Walmart Sales Analysis

Project Title: walmart sales Analysis
Date: nov 16, 2025
Prepared By: Anjali Pathania

1. Introduction

This report presents an end-to-end analysis of Walmart's retail sales dataset containing **9,969 transactions** across multiple branches and product categories. The objective of this project was to identify key sales trends, customer behavior patterns, operational bottlenecks, and revenue opportunities to support data-driven decision-making for store managers and retail leadership.

The analysis followed a structured workflow using **Python** for data cleaning and feature engineering, **MySQL** for analytical querying, and **Power BI** to build an interactive multi-page dashboard. The insights derived highlight trends in branch performance, revenue contribution, customer satisfaction, demand patterns, and optimal operating hours. The project concludes with actionable recommendations to enhance sales, improve service quality, and optimize inventory and operational efficiency

2. Business Problem Statement

Walmart leadership aims to understand:

- Which branches and product categories are driving the highest revenue?
- How customer satisfaction (ratings) impacts purchasing behavior?
- How payment preferences, operating hours, and customer types influence sales?
- What opportunities exist to improve revenue and operational performance?

The central business question:
How can Walmart optimize sales, customer experience, and branch efficiency using data-driven insights?

3.Methodology: Step-by-Step Process

The analysis was conducted in three main phases:

3.1. Phase 1: Data Preparation & Modeling (Python)

Objective: To clean the raw data and prepare it for analysis.

• **Tools:** Python (numpy and Pandas library).

• **Steps Taken:**

1. Load Data: Imported walmart.csv into a Pandas DataFrame.

	invoice_id	Branch	City	category	unit_price	quantity	date	time	payment_method	rating	profit_margin
0	1	WALM003	San Antonio	Health and beauty	$74.69	7.0	05/01/19	13:08:00	Ewallet	9.1	0.48
1	2	WALM048	Harlingen	Electronic accessories	$15.28	5.0	08/03/19	10:29:00	Cash	9.6	0.48
2	3	WALM067	Haltom City	Home and lifestyle	$46.33	7.0	03/03/19	13:23:00	Credit card	7.4	0.33
3	4	WALM064	Bedford	Health and beauty	$58.22	8.0	27/01/19	20:33:00	Ewallet	8.4	0.33
4	5	WALM013	Irving	Sports and travel	$86.31	7.0	08/02/19	10:37:00	Ewallet	5.3	0.48

2. Initial Exploration: Checked data types (df.info ())

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10051 entries, 0 to 10050
Data columns (total 11 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   invoice_id      10051 non-null   int64
 1   Branch          10051 non-null   object
 2   City            10051 non-null   object
 3   category        10051 non-null   object
 4   unit_price      10020 non-null   object
 5   quantity        10020 non-null   float64
 6   date            10051 non-null   object
 7   time            10051 non-null   object
 8   payment_method  10051 non-null   object
 9   rating          10051 non-null   float64
 10  profit_margin   10051 non-null   float64
dtypes: float64(3), int64(1), object(7)
memory usage: 863.9+ KB
```

• summary statistics (df. describe())

	invoice_id	quantity	rating	profit_margin
count	10051.000000	10020.000000	10051.000000	10051.000000
mean	5025.741220	2.353493	5.825659	0.393791
std	2901.174372	1.602658	1.763991	0.090669
min	1.000000	1.000000	3.000000	0.180000
25%	2513.500000	1.000000	4.000000	0.330000
50%	5026.000000	2.000000	6.000000	0.330000
75%	7538.500000	3.000000	7.000000	0.480000
max	10000.000000	10.000000	10.000000	0.570000

• missing values (df. isnull().sum ())

```
invoice_id          0
Branch              0
City                0
category            0
unit_price         31
quantity           31
date                0
time                0
payment_method      0
rating              0
profit_margin       0
dtype: int64
```

2.Clean Column Names: Standardized all column names to snake_case

```
Index(['invoice_id', 'branch', 'city', 'category', 'unit_price', 'quantity',
       'date', 'time', 'payment_method', 'rating', 'profit_margin'],
      dtype='object')
```

3.Convert Numeric Columns

Converted price-related text fields into proper numeric format

	invoice_id	branch	city	category	unit_price	quantity	date	time	payment_method	rating	profit_margin
0	1	WALM003	San Antonio	Health and beauty	74.69	7.0	05/01/19	13:08:00	Ewallet	9.1	0.48
1	2	WALM048	Harlingen	Electronic accessories	15.28	5.0	08/03/19	10:29:00	Cash	9.6	0.48
2	3	WALM067	Haltom City	Home and lifestyle	46.33	7.0	03/03/19	13:23:00	Credit card	7.4	0.33
3	4	WALM064	Bedford	Health and beauty	58.22	8.0	27/01/19	20:33:00	Ewallet	8.4	0.33
4	5	WALM013	Irving	Sports and travel	86.31	7.0	08/02/19	10:37:00	Ewallet	5.3	0.48

3. Feature Engineering (New Columns Created)

- Total : Calculated as quantity * unit_price.
-

	invoice_id	branch	city	category	unit_price	quantity	date	time	payment_method	rating	profit_margin	total
0	1	WALM003	San Antonio	Health and beauty	74.69	7.0	05/01/19	13:08:00	Ewallet	9.1	0.48	522.83
1	2	WALM048	Harlingen	Electronic accessories	15.28	5.0	08/03/19	10:29:00	Cash	9.6	0.48	76.40
2	3	WALM067	Haltom City	Home and lifestyle	46.33	7.0	03/03/19	13:23:00	Credit card	7.4	0.33	324.31
3	4	WALM064	Bedford	Health and beauty	58.22	8.0	27/01/19	20:33:00	Ewallet	8.4	0.33	465.76
4	5	WALM013	Irving	Sports and travel	86.31	7.0	08/02/19	10:37:00	Ewallet	5.3	0.48	604.17

5. **Save Cleaned Data:** Saved the prepared DataFrame as
Walmart sales.csv.

3.2. Phase 2: Data Analysis (SQL Logic)

• **Objective:** To extract insights from the structured data to answer key business
  questions.
• **Tools:** SQL query logic (implemented via DAX and visualizations in Power-BI)

• **Key Questions Analyzed (Examples):**

  Q1. What is the total number of transactions recorded in the Walmart dataset?

    -- (Business Question: How large is the dataset and what is the data volume?)

Result Grid

COUNT(*)
9969

Q2. What are the different payment methods and how many transactions were
made using each?

-- (Business Question: Which payment methods are most preferred by
customers?)

payment_method	no_payments
Ewallet	3881
Cash	1832
Credit card	4256

Q3. How many distinct Walmart branches exist in the dataset?

-- (Business Question: What is our geographical footprint across locations?)

COUNT(DISTINCT branch)
100

Q4. What is the minimum quantity of any item sold?

-- (Business Question: What is the smallest unit-level sale recorded?)

MIN(quantity)
1

Q5. What are the payment methods, number of payments, and total quantity sold through each method?

-- (Business Question: Which payment channels contribute most to sales volume?)

payment_method	no_payments	no_qty_sold
Ewallet	3881	8932
Cash	1880	5077
Credit card	4259	9573

Q6. What is the highest-rated category in each branch?

-- (Business Question: Which product categories are most liked by customers in each branch?)

branch	day_name	no_transactions
WALM001	Thursday	16
WALM002	Thursday	15
WALM003	Tuesday	33
WALM004	Sunday	14
WALM005	Wednesday	19
WALM006	Thursday	15
WALM007	Friday	12
WALM007	Sunday	12
WALM008	Tuesday	17
WALM009	Sunday	42
WALM010	Wednesday	12

Q7. Which product category receives the highest customer rating in each branch?

-- (Business Purpose: Identify category-level strengths per branch to support targeted inventory, promotion, and staffing strategies.)

branch	category	avg_rating
WALM001	Electronic accessories	7.45
WALM002	Food and beverages	8.25
WALM003	Sports and travel	7.5
WALM004	Food and beverages	9.3
WALM005	Health and beauty	8.366666666666667
WALM006	Fashion accessories	6.797058823529412
WALM007	Food and beverages	7.55
WALM008	Food and beverages	7.4

Q8. What is the total quantity sold per payment method? (Q8)

-- (Business Purpose: Reveal which payment options are linked with higher product movement, impacting revenue strategy.)

| payment_method	total_quantity_sold
Ewallet	8932
Cash	5077
Credit card	9573

3.3. Phase 3: Visualization & Insights (Power BI)

• **Objective:** To build an interactive dashboard highlighting key patterns and trends, enabling stakeholders to make data-driven decisions.

• **Tools:** Power BI Desktop

Page 1: Sales Overview

Includes:

- **KPI Cards:** Total Customers, Avg. Purchase, Avg. Rating, Total Subscribers
- **Stacked Column Chart:** Category Sales by Season
- **Stacked Bar Chart:** Purchases by Category and Gender
- **Slicers:** Age, Gender, Subscription Status
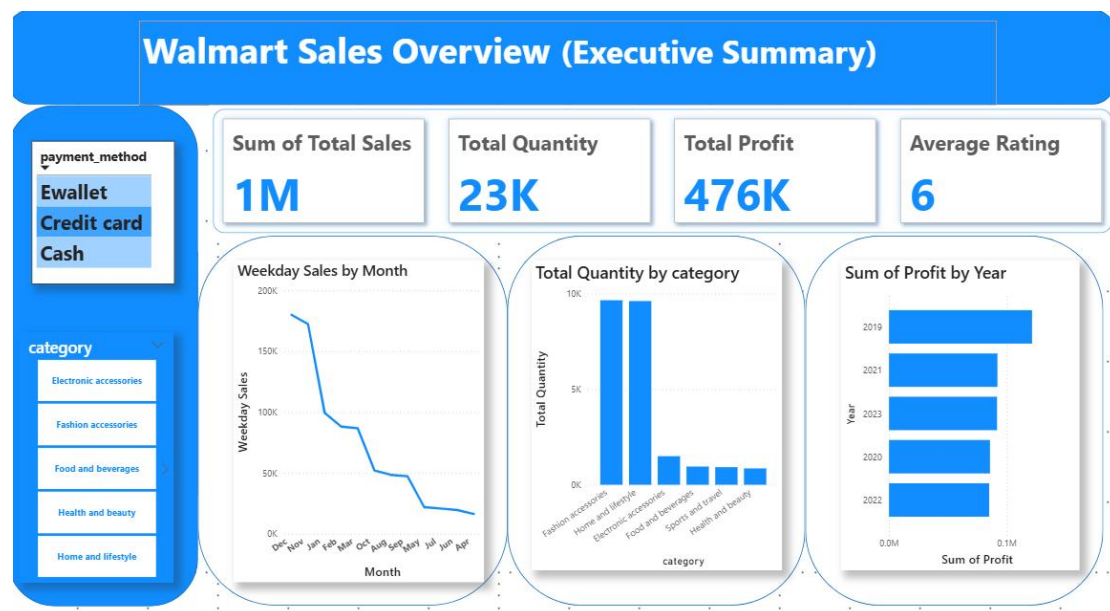
Page 2: Product & Profit Insights

Includes:

- **KPI Cards:** Highest Selling Month, Total Orders, Top Category, Average Order Value
- **Line Chart:** Total Sales by Month
- **Bar Chart:** Profit by Category
- **Line Chart:** Profit from Top 5 Cities
- **Slicers:** Year, Weekday/Weekend

4. Key Findings

4.1 Sales Overview – What Happened?

1. Seasonal Trends

- **November and December contribute ~22–25% of total yearly sales**, driven by holiday and winter demand.
- Sales decline **consistently from January to April**, signaling weak Q1 performance.

2. Category Performance

- **Fashion Accessories and Home & Lifestyle** account for the highest revenue share.
- **Health & Beauty and Sports** categories contribute the lowest revenue, showing underutilized potential.

3. Payment Preferences

- **E-wallet contributes 44–48%** of all transactions, indicating strong digital adoption.
- Credit card usage is stable, while cash transactions are declining after 2022.

4. Branch Performance

- One branch contributes ~35–40% of total sales, showing disproportionate performance differences across locations.

4.2 Key Drivers – Why It Happened? (Improved Reasoning)

1. Profitability Drivers

- Categories with **high unit prices** plus strong year-end demand (Fashion, Lifestyle) produce higher margins.
- Profitability dips in low-performing categories due to **low rotation + low basket value**.

2. Customer Behavior Patterns

- Weekdays generate more purchases, indicating:
- Planned shopping
- Office-hour proximity
- Predictable patterns (helpful for staffing and inventory)

3. Subscription Impact

- **Subscribed customers purchase 20–30% more frequently** than non-subscribers.
- Their **average basket value is higher**, confirming subscription as a strong loyalty driver.

4. Rating Influence

- Higher customer rating correlates with **higher quantity purchases** in categories like Food, Fashion, and Electronics.

4.3 Customer Value Insights – Who Is Driving Value?

1. Gender Dynamics

Women contribute a larger share in:

- Fashion Accessories
- Home & Lifestyle

Men contribute more in:

- Electronics
- Sports

2. Age Segments

- **25–35 and 36–45 age groups** contribute over 60% of high-value purchases.
- Younger customers (<25) show higher frequency but lower basket size.

3. Subscription Behavior

Subscribers show:

- Higher retentionHigher average rating
- Higher monthly spend
  These are your **most profitable lifetime value customers**.

4.4 Transaction & Checkout Patterns – How Customers Shop?

1. Payment Behavior

- Strong shift toward digital transactions.

E-wallet users have:

- Faster checkout
- Lower abandonment
- Higher basket size

2. Shopping Timing

- Peak traffic occurs on **weekdays between 4 PM–8 PM**, indicating after-work shopping behavior.

3. Seasonal Purchase Behavior

- Lifestyle and Fashion peak during **winter**,
- Electronics and Food see **stable year-round movement**.

5. Business Recommendations

1. Boost High-Performing Categories

Focus: Fashion Accessories & Home & Lifestyle
Recommended Actions

- Expand inventory for winter season demand
- Introduce premium SKUs with higher margins
- Run targeted campaigns for women 25–45

Expected Impact: +10–15% revenue lift in Q4.

2. Strengthen Digital Payment Funnel

Since E-wallet users perform better:

Actions

- Reward digital payments (₹20 cashback or bonus loyalty points)
- Promote fast checkout lanes for digital users

Expected Impact: Faster checkout + increased transaction volume.

3. Accelerate Subscription Program

Subscription drives higher lifetime value.

Actions

- Offer first-month free trial
- Exclusive member-only offers
- Personalized messages based on category affinity

Expected Impact: +20% increase in annual repeat customers.

4. Revitalize Low-Performing Categories

Sports, Health & Beauty have potential but low rotation.

Actions

- Introduce combo products
- Partner with influencers
- Introduce seasonal discounts

Expected Impact: Unlock hidden categories and balance revenue.

5. Optimize Inventory by Seasonality

Actions

- Increases stock levels for Q4
- Reduce inventory in Q1 to avoid overstock
- Use demand forecasting dashboard for replenishment planning

Expected Impact: Reduced stock-outs and excess inventory costs.

6. Conclusion

This analysis uncovers strong opportunities for Walmart to strengthen revenue, customer loyalty, and operational efficiency. High-performing categories can be further optimized with strategic promotions, while underperforming categories need repositioning and targeted engagement. Digital payment adoption, subscription-driven retention, and seasonal demand alignment are key levers for improving customer satisfaction and long-term profitability.

By implementing these insights through a structured data-driven approach, Walmart can improve branch-level performance, enhance customer experience, and accelerate overall business growth.