PROJECT 1

SUPERMARKET SALES ANALYSIS REPORT

Domain: Retail Analytics

1. Introduction

Retail organizations operate in highly competitive environments where understanding customer behavior and sales patterns is essential for profitability. Supermarkets, in particular, generate large volumes of transactional data that contain valuable insights about consumer preferences, purchasing frequency, and revenue drivers.

The purpose of this project is to perform an in-depth analysis of supermarket sales data to extract meaningful insights that can help management make informed decisions related to inventory planning, staffing, pricing strategies, and promotional campaigns.

This project demonstrates how data analytics techniques can transform raw transactional data into actionable business intelligence.

2. Problem Statement

Supermarkets often face challenges such as:

- **Overstocking or understocking products**
- **Inefficient staffing during peak hours**
- **Poorly timed promotions**
- **Limited understanding of customer buying behavior**

Without data-driven analysis, these decisions are based on intuition rather than evidence. This project addresses these challenges by analyzing historical sales data to identify trends, patterns, and performance drivers.

3. Objectives

The objectives of this project are:

1. **To analyze overall sales performance over time**
2. **To identify high-performing and low-performing product lines**
3. **To understand customer purchasing behavior**
4. **To determine peak sales days and hours**
5. **To provide actionable recommendations for improving business performance**

4. Dataset Description

The dataset contains transactional-level supermarket sales data. Each record represents a single purchase transaction. Key attributes include:

- **Invoice ID: Unique transaction identifier**
- **Date & Time: Timestamp of purchase**
- **Product Line: Category of the purchased item**
- **Unit Price: Price per unit of product**
- **Quantity: Number of units purchased**
- **Customer Type: Member or normal customer**
- **Gender: Customer gender**
- **Payment Method: Cash, credit card, or e-wallet**

The dataset captures both temporal and behavioral dimensions of customer activity.

5. Data Cleaning and Preparation

Data preprocessing is a critical step to ensure reliable analysis. The following steps were performed:

5.1 Column Standardization

Column names were standardized to lowercase with underscores to ensure consistency and avoid coding errors.

5.2 Date and Time Processing

Date and time columns were converted into appropriate datetime formats to enable time-series analysis.

5.3 Feature Engineering

A new variable Sales was created using:

Sales = Unit Price × Quantity

This derived feature represents the total value of each transaction.

5.4 Data Validation

- **Checked for missing values**
- **Verified numerical ranges**
- **Removed duplicate records**

After cleaning, the dataset was deemed suitable for analysis.

6. Exploratory Data Analysis (EDA)

6.1 Overall Sales Trend

Daily sales aggregation revealed fluctuations across days, indicating varying customer demand. A noticeable increase in sales was observed during weekends.

6.2 Product Performance Analysis

Product-line-wise aggregation showed that certain categories consistently outperform others. This indicates customer preference concentration and highlights opportunities for focused inventory management.

6.3 Time-Based Sales Patterns

Day of Week Analysis

Sales were highest on weekends, particularly Saturdays. This suggests increased shopping activity during non-working days.

Hourly Analysis

Evening hours between 5 PM and 7 PM recorded the highest sales volume, indicating peak shopping hours.

7. Statistical Analysis

- **Mean and Median Sales: Used to understand central tendencies**
- **Standard Deviation: Measured variability in transaction values**
- **Correlation Analysis: Revealed a strong positive correlation between quantity sold and total sales**

These statistical measures help quantify observed patterns and validate visual insights.

8. Key Insights

1. **Weekend sales significantly outperform weekday sales**
2. **Evening hours contribute the largest share of daily revenue**
3. **A small number of product lines generate a majority of sales**
4. **Higher quantity purchases directly drive revenue growth**

9. Business Recommendations

1. **Increase inventory levels for top-performing product lines**
2. **Schedule staff shifts strategically during peak hours**
3. **Run promotional campaigns during weekends**
4. **Introduce bundle pricing for frequently purchased products**

10. Conclusion

This project demonstrates the power of data analytics in retail decision-making. By analyzing transactional data, supermarkets can optimize operations, improve customer satisfaction, and increase profitability.