Supermarket Sales Analysis

Author: B. Vasanth Kumar
Date: 17-12-2024

Table of Contents

Introduction

This project aims to analyze supermarket sales data using SQL to extract meaningful insights about customer behavior, product performance, and sales trends. By leveraging SQL, we address business-critical questions such as revenue trends, customer demographics, and time-based patterns. The analysis is designed to support decision-making for improved operational efficiency and customer satisfaction.

Objectives:

1. Identify revenue distribution across branches and cities.
2. Understand customer demographics and preferences.
3. Analyze time-based sales patterns and product performance.
4. Highlight key business insights for strategic improvements.

Dataset Overview

Source

- **Dataset**: Supermarket Sales Dataset
- **Provider**: Kaggle
- **Link**: Supermarket Sales Dataset

Attributes

The dataset includes the following columns:

- **Invoice ID**: Unique identifier for each transaction.
- **Branch**: Branch location (A, B, or C).
- **City**: City where the branch is located.

- **Customer Type**: Type of customer (Member or Normal).
- **Gender**: Gender of the customer.
- **Product Line**: Category of products purchased.
- **Unit Price**: Price per unit of product.
- **Quantity**: Number of units purchased.
- **Tax 5%**: Tax applied to the transaction.
- **Total**: Total amount paid by the customer (including tax).
- **Date**: Date of the transaction.
- **Time**: Time of the transaction.
- **Payment**: Payment method (Cash, Credit Card, or E-Wallet).
- **COGS**: Cost of goods sold.
- **Gross Margin Percentage**: Gross margin percentage for the transaction.
- **Gross Income**: Gross income from the transaction.
- **Rating**: Customer rating of the service.

Methodology

Approach

1. **Data Exploration**:
    - Analyzed the structure and attributes of the dataset.
    - Identified ambiguities in terms like "Total" and "Quantity."
2. **SQL Queries**:
    - Formulated SQL queries to answer 20 predefined business questions.
    - Utilized aggregation functions, grouping, and filtering for analysis.
3. **Data Validation**:
    - Verified query outputs for accuracy.
    - Cross-checked results to ensure consistency.

SQL Concepts Used

- **Basic SQL**: SELECT, WHERE, GROUP BY, ORDER BY, COUNT, SUM, AVG.
- **Intermediate SQL**: Subqueries, joins, and filtering.
- **Advanced SQL**: Window functions, CASE statements.

Analysis Questions

1. What is the total revenue generated by each branch?
2. Which city has the highest total sales?
3. Identify the most popular product line based on the total sales amount.
4. What is the average gross income per transaction for each branch?
5. Calculate the total quantity of products sold in each product line.
6. What is the gender distribution of customers for each branch?
7. Determine the average rating given by "Member" vs. "Normal" customers.
8. Find the city with the highest number of "Normal" customers.
9. Identify the top three payment methods used by customers.
10. What is the average purchase amount for male and female customers?
11. Which month had the highest sales revenue?
12. Find the peak sales hour for each branch.
13. Calculate the total revenue for weekends vs. weekdays.
14. Determine the branch with the highest revenue on weekends.
15. What are the top three most profitable days based on gross income?
16. Which product line has the highest average rating?
17. Find the most frequently purchased product line by "Normal" customers.
18. Calculate the total revenue generated by each product line in each city.
19. Determine the product line with the highest unit price on average.
20. What is the correlation between quantity sold and total revenue for each branch?

Key Insights

1. **Branch Revenue**: Branch C generated the highest revenue, with a total of 110568.
2. **City Insights**: City “Naypyitaw” accounted for the largest share of total sales.
3. **Product Line Popularity**: Product Line "Food and Beverages" emerged as the most purchased line based on total sales.
4. **Customer Demographics**:
    - Normal had a higher average purchase value than Member customers.
5. **Time-Based Trends**:
    - Peak sales occurred on weekends, especially during 6 PM.

    - Month January recorded the highest revenue.
6. **Payment Preferences**: Ewallet were the most preferred payment method.

Conclusion

This project provided a detailed analysis of supermarket sales data, uncovering key insights into branch performance, customer behavior, and product trends. The findings can be utilized to:

- Optimize inventory management for popular product lines.
- Tailor marketing strategies for different customer segments.
- Improve operational efficiency by aligning staffing with peak sales hours.

Through this project, I gained hands-on experience with SQL queries, learned to address real-world business problems, and improved my ability to analyze datasets systematically. The experience lays a strong foundation for future projects in domains like e-commerce, finance, and healthcare.

Note: For SQL queries and detailed outputs, please refer to the accompanying SQL script file and insights folder in the repository.