Walmart Sales Data Analysis

This project explores Walmart's sales data to gain insights into top-performing branches and products, analyse sales trends across departments, and understand customer behaviour. The objective is to identify key factors influencing sales performance and propose strategies for improving and optimizing sales outcomes.

The dataset, originally from a Kaggle competition, contains historical sales records from 45 Walmart stores across various regions. Each store includes multiple departments, and the challenge involves forecasting departmental sales while considering the effects of holiday markdown events. These markdowns significantly impact sales patterns, making it crucial to determine which departments are affected and to what extent.

Through detailed analysis, this project aims to:

- **Identify top-performing stores and departments.**
- **Uncover seasonal and product-based sales trends.**
- **Understand the influence of promotional events and holidays.**
- **Develop data-driven insights to optimize sales strategies.**

About the Data

The dataset was obtained from the Kaggle Walmart Sales Forecasting Competition.

It contains sales transactions from three different Walmart branches located in Mandalay, Yangon, and Naypyitaw.

Dataset Overview:

- Rows: 1000
- Columns: 17

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	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)

Analysis Overview

1. Product Analysis

Understand different product lines.

Identify best-performing and underperforming product lines.

Suggest improvements for weaker product categories.

2. Sales Analysis

Analyse sales trends across months, cities, and departments.

Measure the effectiveness of sales strategies.

Recommend modifications to enhance future sales.

3. Customer Analysis

Identify different customer segments and their buying behaviour.

Determine profitability per segment.

Understand customer demographics and preferences.

Approach Used

1. Data Wrangling

Inspect data for NULL or missing values.

Replace or handle missing values appropriately.

Create and populate database tables with the dataset.

Ensure all fields are defined with NOT NULL constraints to avoid missing data.

2. Feature Engineering

New features were generated from existing data to provide deeper insights:

time_of_day – Categorizes sales into Morning, Afternoon, and Evening.

day_name – Extracts the weekday (Mon–Fri) of each transaction to identify busy days.

month_name – Extracts the month (Jan–Dec) to analyse monthly sales and profit trends.

3. Exploratory Data Analysis (EDA)

EDA was performed to answer the business questions and objectives listed below.

Business Questions

Generic Questions

* How many unique cities does the data have?

city
Yangon
Naypyitaw
Mandalay

* In which city is each branch located?

city	branch
Yangon	A
Naypyitaw	C
Mandalay	B

Product Analysis

* How many unique product lines exist?

product_line
Food and beverages
Health and beauty
Sports and travel
Fashion accessories
Home and lifestyle
Electronic accessories

* What is the most common payment method?

payment_method	cnt
Cash	344
Ewallet	342
Credit card	309

* What is the best-selling product line?

product_line	cnt
Fashion accessories	178
Food and beverages	174
Electronic accessories	169
Sports and travel	163
Home and lifestyle	160
Health and beauty	151

* What is the total revenue by month?

MONTH	total_revenue
January	116291.8680
March	108867.1500
February	95727.3765

* Which month had the largest COGS?

month	cogs
January	110754.16
March	103683.00
February	91168.93

* Which product line had the highest revenue?

product_line	total_revenue
Food and beverages	56144.8440
Fashion accessories	54305.8950
Sports and travel	53936.1270
Home and lifestyle	53861.9130
Electronic accessories	53783.2365
Health and beauty	48854.3790

* Which city generated the highest revenue?

branch	city	total_revenue
C	Naypyitaw	110490.7755
A	Yangon	105861.0105
B	Mandalay	104534.6085

* Which product line had the largest VAT?

product_line	avg_tax
Home and lifestyle	16.03033124
Sports and travel	15.75697549
Health and beauty	15.40661591
Food and beverages	15.36531029
Electronic accessories	15.15447632
Fashion accessories	14.52806181

* Which branch sold more products than the average quantity sold?

branch	qty
A	1849
C	1828
B	1795

* What is the most common product line by gender?

gender	product_line	total_cnt
Female	Fashion accessories	96
Female	Food and beverages	90
Male	Health and beauty	88
Female	Sports and travel	86
Male	Electronic accessories	86
Male	Food and beverages	84
Female	Electronic accessories	83
Male	Fashion accessories	82
Male	Home and lifestyle	81
Female	Home and lifestyle	79
Male	Sports and travel	77
Female	Health and beauty	63

* What is the average rating of each product line?

product_line	avg_rating
Food and beverages	7.11322
Fashion accessories	7.02921
Health and beauty	6.98344
Electronic accessories	6.90651
Sports and travel	6.85951
Home and lifestyle	6.83750

Sales Analysis

* Number of sales made during each time of the day per weekday.

time_of_day	total_sales
Evening	58
Afternoon	52
Morning	22

* Which customer type generates the most revenue?

customer_type	total_rev
Member	163625.1015
Normal	157261.2930

* Which city has the highest VAT percentage?

city	tax_perc
Naypyitaw	16.09010850
Mandalay	15.13020824
Yangon	14.87020798

* Which customer type pays the most VAT?

customer_type	VAT
Member	15.61457214
Normal	15.09805040

Customer Analysis

* How many unique customer types exist?

customer_type
Normal
Member

* How many unique payment methods exist?

payment_method
Credit card
Ewallet
Cash

* What is the most common customer type?

customer_type	cnt
Member	499
Normal	496

* What is the gender of most customers?

gender	cus
Male	498
Female	497

* What is the gender distribution per branch?

For Branch A, B, C

	gender	cus
▶	Male	179
	Female	160

	gender	cus
▶	Male	169
	Female	160

	gender	cus
▶	Female	177
	Male	150

* Which day of the week has the highest sales?

	day_name	total_sales
▶	Saturday	56120.8095
	Tuesday	51482.2455
	Thursday	45349.2480
	Sunday	43937.4810
	Friday	43848.4095
	Wednesday	42803.4180
	Monday	37344.7830

* During which time of the day do customers give the most ratings?

	time_of_day	avg_rating
▶	Afternoon	7.02340
	Morning	6.94474
	Evening	6.90536

* Which time of the day receives the highest ratings per branch?

For Branch A, B, C

	time_of_day	avg_rating
▶	Afternoon	7.18889
	Morning	7.00548
	Evening	6.87143

	time_of_day	avg_rating
▶	Morning	6.83793
	Afternoon	6.81129
	Evening	6.75102

	time_of_day	avg_rating
▶	Evening	7.09859
	Afternoon	7.06667
	Morning	6.97458

* Which day of the week has the best average ratings?

	day_name	avg_rating
▶	Monday	7.13065
	Friday	7.05507
	Tuesday	7.00316
	Sunday	6.98864
	Saturday	6.90183
	Thursday	6.88986
	Wednesday	6.76028

* Which day of the week has the best average ratings per branch?

For Branch A, B, C

day_name	avg_rating
Friday	7.31200
Monday	7.09792
Sunday	7.07885
Tuesday	7.05882
Thursday	6.95870
Wednesday	6.84286
Saturday	6.74600

day_name	avg_rating
Monday	7.26579
Tuesday	7.00189
Sunday	6.79706
Thursday	6.75227
Saturday	6.73667
Friday	6.69412
Wednesday	6.37959

day_name	avg_rating
Saturday	7.22963
Friday	7.20541
Wednesday	7.06400
Monday	7.03684
Sunday	7.02826
Tuesday	6.95185
Thursday	6.95000