INTERNSHIP PROJECT REPORT

Retail Sales, Revenue & Customer Analysis Using SQL and Power BI

Submitted in partial fulfillment of the requirements
for the award of Internship Completion

Submitted by

(sivapradeep)
Internship Duration:
(Start Date – End Date)

Abstract

In the era of digital transformation, data analytics has become a crucial component of business strategy, especially in the retail sector. Retail organizations generate vast amounts of transactional data daily which, when analyzed effectively, provide valuable insights into customer behavior, sales performance, and revenue optimization.

This internship project focuses on conducting a comprehensive analysis of retail sales data using Structured Query Language (SQL) for data processing and analysis, and Power BI for visualization and reporting. The project involves data import, data cleaning, discount calculation, net revenue computation, descriptive statistical analysis, trend analysis, anomaly detection, and dashboard development. The final outcome is an interactive business intelligence solution that supports data-driven decision-making.

Introduction

Retail businesses operate in a highly competitive environment where understanding customer preferences and sales trends is essential for sustaining growth. Traditional reporting systems often fail to extract meaningful insights from large datasets. Data analytics bridges this gap by transforming raw transactional data into actionable business intelligence.

SQL is widely used for handling structured data due to its efficiency in querying and aggregating large datasets. Power BI complements SQL by enabling interactive dashboards and intuitive data visualizations. This internship project demonstrates an end-to-end retail analytics workflow using SQL and Power BI.

Internship Objectives

The primary objectives of this internship project are:

- To understand real-world retail transaction data

- To design and manage a relational database using SQL

- To perform data cleaning and preparation

- To calculate discounts and net revenue accurately

- To conduct descriptive statistical analysis

- To identify sales and revenue trends

- To detect anomalies and unusual sales patterns

- To create an interactive Power BI dashboard

- To gain hands-on experience with SQL and DAX

- To generate meaningful business insights

Organization and Domain Overview

Domain: Retail Analytics

Retail analytics focuses on analyzing sales transactions to improve business performance. It enables organizations to answer key business questions such as:

- Which store locations generate the highest revenue?

- How do discounts affect net revenue?

- What are customer purchasing patterns?

- Which time periods show high or low demand?

This project simulates real-world retail analytics scenarios commonly encountered by data analysts in industry environments.

Dataset Description

The dataset used in this project consists of 1,000 retail sales transactions. Each record represents a single customer purchase.

*Dataset Fields

Column Name	Description
transaction_id	Unique transaction identifier
customer_id	Unique customer identifier
customer_age	Age of the customer
gender	Gender of the customer
product_category	Product category purchased
store_location	Store or city location
quantity	Number of units sold
unit_price	Price per unit
discount_percent	Discount percentage applied
payment_method	Mode of payment
transaction_date	Date of transaction
total_amount	Gross transaction amount

Tools and Technologies Used

Tool	Purpose
Excel	data Handling
MySQL 8	Database creation and SQL analysis
MySQL Workbench	Query execution and data import
Power BI Desktop	Dashboard development and visualization
SQL	Data cleaning and aggregation
DAX	KPI and measure calculations

Database Design and Implementation

Database Creation

```sql
CREATE DATABASE retail_db;
USE retail_db;
```

Table Structure

```sql
CREATE TABLE retail_analytics_dataset_column_based_filled (
    transaction_id INT PRIMARY KEY,
    customer_id INT,
    customer_age INT,
    gender VARCHAR(15),
    product_category VARCHAR(50),
    store_location VARCHAR(50),
    quantity INT,
    unit_price DECIMAL(10,2),
    discount_percent DECIMAL(5,2),
    payment_method VARCHAR(20),
    transaction_date DATE,
    total_amount DECIMAL(12,2)
);
```

Data Import Verification

```sql
SELECT COUNT(*)
FROM retail_analytics_dataset_column_based_filled;
```

A total of 1,000 records were successfully imported.

Data Cleaning and Preparation

Data cleaning was performed to ensure consistency and accuracy of the dataset. Missing discount values were replaced with zero, numeric consistency was verified, and date formats were validated.

```sql
COALESCE(discount_percent, 0)
```

This ensured accurate revenue and discount calculations.

Discount and Net Revenue Analysis

Discount Calculation

Formula:
$$\text{Discount Amount} = \text{Total Amount} \times \frac{\text{Discount Percent}}{100}$$

```sql
SELECT transaction_id,
       total_amount,
       discount_percent,
       total_amount * COALESCE(discount_percent,0)/100 AS
          discount_amount
FROM retail_analytics_dataset_column_based_filled;
```

Net Revenue Calculation

Formula:
$$\text{Net Revenue} = \text{Total Amount} - \text{Discount Amount}$$

```sql
SELECT transaction_id,
       total_amount,
       total_amount -
       (total_amount * COALESCE(discount_percent,0)/100) AS
          net_revenue
FROM retail_analytics_dataset_column_based_filled;
```

Descriptive Statistical Analysis

Statistical analysis was performed to understand the distribution of sales values.

```sql
SELECT AVG(total_amount) AS mean_sales
FROM retail_analytics_dataset_column_based_filled;
```

```sql
SELECT MIN(total_amount), MAX(total_amount)
FROM retail_analytics_dataset_column_based_filled;
```

Trend Analysis

Monthly Sales Trend

```sql
SELECT MONTH(transaction_date) AS month,
       SUM(total_amount) AS monthly_sales
FROM retail_analytics_dataset_column_based_filled
GROUP BY MONTH(transaction_date)
ORDER BY month;
```

Month-over-Month Growth

```sql
SELECT month,
       monthly_sales,
       monthly_sales -
       LAG(monthly_sales) OVER (ORDER BY month) AS sales_growth
FROM (
    SELECT MONTH(transaction_date) AS month,
           SUM(total_amount) AS monthly_sales
    FROM retail_analytics_dataset_column_based_filled
    GROUP BY MONTH(transaction_date)
) t;
```

Anomaly Detection

High-value transactions were identified using statistical thresholds.

```sql
SELECT *
FROM retail_analytics_dataset_column_based_filled
WHERE total_amount >
(SELECT AVG(total_amount) + 2*STDDEV(total_amount)
 FROM retail_analytics_dataset_column_based_filled);
```

Power BI Monthly KPI Measures

Monthly Total Revenue

```
Monthly Total Revenue =
SUM ( pro[total_amount] )
```

Monthly Net Revenue

```
Monthly Net Revenue =
SUM ( pro[net_revenue] )
```

Monthly Transactions

```
Monthly Transactions =
COUNTROWS ( pro )
```

Monthly Unique Customers

```
Monthly Unique Customers =
DISTINCTCOUNT ( pro[customer_id] )
```

Average Revenue per Customer

```
Avg Revenue per Customer (Monthly) =
DIVIDE (
    [Monthly Total Revenue],
    [Monthly Unique Customers],
    0
)
```

Month-over-Month Growth

```
Revenue MoM % =
VAR CurrentMonth = [Monthly Total Revenue]
VAR PreviousMonth =
    CALCULATE (
        [Monthly Total Revenue],
        DATEADD ( pro[transaction_date], -1, MONTH )
    )
RETURN
DIVIDE ( CurrentMonth - PreviousMonth, PreviousMonth, 0 )
```

Key Insights

- Houston and Phoenix emerged as high-performing store locations

- Discounts significantly impacted net revenue

- A small segment of customers contributed disproportionately to sales

- Monthly sales trends revealed seasonal variation

- New York showed comparatively lower average customer value

Learning Outcomes

This internship enhanced practical skills in SQL, Power BI, DAX, and statistical analysis. It provided exposure to real-world retail analytics workflows and strengthened data-driven decision-making abilities.

Limitations

- Synthetic dataset

- Limited historical time span

- Absence of profit and cost data

- No customer loyalty or retention metrics

Future Enhancements

- Predictive sales forecasting

- Customer segmentation using clustering

- Profit margin analysis

- Real-time dashboard integration

- Cloud database connectivity

Conclusion

This internship project successfully demonstrates an end-to-end retail data analytics solution using SQL and Power BI. The project highlights how structured data analysis and interactive dashboards can support strategic business decisions. The experience gained has prepared the intern for real-world data analytics roles.

Internship Declaration

This project was completed as part of my internship, where I independently performed data analysis using SQL and Power BI to extract meaningful insights from retail sales data.