**Walmart Sales Data Analysis**

**Project Overview**

This project delves into the analysis of Walmart sales data, aiming to uncover insights into sales patterns, product performance, customer behaviour, and revenue trends across different branches and cities. 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.

**Dataset**

The project utilizes a comprehensive Walmart sales dataset containing information such as invoice details, product lines, customer types, gender distribution, payment methods, and more. The SQL queries provided were employed to extract valuable insights from this dataset.

**Data Preparation Approach**

**1. Data Wrangling:** Conducted initial data inspection to detect and handle null values. The database was built, tables created, and data inserted. Null values were eliminated by setting 'NOT NULL' constraints for each field during table creation.

**2. Feature Engineering:** Generated new columns such as 'time\_of\_day', 'day\_name', and 'month\_name' from existing data to provide insights into sales patterns by time, day, and month.

1. **Exploratory Data Analysis (EDA):** Utilized SQL queries to perform exploratory analysis, addressing the outlined business questions and project objectives.

**Analysis and Insights**

**1. Product Analysis**

To analyze product lines, identify top-performing products, and suggest improvements for underperforming product lines.

**2. Sales Analysis**

To uncover sales trends, assess the effectiveness of sales strategies, and recommend modifications for increased sales.

**3. Customer Analysis**

To segment customers, analyze their purchase trends, and evaluate profitability across customer segments.

**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. 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?
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 fo the week has the best avg ratings?
10. Which day of the week has the best average ratings per branch?

**Key Findings and Future Analysis**

The analysis revealed crucial insights, including revenue patterns, product performance, and customer preferences. Further explorations could include predictive modeling, sales forecasting, and targeted marketing strategies based on the identified trends and patterns.