



Sales Forecasting for Walmart

Unlocking Retail Success: Analysing Walmart's Sales Data for Strategic Insights & Enhance Customer Experience

Presented By - Binoy Patra



CONTENTS

01

Introduction

Overview of the project and its significance in leveraging data for strategic decision-making.

02

Objective

Goals aimed at optimizing marketing strategies, customer satisfaction, and sales efficiency.

03

Dataset Overview

Summary of the dataset's key features and components.

04

Analysis List

Key areas of analysis including product, sales, and customer insights.

05

Approach Used

Description of the methodologies and tools employed in the analysis.



CONTENTS

06 Business Questions to Answer
Specific questions targeted to uncover critical business insights.

07 Overall Insights
Key findings from the data analysis.

08 Recommendations
Strategic suggestions based on the analysis to drive business improvements.



01. Introduction



In the contemporary retail landscape, data-driven decision-making is pivotal for maintaining competitive advantage. This analysis of Walmart's sales data employs Structured Query Language (SQL) to unearth key insights that can inform strategic decisions. By dissecting various dimensions such as customer types, product lines, and sales performance across different times and days, this report aims to provide a comprehensive understanding of consumer behavior and sales dynamics.

02. Objective



The objective is to identify trends and patterns that can help optimize marketing strategies, improve customer satisfaction, and boost overall sales efficiency. Through rigorous querying and analysis of Walmart's extensive sales records, we seek to answer critical business questions and support the company's operational and strategic goals.

03. Dataset Overview



- ❖ *The dataset was obtained from the Kaggle Walmart Sales Forecasting Competition.*
- ❖ *This dataset contains sales transactions from a three different branches of Walmart,*
- ❖ *respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows.*

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product sold	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The total amount paid	DECIMAL(10, 2)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

04. Analysis List



1. Product Analysis

Conduct analysis on the data to understand the different product lines, the products lines performing best and the product lines that need to be improved.

2. Sales Analysis

This analysis aims to answer the question of the sales trends of product. The result of this can help use measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

3. Customer Analysis

This analysis aims to uncover the different customers segments, purchase trends and the profitability of each customer segment.

05. Approach Used



1. Data Wrangling:

This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace, missing or NULL values.

- 1. Build a database*
- 2. Create table and insert the data.*
- 3. Select columns with null values in them. There are no null values in our database as in creating the tables, we set NOT NULL for each field, hence null values are filtered out.*

2. Feature Engineering:

This will help use generate some new columns from existing ones.

- 1. Add a new column named time_of_day to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.*
- 2. Add a new column named day_name that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.*
- 3. Add a new column named month_name that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.*

3. Exploratory Data Analysis (EDA):

Exploratory data analysis is done to answer the listed questions and aims of this project.

06. Business Questions To Answer (Generic Question)



- 1. How many unique cities does the data have?*
- 2. In which city is each branch?*
- 3. What is the total revenue?*
- 4. What is the total CoG?*

Summary for Generic Question



Summary for Generic Questions:

- 1. The dataset reveals that Walmart operates in three unique cities: Yangon, Naypyitaw, and Mandalay.*
- 2. Each city hosts a specific branch, with Branch A located in Yangon, Branch B in Mandalay, and Branch C in Naypyitaw.*
- 3. The analysis shows that the total revenue generated across all branches amounts to \$320,886.39, while the total Cost of Goods Sold (CoG) is \$305,606.09.*

These figures provide a foundational understanding of the geographic distribution of operations and the overall financial performance of the company.

06. Business Questions To Answer (Product Analysis)



1. How many unique product lines does the data have?
2. What is the most common payment method?
3. What is the most selling product line?
4. What is the total revenue by month?
5. What month had the largest COGS?
6. What product line had the largest revenue?
5. What is the city with the largest revenue?
6. What product line had the largest VAT?
7. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales.
8. Which branch sold more products than average product sold?
9. What is the most common product line by gender?
12. What is the average rating of each product line?
13. What is the Total count of rating of each product line?

Summary for Product Analysis



Summary for Product Analysis

1. The dataset includes **six** unique product lines: **Food and Beverages, Health and Beauty, Sports and Travel, Fashion Accessories, Home and Lifestyle, and Electronic Accessories.**
2. The most common payment method is **cash**, with **344 transactions**, followed closely by **e-wallets and credit cards.**
3. Among the product lines, **Electronic Accessories** lead in quantity sold with 961 units, followed by **Food and Beverages** at 952 units.
4. Monthly revenue peaks in January with \$116,291.87, while February records the highest Cost of Goods Sold (CoGS) at \$91,168.93.
5. **Food and Beverages** generate the highest revenue at \$56,144.84, and **Naypyitaw (Branch C)** contributes the most to overall revenue, totaling \$110,490.78.
6. The product line with the largest VAT is **Home and Lifestyle**, at 16.03%.
7. In terms of sales performance, **Health and Beauty, Sports and Travel, Home and Lifestyle, and Electronic Accessories** are deemed "Good" for having above-average sales.
8. . Branches A and C have sold more products than the average, with Branch A selling 1,849 products and Branch C selling 1,828 products.
9. Gender-wise, **Fashion Accessories** are the most common product line among females, while **Health and Beauty** is most popular among males.
10. The average rating across product lines ranges from 6.84 to 7.11, with **Fashion Accessories** receiving the highest total count of ratings at 178.

06. Business Questions To Answer (Sales Analysis)



- 1. Number of sales made in each time of the day per weekday.*
- 2. Which of the customer types brings the most revenue?*
- 3. Which city has the largest tax percent/ VAT (Value Added Tax)?*
- 4. Which customer type pays the most in VAT?*

Summary for Sales Analysis



Summary for Sales Analysis:

- 1. Sales distribution throughout the day shows that the evening period sees the highest sales with 58 transactions, followed by the afternoon with 52 transactions, and the morning with 22 transactions.*
- 2. Customer type analysis reveals that Members bring in the most revenue, contributing \$163,625.10 compared to Normal customers' \$157,261.29.*
- 3. In terms of VAT, Naypyitaw has the highest average tax percent at 16.09%, followed by Mandalay at 15.13%, and Yangon at 14.87%.*
- 4. When examining VAT contributions by customer type, Members again lead, paying a total VAT of 15.61, slightly more than Normal customers, who pay 15.10. These insights highlight the significance of targeting evening sales periods and focusing on Member customers for revenue and VAT optimization.*

06. Business Questions To Answer (Customer Analysis)



- 1. How many unique customer types does the data have?*
- 2. How many unique payment methods does the data have?*
- 3. What is the most common customer type?*
- 4. Which customer type buys the most?*
- 5. What is the gender of most of the customers?*
- 6. What is the gender distribution per branch?*
- 7. Which time of the day do customers give most ratings?*
- 8. Which time of the day do customers give most ratings per branch?*
- 9. Which day of the week has the best avg ratings?*
- 10. Which day of the week has the best average ratings per branch?*

Summary for Customer Analysis)



Summary for Customer Analysis:

- 1. The data reveals two unique customer types: Normal and Member, with a nearly equal distribution of 499 Members and 496 Normal customers. Among these, Members are slightly more prevalent.*
- 2. The gender distribution is also quite balanced, with 498 male customers and 497 female customers. Analyzing the gender distribution per branch, Branch A has the most male customers (179), while Branch C has the most female customers (177).*
- 3. Customers tend to provide ratings most frequently in the afternoon, with an average rating of 7.02, followed by the morning at 6.94 and the evening at 6.91. Branch-specific analysis shows that Branch A receives the highest ratings in the afternoon (7.19), Branch B has more consistent ratings throughout the day, and Branch C sees the highest evening ratings (7.10).*
- 4. Regarding weekly ratings, Monday has the best average rating of 7.13, indicating higher customer satisfaction at the start of the week. This is followed by Friday (7.06) and Tuesday (7.00), with Wednesday having the lowest average rating (6.76). These insights can guide targeted strategies to enhance customer experiences based on gender, branch, and timing preferences.*

06 Overall Insights



The analysis of Walmart's sales data offers several key insights:

1. Geographic and Branch Insights:

- *There are three cities covered: Yangon, Naypyitaw, and Mandalay.*
- *Branch A is in Yangon, Branch B is in Mandalay, and Branch C is in Naypyitaw.*
- *Naypyitaw (Branch C) generates the highest revenue.*

2. Revenue and Costs:

- *The total revenue across all branches is \$320,886.39.*
- *The total Cost of Goods Sold (CoG) is \$305,606.09.*

3. Product Analysis:

- *There are six unique product lines, with "Electronic accessories" being the top seller by quantity.*
- *"Food and beverages" is the product line with the highest revenue.*
- *Cash is the most common payment method.*
- *January recorded the highest revenue and CoG among the months analyzed.*

06 Overall Insights



4. Sales Analysis:

- The evening time sees the highest sales, followed by the afternoon and morning.
- Members contribute slightly more to revenue than Normal customers.
- Naypyitaw has the highest average tax percentage.

5. Customer Insights:

- The customer base is evenly split between Normal and Member types.
- Gender distribution is almost equal, with slightly more male customers.
- Afternoon is the time of day when customers give the highest ratings, with Monday being the day with the best average ratings.

6. Product Line Performance:

- The top three product lines by revenue are "Food and beverages," "Fashion accessories," and "Sports and travel."
- "Health and beauty" shows the highest VAT, followed by "Sports and travel."

These insights can help Walmart optimize its operations, tailor marketing strategies, and improve customer satisfaction and sales efficiency across different branches and product lines.

07. Recommendations



Based on the analysis of Walmart's sales data, the following recommendations are proposed to optimize operations, enhance customer satisfaction, and boost sales:

1. Optimize Inventory for High-Performing Product Lines:

- *Electronic accessories and Food and beverages are the top-performing product lines. Ensure these items are always well-stocked to meet demand.*
- *Focus marketing and promotional efforts on these categories to capitalize on their popularity.*

2. Enhance Marketing for Underperforming Categories:

- *Fashion accessories and Health and beauty have lower sales compared to other categories. Develop targeted marketing campaigns to boost sales in these areas.*
- *Consider customer feedback and trends to adjust the product mix within these categories.*

3. Leverage High Revenue Periods:

- *January has been identified as the month with the highest revenue. Plan special promotions and sales events during this period to maximize sales.*
- *Analyze reasons behind the January peak and apply similar strategies to other months.*

4. Tailor Customer Experience:

- *Evening is the peak time for sales. Ensure sufficient staffing and inventory during these hours to provide excellent customer service.*
- *Enhance the shopping experience during evenings with in-store events, discounts, or special offers.*

07. Recommendations



5. Focus on High Revenue Branches:

- Naypyitaw (Branch C) generates the highest revenue. Consider using successful strategies from this branch as a model for others.
- Conduct a detailed analysis of customer preferences and behaviors in Naypyitaw to replicate success in other cities.

6. Improve Payment Method Options:

- Cash is the most common payment method, but a significant number of customers use E-wallets and credit cards. Ensure all payment systems are quick, secure, and user-friendly to enhance customer convenience.

7. Enhance Customer Loyalty Programs:

- Members contribute slightly more to revenue than normal customers. Enhance loyalty programs to encourage more customers to become members.
- Offer exclusive discounts, early access to sales, and special promotions to members to increase retention and attract new sign-ups.

8. Utilize Customer Feedback for Improvements:

- Afternoon has the highest customer ratings, followed by the morning and evening. Investigate the reasons for higher satisfaction during these periods and apply similar strategies to other times.
- Regularly collect and analyze customer feedback to identify areas for improvement and implement changes promptly.

07. Recommendations



9. Capitalize on High-Rating Days:

- Monday has the best average ratings. Consider extending successful Monday strategies to other days to improve overall customer satisfaction.

10. Monitor and Adjust VAT Strategies:

- Naypyitaw has the highest average tax percentage. Monitor the impact of VAT on sales and adjust pricing strategies if necessary to maintain competitiveness without sacrificing profitability.

By implementing these recommendations, Walmart can optimize its operations, improve customer satisfaction, and drive higher sales across its branches.

THANK YOU

- LinkedIn: [Binoy Patra](#)
GitHub: [Binoy Patra](#)

