SALES ANALYSIS AND REPORTING FOR A RETAIL CHAIN

Project Report

Name: Vaishnavi Desale
Course: B.Sc. Computer Science – 3rd Year
University: Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon
Submission Date: 06 January 2026

Abstract

This project focuses on the analysis of retail transaction data to identify sales trends, customer behavior, and key business insights. The dataset consists of approximately 125,000 retail transactions collected from an online source. Tools such as MySQL, Python, and Excel were used for data storage, cleaning, analysis, and visualization. The project aims to support data-driven decision-making by providing meaningful insights through charts, dashboards, and summarized findings.

---

1. Introduction

Retail organizations generate large volumes of transactional data on a daily basis. Analyzing this data is essential to understand customer purchasing patterns, sales performance, and overall business health. This project demonstrates an end-to-end data analytics workflow, starting from raw data collection to final reporting and visualization.

---

2. Problem Statement

Raw retail transaction data is complex and difficult to interpret without proper cleaning, analysis, and visualization. Businesses require structured analysis to identify sales trends, top customers, and customer engagement patterns.

---

3. Objectives

- To analyze sales trends over time
- To identify high-performing months and customers
- To perform customer segmentation and churn analysis
- To visualize insights using charts and dashboards

4. Data Description

The dataset used in this project was obtained from Kaggle and contains retail transaction records.

Key Attributes: - Customer ID - Transaction Date - Transaction Amount - Response (Customer Activity Indicator)

The dataset contains approximately 125,000 records.

---

5. Tools and Technologies

Tool	Purpose
MySQL	Database creation and data storage
Python (Pandas)	Data cleaning and analysis
Matplotlib / Seaborn	Data visualization
Excel	Dashboard and summary reporting
Jupyter Notebook	Development environment

---

6. Methodology

Phase 1: Data Collection and Database Setup

- Retail transaction data downloaded in CSV format
- Database and tables created using MySQL
- Data imported into the database

Phase 2: Data Cleaning and Preparation

- Removed missing values
- Converted date columns to datetime format
- Created additional features such as month and year
- Checked for outliers using statistical methods

Phase 3: Data Analysis

- Monthly and yearly sales analysis
- Identification of top customers by transaction count and value
- Customer churn analysis
- RFM (Recency, Frequency, Monetary) segmentation

Phase 4: Reporting

- Charts and graphs generated using Python and Excel
- Dashboard created for summary insights

- Results compiled into a project report

---

7. Data Analysis and Results

This section presents the results obtained from data analysis.

- Monthly sales trends showed seasonal patterns
- February, October, and January recorded the highest sales
- A small group of customers contributed significantly to total revenue
- Customer response rate was approximately 11%

(Charts and dashboards are included for visual representation.)

---

8. Key Findings

- Peak monthly sales reached approximately 726,775 units
- Customer CS4424 generated the highest revenue
- Majority of customers were non-responders
- RFM analysis helped identify high-value customers

---

9. Conclusion

The project successfully demonstrates how retail data can be transformed into actionable business insights using data analytics techniques. The analysis supports better decision-making and provides a scalable approach for future retail analytics projects.

---

10. Future Scope

- Sales forecasting using machine learning
- Real-time dashboard development
- Advanced customer recommendation systems

---

11. References

- Kaggle Retail Transaction Dataset
- Python Documentation (pandas, matplotlib)
