

## Project : Walmart Sales Analysis Using SQL

**Objective** :To gain insights into Walmart's sales performance and trends.

### Dataset Schema:

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)

## Functionality:

The Walmart Sales Analysis project aims to provide comprehensive insights into Walmart's sales performance through the analysis of its sales data. The functionality of the project includes:

### 1. Data Retrieval:

- Extracting relevant sales data from the kaggle.
- Obtaining data such as sales revenue, product information, sales dates, and store locations.

### 2. Data Cleaning and Preparation:

- Cleaning and preprocessing the extracted data to ensure accuracy and consistency.
- Handling missing values, duplicates, and inconsistencies in the dataset.
- Formatting the data for analysis.

### 3. Sales Analysis:

- Performing various analyses on the sales data to uncover patterns, trends, and insights.
- Calculating key performance indicators (KPIs) such as total sales revenue, average sales per store, and sales by product category.
- Identifying top-selling products, popular store locations, and seasonal sales trends.

### 4. Querying and Reporting:

- Writing SQL queries to retrieve specific information and answer analytical questions.

## Project implementation:

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:

## Analysis Done

### 1.Examination of Products

Analyse the data to identify the various product lines, the ones that are doing well, and the ones that want improvement.

### 2.Analysis of Sales

The goal of this investigation is to provide an answer to the query about product sales trends. The outcome of this can assist us in gauging the success of every sales technique the company employs and the adjustments required to increase sales.

### 3.Client Evaluation

The many customer segments, purchasing patterns, and profitability of each customer category are the focus of this investigation.

## Tools and technologies used : MySQL Workbench

### Some key insights that can be derived from the Walmart sales data analysis:

#### Cities and Branches:

- There are 3 unique cities represented in the data.
- The data provides insights into the location of each Walmart branch , branches are A,B,C, allowing for targeted analysis and decision-making.

#### Products:

- There are 6 unique product lines available at Walmart, showcasing the diversity of offerings.
- The most common payment method is Ewallet, indicating preferred transactional behavior among customers.
- The top-selling product line is Fashion and Acessories, highlighting consumer demand and popularity.
- Revenue varies by month, with January seeing the highest total revenue, indicating seasonal trends.
- January had the largest Cost of Goods Sold (COGS), suggesting potential areas for cost optimization.
- Food and beverages generated the largest revenue, underscoring its importance to Walmart's sales performance.
- The city with the largest revenue is Naypyitaw, indicating significant economic activity.

#### Customer Insights:

- Home and Lifestyle had the largest Value Added Tax (VAT), suggesting potential areas for tax optimization.
- Products classified as "Good" outperformed the average sales, indicating strong performance in certain product lines.
- C Branch sold more products than the average, suggesting potential areas for sales improvement.
- The most common product line purchased by females is Fashion accessories, indicating and by male is health and beauty preferences in purchasing behavior.
- The average rating of each product line varies, with Food and beverages having the highest rating, indicating customer satisfaction.
- Customer types vary in their contribution to revenue, with Membership bringing the most revenue, guiding marketing and sales strategies.
- The city with the largest VAT is Naypyitaw, highlighting tax implications across locations.
- Member, customer type pays the most in VAT, indicating differential tax burdens among customer segments.

## References:

- Links to resources used :
- <https://www.youtube.com/watch?v=hlGoQC332VM>
- <https://www.youtube.com/watch?v=36fBGMT0tuE>