**Walmart Sales Data Analysis**

**About**

This project aims to explore the Walmart Sales data to understand top performing branches and products, sales trend of of different products, customer behaviour. The aims is to study how sales strategies can be improved and optimized. The dataset was obtained from the [Kaggle Walmart Sales Forecasting Competition](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting).

"In this recruiting competition, job-seekers are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and participants must project the sales for each department in each store. To add to the challenge, selected holiday markdown events are included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact." [source](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting)

**Purposes Of The Project**

The major aim of thie project is to gain insight into the sales data of Walmart to understand the different factors that affect sales of the different branches.

**About Data**

The dataset was obtained from the [Kaggle Walmart Sales Forecasting Competition](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting). 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:

| **Column** | **Description** | **Data Type** |
| --- | --- | --- |
| invoice\_id | Invoice of the sales made | VARCHAR(25) |
| branch | Branch at which sales were made | VARCHAR(5) |
| city | The location of the branch | VARCHAR(25) |
| customer\_type | The type of the customer | VARCHAR(15) |
| gender | Gender of the customer making purchase | VARCHAR(15) |
| product\_line | Product line of the product solf | VARCHAR(30) |
| 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) |
|  |  |  |
|  |  |  |
| payment\_method | The total amount paid | DECIMAL(10, 2) |
| cogs | Cost Of Goods sold | DECIMAL(7, 2) |
| gross\_margin\_percentage | Gross margin percentage | FLOAT(11, 9) |
| gross\_income | Gross Income | DECIMAL(7, 4) |
| rating | Rating | FLOAT(2, 1) |

### Analysis List

1. Product Analysis

Conduct analysis on the data to understand the different product lines, the products lines performing best and the product lines that need to be improved.

1. Sales Analysis

This analysis aims to answer the question of the sales trends of product. The result of this can help use measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

1. Customer Analysis

This analysis aims to uncover the different customers segments, purchase trends and the profitability of each customer segment.

## Approach Used

1. **Data Wrangling:** This is the first step where inspection of data is done to make sure **NULL** values and missing values are detected and data replacement methods are used to replace, missing or **NULL** values.
2. Build a database
3. Create table and insert the data.
4. Select columns with null values in them. There are no null values in our database as in creating the tables, we set **NOT NULL** for each field, hence null values are filtered out.
5. **Exploratory Data Analysis (EDA):** Exploratory data analysis is done to answer the listed questions and aims of this project.

## Business Questions To Answer

### Generic Question

1. How many unique cities does the data have?
2. In which city is each branch?

### Product

1. How many unique product lines does the data have?
2. What is the most common payment method?
3. What is the most selling product line?
4. What month had the largest COGS?
5. What product line had the largest revenue?
6. What is the city with the largest revenue?
7. What product line had the largest VAT?
8. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales
9. Which branch sold more products than average product sold?
10. What is the most common product line by gender?
11. What is the average rating of each product line?

### Sales

1. Which of the customer types brings the most revenue?
2. Which city has the largest tax percent/ VAT (**Value Added Tax**)?
3. Which customer type pays the most in VAT?

### Customer

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

\