**Sales Data Analysis**

**Overview:**

This project explores Walmart sales data to understand branch performance, popular products, and customer behaviour. The dataset is from the Kaggle Walmart Sales Forecasting Competition(https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting).

**Purpose**

Gain insights into Walmart sales for better strategies and optimizations. The dataset includes historical sales data from 45 stores, each with multiple departments, making sales forecasting challenging due to included holiday markdown events.

**Key Points**

The data contains 17 columns and 1000 rows:

|  |  |  |
| --- | --- | --- |
| **Column** | **Description** | **Data Type** |
| Invoice\_Id | Invoice of the sales | VARCHAR(30) |
| Branch | Branch at which sales were made | VARCHAR(5) |
| City | The location of the branch | VARCHAR(20) |
| Customer\_type | The type of customer | VARCHAR(30) |
| Gender | Gender of the customer | VARCHAR(10) |
| Product\_line | Product line of the product sold | VARCHAR(10) |
| Unit\_price | Price of each product | DECIMAL(10,2) |
| Quantity | Number of product sold | INT |
| VAT | The amount tax of the purchase | FLOAT(6,4) |
| Total | The total cost of the purchase | DECIMAL(10,2) |
| Date | Product purchase date | DATE |
| Time | Product purchase time | TIME |
| Payment\_method | Mode of the payment | 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) |

**Steps**

**1. Data Wrangling:** This is the first step where data is inspected to make sure no value is missing or NULL

- Build database

- Create tables

**2. Feature Engineering:** This will help use generate some new columns from existing ones.

> 1. 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.

> 2. 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.

> 3. 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.

**Business Questions Answered**

**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?

5. What is the city with the largest revenue?

6. What product line had the largest VAT?

7. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales

8. Which branch sold more products than average product sold?

9. 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?

**Revenue and Profit Calculations**

1.COGS: `unit\_price x quantity`

2. VAT: 5% of COGS.( VAT is added to the $COGS$ and this is what is billed to the customer.)

3. Total Revenue(gross sales) : VAT + COGS.

4. Gross Income: Total revenue - COGS.

5. Gross Margin Percentage:

**Conclusion**

This analysis provides actionable insights for Walmart, guiding decisions for improved sales strategies and operations. The findings cover product performance, customer trends, and overall revenue, contributing to informed decision-making.