**Walmart Sales Analysis**

**About**

The aim of this project is to explore Walmart sales data in order to better understand top performing branches and items, sales trends for various categories, and consumer behaviour. The goal is to investigate how sales strategies might be enhanced and maximized. The data was obtained from the Kaggle Walmart Sales Forecasting Competition.  
  
"In this recruitment challenge, applicants are given historical sales data for 45 Walmart shops located in different regions. Each business has several sections, and players must forecast sales for each one. The collection includes chosen holiday markdown events, which adds to the complexity. These markdowns are known to have an influence on sales, but it is difficult to forecast which departments will be affected and how much.

**Project Purpose**

The primary goal of this project is gaining insight into Walmart's sales data in order to better understand all the variables that influence sales across various branches.

**About Data**

The data has been collected from the Kaggle Walmart Sales Forecasting Competition. This dataset includes sales transactions from three separate Walmart branches, situated in Mandalay, Yangon, and Naypyitaw. The data consists of 1000 rows and 17 columns.

|  |  |  |
| --- | --- | --- |
| **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 branch | 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 the 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 purchase was made | DATE |
| Time | The time at which 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) |

**Analysis List**  
**1. Product Analysis:**   
  
Analyse the data to better understand the different product lines, which are working well and which need to be enhanced.  
  
**2. Sales Analysis:**   
  
The purpose of this analysis is to address the question about product sales patterns. As a consequence, we can better analyse the performance of each sales technique used by the company and identify what changes are required to increase sales.  
  
**3. Customer Analysis:**   
  
The analysis seeks to identify the various client segments, buying trends, and profitability for each customer category.

**Approach used**  
 **1. Data Wrangling:** This is the initial phase in which data is inspected to ensure that NULL and missing values are discovered, and data replacement methods are employed to replace any missing or NULL values.  
1. Create a database.  
2. Create a table and input the data.  
3. Select columns that contain null values. Our database contains no null values since when we created the tables, we specified NOT NULL for each column, therefore null values are filtered out.

**2. Feature Engineering:** This will allow us to generate new columns from current ones.

1. Add a new column titled time\_of\_day to track sales in the morning, afternoon, and evening. This will assist to answer the issue of when the greatest sales occur.  
  
2. Add a new column entitled day name that includes the extracted days of the week when the specified transaction occurred (Mon, Tue, Wed, Thurs, Fri). This will assist to answer the issue of which days of the week each branch is busiest.  
  
3. Add a new column called month\_name that holds the extracted months of the year when the supplied transaction occurred (Jan, Feb, Mar). Help identify which month of the year has the most sales and profit.

**2. 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 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 lines showing "Good", "Bad". Good if it’s 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 of the week has the best avg ratings?

**Conclusion**

In summary, while gender distribution and time of day may not significantly impact sales and ratings across branches, it's clear that certain branches outperform others in terms of ratings. Branches A and C are currently excelling, while Branch B could benefit from improvement efforts to enhance its ratings. Interestingly, Mondays, Tuesdays, and Fridays emerge as the top-performing days for achieving higher ratings. Moving forward, focusing on areas of improvement for Branch B and leveraging peak rating days across all branches could be key strategies for maximizing overall performance and customer satisfaction.