# **Walmart Sales Data Analysis**

## **Overview**

The objective of this project is to analyze Walmart sales data to gain insights into the performance of different branches and products, understand sales trends, and identify opportunities for improvement.

The dataset used for analysis is sourced from the [Kaggle Walmart Sales Forecasting Competition](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting), containing historical sales data from 45 Walmart stores across various regions, including information on departments and holiday markdown events.



## **Project Goals**

The primary goals of this project include:

* Analyzing product lines to identify top-performing and underperforming categories.
* Examining sales trends to evaluate the effectiveness of sales strategies.
* Conducting customer segmentation analysis to understand purchasing behavior and profitability.

## **Data Description**

The dataset consists of sales transactions from three Walmart branches located in Mandalay, Yangon, and Naypyitaw, comprising 17 columns and 1000 rows.

## **Analysis Objectives**

**Product Analysis:** Investigate product categories to determine sales performance and areas for improvement.

**Sales Analysis:** Evaluate sales trends to assess the impact of different strategies and identify opportunities for optimization.

**Customer Analysis:** Segment customers based on purchasing behavior to enhance targeting and profitability.

## **Methodology**

**Data Wrangling:** Ensure data quality by detecting and handling missing values appropriately.

**Feature Engineering:** Create new features such as time\_of\_day, day\_name, and month\_name to provide additional insights into sales patterns.

**Exploratory Data Analysis (EDA):** Utilize various visualization techniques to explore relationships between variables and uncover trends in the data.

## **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 is the total revenue by month?
5. What month had the largest COGS?
6. What product line had the largest revenue?
7. What is the city with the largest revenue?
8. What product line had the largest VAT?
9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales
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**

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

### **Customer**

1. How many unique customer types does the data have?
2. What is the most common customer type?
3. Which customer type buys the most?
4. What is the gender of most of the customers?
5. What is the gender distribution per branch?
6. Which time of the day do customers give most ratings?
7. Which time of the day do customers give most ratings per branch?
8. Which day fo the week has the best avg ratings?
9. Which day of the week has the best average ratings per branch?

## **CALCULATIONS**

* *To calculate the Cost of Goods Sold (COGS), we use the formula:*

$ COGS = \text{unitPrice} \times \text{quantity} $

* *Additionally, we calculate the Value Added Tax (VAT) as 5% of the COGS:*

$ VAT = 5% \times COGS $

*The VAT is then added to the COGS to determine the total amount billed to the customer:*

$ \text{total (gross sales)} = VAT + COGS $

*Next, we compute the gross profit (gross income) by subtracting the COGS from the total:*

$ \text{grossProfit (gross income)} = \text{total (gross sales)} - COGS $

*The Gross Margin is expressed as a percentage of the total revenue and is calculated as:*

$ \text{Gross Margin} = \frac{\text{gross income}}{\text{total revenue}} $

For example, using the data from the first row in our database:

Given:

**$ \text{Unit Price}** = 45.79 $

**$ \text{Quantity}** = 7 $

We calculate the COGS as follows:

**$ COGS** = 45.79 \times 7 = 320.53 $

Then, we compute the VAT:

**$ VAT** = 5% \times 320.53 = 16.0265 $

Adding VAT to COGS gives us the total:

**$ \text{total}** = VAT + COGS = 16.0265 + 320.53 = 336.5565 $

Finally, we determine the Gross Margin Percentage:

$ \text{Gross Margin Percentage} = \frac{\text{gross income}}{\text{total revenue}} = \frac{16.0265}{336.5565} \approx 4.7619% $

This process allows us to accurately assess the profitability of each transaction and analyze the overall financial performance of the business.

## **Conclusion**

By conducting thorough analysis of Walmart sales data, this project aims to provide actionable insights that can inform strategic decision-making and drive improvements in sales performance and customer satisfaction.