[](https://camo.githubusercontent.com/148f6ed9b3519f98fa309d3b30ca09d1100a85467cf2e1151a2aa0e6a9839a47/68747470733a2f2f656e637279707465642d74626e302e677374617469632e636f6d2f696d616765733f713d74626e3a414e64394763527a744c30577a53487859647a6338466d465a3870435941384778734154685668696f416f51497266506d6e4d32353237694d6e4c76525879645f35564f31785964526726757371703d434155)

**Walmart-Sales-Data-Analysis--SQL-Project**

**About**

We are analysing Walmart's sales data to identify high-performing branches and products, analyze the sales patterns of various products, and understand customer behavior. The primary objective is to enhance and optimize sales strategies. The dataset utilized in this project is sourced from the Kaggle Walmart Sales Forecasting Competition.

**Purposes of the Project**

The main goal of this project is to gain understanding from Walmart's sales data, exploring the various factors that influence sales across different branches.

**About Data**

This project's data was obtained from the Kaggle Walmart Sales Forecasting Competition and it encompasses sales transactions from three Walmart branches situated in Mandalay, Yangon, and Naypyitaw, respectively. 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(12, 4) |
| date | The date on which the purchase was made | DATETIME |
| time | The time at which the purchase was made | TIME |
| payment | The total amount paid | DECIMAL(10, 2) |
| cogs | Cost Of Goods sold | DECIMAL(10, 2) |
| gross\_margin\_pct | Gross margin percentage | FLOAT(11, 9) |
| gross\_income | Gross Income | DECIMAL(12, 4) |
| rating | Rating | FLOAT(2, 1) |

**Analysis List:**

1. Product Analysis

Perform an analysis on the data to gain insights into different product lines, determine the top-performing product lines, and identify areas for improvement in other product lines.

1. Sales Analysis

The objective of this analysis is to address the inquiry regarding the sales trends of the product. The outcomes of this analysis can assist in evaluating the efficiency of each applied sales strategy in the business and determining necessary modifications to increase sales.

1. Customer Analysis

This analysis is focused on identifying various customer segments, understanding purchasing trends, and evaluating the profitability associated with each of these customer segments.

**Approach Used**

***1. Data Wrangling***

During this initial phase, the data is examined to detect any NULL or missing values, and strategies for data replacement are implemented to address and substitute these values effectively.

* Build a database
* Create a table and insert the data.
* Select columns with null values in them. Null values are not present in our database because, in creating the tables, NOT NULL was specified for each field, effectively filtering out any null values.

***2. Feature Engineering***

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

* 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.
* 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.
* 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)***

Conducting exploratory data analysis is essential to address the project's listed questions and objectives.

**Business Questions to Answer**

**Generic Questions**

1. How many distinct cities are present in the dataset?
2. In which city is each branch situated?

**Product Analysis**

1. How many distinct product lines are there in the dataset?
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. Which month recorded the highest Cost of Goods Sold (COGS)?
6. Which product line generated the highest revenue?
7. Which city has the highest revenue?
8. Which product line incurred the highest VAT?
9. Retrieve each product line and add a column product\_category, indicating 'Good' or 'Bad,' based on whether its sales are above the average.
10. Which branch sold more products than average product sold?
11. What is the most common product line by gender?
12. What is the average rating of each product line?

**Sales Analysis**

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

**Customer Analysis**

1. How many unique customer types does the data have?
2. How many unique payment methods does the data have?
3. Which 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?