of 847.62, exceeding other categories. We must know that higher sales do not mean more profit; for example, product lines such as home and lifestyle may have a high sales contribution but are less profitable than fashion accessories or any other highly profitable category.

As for the donut chart, it represents the sales and profit by branch. Both branches A and B accounted for 31.39% of sales and 1.49% of profit. On the other hand, branch C contributed the highest sales and profit volume by making 32.68% of the sales and a gross margin percentage of 1.56%. Even though branch C had higher sales and profit volumes, branch A and branch B did not have significant low sales and profit volumes.

Regarding the stacked column chart, it represents how customers in each city in the dataset usually pay. We have three payment methods: cash, e-wallet, and credit card. Interestingly, the preferred payment method differed across the three cities. In Mandalay, its population mostly pays using credit cards, with a sales volume of over USD 37K. Moreover, the second preferred payment method in Mandalay was e-wallet, and the least favorite payment method was cash. As for Naypyitaw, the most preferred payment method was cash with a sales volume of over USD 43K, followed by e-wallet and credit card ranking last. In Yangon, e-wallet dominated as the preferred payment method, sales surpassing 39K, while cash took second place and credit cards came in last.

In terms of the stacked bar chart, it shows sales by product line across cities. There are six product lines, which are electronic accessories, fashion accessories, food and beverages, health and beauty, home and lifestyle, and sports and travel. As for the electronic accessories' category, we can see that Naypyitaw contributed the most sales volume with over USD 18.9K, followed by slightly lower sales volume of USD 18.4K in Yangon and USD 17K sales volume in Mandalay. Regarding the fashion accessories product line, the highest sales volume was generated in Naypyitaw, making up sales of USD 21.5K. Moreover, both Yangon and Mandalay generated sales volume between USD 16.3K and USD 16.4K. Concerning the food and beverages line, Naypyitaw generated the highest sales in this category with sales of over USD 23.76K, followed by Yangon, and Mandalay trailing behind in last place. As for the health and beauty line, Mandalay's population outspent Naypyitaw's population and Yangon's population, with sales volume exceeding USD 19.9K. Regarding the home and lifestyle category, Yangon's residents made higher purchases than those in Naypyitaw and Mandalay, with total sales surpassing USD 22.4K. As for sports and travel product line, both Mandalay's population and Yangon's population had relatively close sales volumes between USD 19.9K and USD 19.3K; however, Naypyitaw' had moderately lower total sales of USD 15.7K.




As for the pie chart, it shows the distribution by customer type and gender. For each gender, we have to type of customers: normal and member. In general, females made more purchases than men, with a percentage of 51.98% for both types, normal and member. Moreover, female members contributed more to the total sales, making up 27.29%, than normal females, which generated sales with a proportion of 24.69%. On the other hand, males of both types accounted for 48.02%. Interestingly, normal males generated more sales of USD 79.01K, just under a quarter of total sales. Besides this, males of type members had a slightly lower percentage of 23.56%.

Regarding the line chart, it clarifies sales trends by month. As shown in the figure, January had the highest sales with around USD 116K, followed by March, which had a moderate decline, close to USD 109K. However, before March, in February 2019, Walmart experienced a steep decline, nearing USD 100K, before moderately increasing to approximately USD 116K.

The last figure, which shows total sales by time and day of week, clarifies that people made regular purchases during the days of the week without any spikes on specific days. Despite that, evening had the highest sales during the whole week, with accumulated sales of USD 138.37K. Besides evening, afternoon also significantly

contributes to the sales but less than evening time. Moreover, morning shows the least sales volume, maybe due to customers not being able to do some activities.

## Insights

Overall, higher sales volume does not equal high profitability. So, when we take the fashion accessories product line as an example, this category was the highest profitable although there are some other product lines, such as sports and travel, that generated higher sales than fashion accessories. Walmart should focus on high-profitability product lines rather than focusing on product lines that generate high sales volumes. Concentrating on high-profitability product lines leads to improved overall profitability.

It is noticeable that Naypyitaw's residents have purchasing power due to leading sales volumes in three product lines: electronic accessories, fashion accessories, and food and beverages. Walmart should focus on Naypyitaw for several reasons:

- Because it can generate higher revenue in a short time.
- Cities that have strong purchasing power are often willing to pay for high quality or premium products, which leads to have better profit margin.
- Strengthen customer relationship by offering the preferred product lines for Naypyitaw's residents, which Walmart can earn and build customer's loyalty.

Walmart should concentrate on applying promotions and campaigns, especially to females, to maximize sales during evening and afternoon times. Females are more affected to the marketing campaigns than men. Therefore, Walmart can increase sales and profit margin.

Walmart experienced a significant decline from January to March in sales, which implies there is something wrong with the marketing campaigns, customer engagement, and seasonality. January experienced sales peak probably due to the holiday season (e.g. Christmas and New Year) and the increasing marketing efforts that most companies put in during the holiday season. So, Walmart should improve promotions and campaigns even after the holiday season to boost overall profitability.

Branch C had the highest sale volumes and gross income percentage, yet the other two branches A and B were not significantly lower than branch C, so Walmart may need to use some strategies to improve sale volumes and profit such as:

- Pricing Strategy.
- Forecasting.
- Revenue Management.
- Reviewing the performance of each branch's employees to determine if there is an issue regarding customer service.

## References


- Walmart. (n.d.). About us. Walmart Corporate. Retrieved November 23, 2024, from https://corporate.walmart.com/about.
- Yadav, A. (2024, February 15). Leading Global Retailers' Strategies for Business Intelligence: Case Study Highlights: Short case studies. FabricHQ. Retrieved November 23, 2024,, from https://blog.fabrichq.ai/this-is-how-walmart-amazon-and-ikea-lead-with-business-intelligence-short-case-studies-418d10ad348d.
- Costa, I. (2024, November 27). *Optimizing business intelligence for strategic advantage*. Kaizen Institute Consulting Group. https://kaizen.com/insights/business-intelligence-strategic-advantage/
- Wailgum, T. (2007, October 17). 45 Years of Wal-Mart History: A Technology Timeline. CIO. https://www.cio.com/article/274537/infrastructure-45-years-of-wal-mart-history-a-technology-time-line.html
- Projectpro. (2023, October 11). How big data analysis helped increase walmart's sales turnover? ProjectPro. https://www.projectpro.io/article/how-big-data-analysis-helped-increase-walmarts-sales-turnover/109

- Musani, P. (2023, October 25). Decking the aisles with data: How Walmart's AI-powered inventory system brightens the holidays. Walmart Global Tech. https://tech.walmart.com/content/walmart-global-tech/en_us/blog/post/walmarts-ai-powered-inventory-system-brightens-the-holidays.html