

Walmart Sales Data Analysis

1. Project Overview

This project analyzes Walmart sales data to identify top-performing branches and products, uncover sales trends, and understand customer behavior. The goal is to derive actionable insights to enhance and optimize sales strategies.

The dataset was sourced from the [Kaggle Walmart Sales Forecasting Competition](#).

As described in the competition:

"Participants are given historical sales data for 45 Walmart stores across different regions. Each store has multiple departments, and participants must forecast sales for each department. The dataset includes selected holiday markdown events, which significantly influence sales. However, predicting their impact across departments is challenging."

2. Dataset Information

The dataset comes from the [Kaggle Walmart Sales Forecasting Competition](#) and contains sales transactions from three Walmart branches located in Mandalay, Yangon, and Naypyitaw. It comprises 17 columns and 1000 rows:

3. SQL Workflow and Key Steps

Step 1: Database and Table Creation

- Created the **walmartSales** database and defined the **sales** table with appropriate data types.
- Ensured that critical columns like **invoice_id** were set as primary keys to maintain data integrity.

```
CREATE DATABASE IF NOT EXISTS walmartSales;
```

```
CREATE TABLE IF NOT EXISTS sales(  
    invoice_id VARCHAR(30) NOT NULL PRIMARY KEY,  
    branch VARCHAR(5) NOT NULL,  
    city VARCHAR(30) NOT NULL,  
    customer_type VARCHAR(30) NOT NULL,  
    gender VARCHAR(30) NOT NULL,
```

```
product_line VARCHAR(100) NOT NULL,  
unit_price DECIMAL(10,2) NOT NULL,  
quantity INT NOT NULL,  
tax_pct FLOAT(6,4) NOT NULL,  
total DECIMAL(12,4) NOT NULL,  
date DATETIME NOT NULL,  
time TIME NOT NULL,  
payment VARCHAR(15) NOT NULL,  
cogs DECIMAL(10,2) NOT NULL,  
gross_margin_pct FLOAT(11,9),  
gross_income DECIMAL(12,4),  
rating FLOAT(2,1)  
);
```

Step 2: Data Cleaning

- Queried the raw data to ensure accurate loading into the **sales** table.
- Examined all rows and columns to identify anomalies or missing data.

Step 3: Data Transformation

- Added a **time_of_day** column to categorize transactions into "Morning," "Afternoon," and "Evening" for time-based analysis.

```
SELECT  
    time,  
    (CASE  
        WHEN `time` BETWEEN '00:00:00' AND '12:00:00' THEN 'Morning'  
        WHEN `time` BETWEEN '12:01:00' AND '16:00:00' THEN 'Afternoon'  
        ELSE 'Evening'  
    END) AS time_of_day  
FROM sales;
```

```
ALTER TABLE sales ADD COLUMN time_of_day VARCHAR(10);
```

Step 4: Metrics Calculation

- Calculated gross income and gross margin percentages to analyze profitability.

```
UPDATE sales  
SET gross_income = total - cogs,  
    gross_margin_pct = (gross_income / total) * 100;
```

4. Key Insights

Time-Based Trends:

- The **time_of_day** column allows identification of peak sales periods, helping businesses optimize staffing and promotional activities.

Profitability Analysis:

- The **gross_income** and **gross_margin_pct** fields provide insights into the most and least profitable product lines.

Customer Behavior:

- Analyzing fields like **customer_type** and **payment** helps understand customer preferences and trends.
-

5. Optimization Techniques

- Used appropriate data types (e.g., DECIMAL for financial fields) to ensure precision and minimize storage overhead.
 - Structured queries using CASE statements to derive categorizations efficiently.
 - Indexed key columns (e.g., **invoice_id**) to enhance query performance.
-

6. Business Impact

- **Improved Decision-Making:** Insights from time-based and profitability analyses empower targeted marketing and inventory strategies.
 - **Operational Efficiency:** Automation of data transformations reduced manual errors and improved reporting accuracy.
 - **Strategic Growth:** Highlighted customer preferences and store performance trends to support long-term planning.
-

7. Tools and Techniques

- **Database Management System:** MySQL.
- **Key Techniques:** Data cleaning, categorization using CASE statements, and profitability calculations.
- **Focus Areas:** Structured analysis and efficient data transformations for actionable insights.