Kumar Boste

https://github.com/KumarBoste/Retail_Sales_Analysis_Project

Abstract

This project demonstrates how SQL can be effectively used to explore, clean, and analyze retail sales data. The main goal is to uncover insights about customer behavior, sales performance, and product categories by performing SQL-based data manipulation and analysis. Through database setup, data cleaning, exploratory data analysis (EDA), and business queries, this project highlights key analytical skills essential for retail business intelligence.

Contents :

1. Project Overview

This Retail Sales Analysis project aims to simulate real-world analytical tasks performed by data analysts in retail organizations. The project covers the complete workflow — from creating and managing a retail sales database to deriving actionable insights using SQL queries.

• Database: retail_db

• Table: retail_sales

• Tools Used: SQL (PostgreSQL/MySQL), Excel (optional for data import)

• Key Techniques: Data Cleaning, EDA, Aggregation, Ranking, Filtering, and Conditional Logic

2. Business Problem

Retail businesses need data-driven strategies to optimize sales, identify customer preferences, and track performance. The primary business questions include:

• Which product categories generate the most revenue?
• Who are the top customers?
• What are the seasonal or monthly sales trends?
• Which time shifts drive higher sales?
• How do demographics like gender and age impact purchases?

3. Project Workflow

Step 1: Database Setup

A retail database was created with transactional-level sales data.

CREATE DATABASE retail_db;
CREATE TABLE retail_sales (...);

Key fields include transaction details, customer demographics, product category, and financial metrics like COGS and total_sale.

Step 2: Data Cleaning
Ensured data quality by removing records with missing values, checking for nulls and duplicates, and validating transaction IDs.

DELETE FROM retail_sales WHERE sale_date IS NULL OR customer_id IS NULL OR total_sale IS NULL;

Step 3: Exploratory Data Analysis (EDA)
Performed initial checks such as record count, unique customers, category distribution, and sales range. These steps helped understand dataset structure and potential outliers.

4. Data Analysis & Insights

A series of SQL queries were developed to answer critical business questions regarding sales patterns, customer demographics, and performance trends.

5. Statistical & Business Insights

a. Customer Demographics: Average age of shoppers lies between 25–40 years.
b. Sales Distribution: Clothing and Beauty are top-performing categories.
c. Time-Based Trends: November emerged as a strong month for sales.
d. Customer Loyalty: Top 5 customers generate a significant share of total sales.

6. SWOT Analysis

Strengths:
• Robust SQL-based analytical framework
• Clean, structured data

Weaknesses:
• Limited demographic details
• No regional segmentation

Opportunities:
• Cross-sell and loyalty campaigns
• Predictive modeling potential

Threats:
• Seasonal dependency on revenue
• Market competition

7. Reports Summary

A. Sales Summary Report – Total sales volume by category and customer demographics.
B. Trend Analysis Report – Monthly and time-based performance trends.
C. Customer Insight Report – Top customers and unique customer distribution per category.

8. Conclusion

This Retail Sales Analysis Project demonstrates how SQL can transform raw data into actionable insights. It provides a clear view of sales trends, customer segmentation, and revenue drivers. The project also establishes a strong foundation for aspiring data analysts to master SQL-based business analysis.