Customer Insights Analysis Using SQL — Portfolio Project

Objective

The goal of this project is to apply your SQL knowledge to analyze real-world retail data and uncover meaningful customer and business insights. By completing this project, you'll gain practical experience working with SQL for data cleaning, transformation, querying, and reporting, preparing you to confidently showcase your skills to potential employers. This project will become part of your professional portfolio, highlighting your competency in end-to-end SQL-based data analysis.

Project Background

You have been hired as a junior data analyst at a retail firm. The company has collected detailed transactional data, including information about customers, their purchases, and product categories. Your task is to analyze these transactions to uncover key insights that will guide business decisions. This includes customer behavior analysis, product performance, seasonal trends, segmentation, and more.

Dataset Description

You will be working with the **'Retail Sales and Customer Demographics Dataset'** available on Kaggle. Though synthetic, the dataset is designed to mimic real-world data. It contains approximately 50,000 transaction records with the following key columns:

- **Transaction ID:** A unique identifier for each transaction, allowing tracking and reference.
- **Date:** The date when the transaction occurred, providing insights into sales trends over time.
- **Customer ID:** A unique identifier for each customer, enabling customer-centric analysis.
- **Gender:** The gender of the customer (Male/Female), offering insights into gender-based purchasing patterns.

- **Age:** The age of the customer, facilitating segmentation and exploration of age-related influences.
- **Product Category:** The category of the purchased product (e.g., Electronics, Clothing, Beauty), helping understand product preferences.
- **Quantity:** The number of units of the product purchased, contributing to insights on purchase volumes.
- **Price per Unit:** The price of one unit of the product, aiding in calculations related to total spending.
- **Total Amount:** The total monetary value of the transaction, showcasing the financial impact of each purchase.

This dataset allows for customer-centric analysis, sales and marketing trend discovery, and product-level evaluation.

Analysis Questions

Use SQL to answer the following 10 questions. Each question is designed to test specific SQL skills such as aggregations, joins, subqueries, window functions, and time-series analysis. You must structure your SQL code clearly and provide screenshots of your output for the final report:

1. Retrieve all transactions with valid customer and product data.
   Order by transaction date to understand the chronological flow of purchases.
2. Clean the dataset by ensuring that numeric fields like Quantity, Price per Unit, and Total Amount are properly formatted.
   Remove duplicates or null values if any exist.
3. Calculate the total and average revenue for each product category.
   Which categories bring in the most and least revenue?
4. Analyze the monthly sales trend over the entire dataset period.
   Summarize total revenue per month and order the results chronologically.
5. Identify the top 10 customers by total spending.
   Rank customers based on how much they've spent across all transactions.
6. Calculate the average transaction value for each customer.
   How much does each customer spend per transaction on average?

7. Group customers by gender and age brackets (e.g., 18–25, 26–35, 36–50, etc.).
   Summarize total revenue and transaction count for each group.

8. Compare the number of one-time buyers versus repeat buyers.
   Group customers by purchase frequency to determine repeat behavior.

9. Identify inactive customers who have not made a purchase in the last 6 months.
   Use the most recent date in the dataset as the reference point.

10. Perform RFM (Recency, Frequency, Monetary) analysis for customer segmentation.
    Recency: Days since last purchase; Frequency: Number of purchases; Monetary: Total amount spent.

11. Find the product categories with the highest average quantity per transaction.
    Which product types are purchased in bulk?

12. Identify the busiest sales day of the week.
    Which day(s) consistently have the highest transaction volume or revenue?

13. Calculate total revenue and average spend per transaction by gender.
    Are there differences in spending patterns across genders?

14. Find the top 5 most frequently purchased product categories.
    Based on number of transactions involving each category.

15. Determine the percentage of total revenue contributed by each age group.
    Which customer age brackets are most valuable to the business?

Deliverables

Each student must submit the following individually:

- ✓ A well-structured, well-commented SQL script used for the analysis
- ✓ A final PDF report that includes project overview, data insights, screenshots of SQL output, and conclusions
- ✓ A GitHub repository with all project files (SQL script and report)

✓ A LinkedIn post summarizing the project, your role, and what you learned during the process

Project Format & Submission Guidelines

✓ This is a group project (3 students per group), but each student must submit their own deliverables.

✓ Collaborate on code development and insight generation, but ensure your individual report is uniquely written.

✓ Include interpretations of each SQL output, why the query was written the way it was, and what the results mean for the business.

✓ Ensure your GitHub and LinkedIn submissions are public, clear, and professional. For GitHub please ensure that the files are saved in you Data Analytics Repository in a Folder named "SQL".

✓ After uploading and writing a LinkedIn post about the project, please submit the **links [here](#)**

✓ The deadline for all submissions is 21st May 2025.

Final Note

This project is your chance to demonstrate end-to-end analytical thinking and technical skill using SQL. Treat it as a real client assignment. Write clean, professional code and explanations. You will be evaluated on creativity, problem-solving, correctness of queries, clarity of explanations, and your ability to present results in a business-friendly format. Use this opportunity to build a standout project for your Data Analytics career portfolio.

Prepared by: **Samuel Mati**

Data Analytics Instructor