Walmart-Sales-Data-Analysis—Using SQL

About

We are examining Walmart's sales data to identify top-performing branches and products, analyze sales trends across different items, and gain insights into customer behavior. The main goal is to refine and optimize sales strategies. The dataset used in this project is from the Kaggle Walmart Sales Forecasting Competition.

Purpose of the Project

The main goal of this project is to gain understanding from Walmart's sales data, exploring the various factors that influence sales across different branches.

About Data

This project's data was obtained from the Kaggle Walmart Sales Forecasting Competition and it encompasses sales transactions from three Walmart branches situated in Mandalay, Yangon, and Naypyitaw, respectively. The data contains 17 columns and 1000 rows:

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(40)
branch	Branch at which sales were made	VARCHAR(15)
city	The location of the branch	VARCHAR(40)
customer_type	The type of the customer	VARCHAR(40)
gender	Gender of the customer making purchase	VARCHAR(15)
product_line	Product line of the product sold	VARCHAR(150)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT

Column	Description	Data Type
VAT	The amount of tax on the purchase	DECIMAL(6, 4)
total	The total cost of the purchase	DECIMAL(12, 4)
date	The date on which the purchase was made	DATETIME
time	The time at which the purchase was made	TIME
payment	The total amount paid	VARCHAR(25)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_pct	Gross margin percentage	DECIMAL(15, 9)
gross_income	Gross Income	DECIMAL(15, 4)
rating	Rating	DECIMAL(5, 3)

Analysis List:

□ Product Analysis

Conduct an analysis of the data to uncover insights into different product lines, identify the top performers, and pinpoint areas that need improvement in other lines.

□ Sales Analysis

This analysis aims to explore sales trends for the product, helping assess the effectiveness of current sales strategies and highlight areas that require adjustments to boost sales.

□ Customer Analysis

The focus of this analysis is to identify distinct customer segments, analyze purchasing behaviours, and assess the profitability linked to each segment.

Approach Used

1. Data Wrangling

During this initial phase, the data is examined to detect any NULL or missing values, and strategies for data replacement are implemented to address and substitute these values effectively.

- Build a database
- Create a table and insert the data.

- Select columns with null values in them. Null values are not present in our database because, in creating the tables, NOT NULL was specified for each field, effectively filtering out any null values.

2. Exploratory Data Analysis (EDA)

Conducting exploratory data analysis is essential to address the project's listed questions and objectives.

Business Questions to Answer

General Questions

1. How many distinct cities are present in the dataset?
2. In which city is each branch situated?

Product Analysis

1. How many distinct product lines are there in the dataset?
2. What is the most common payment method?
3. What is the most selling product line?
4. Which product line generated the highest revenue?
5. Which city has the highest revenue?
6. Which product line incurred the highest VAT?

7. Which branch sold more products than average product sold?
8. What is the most common product line by gender?
9. What is the average rating of each product line?

Sales Analysis

1. Identify the customer type that generates the highest revenue.
2. Which city has the largest tax percent/ VAT (Value Added Tax)?
3. Which customer type pays the most VAT?

Customer Analysis

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