RETAIL STORE SALES ANALYTICS REPORT

Revenue Optimisation & Operational Performance Analysis

Title:
Retail Store Sales Analytics for Revenue Optimisation and Operational Efficiency

Sector:
Retail Industry

Team Details:

- **Niharika Choudhary - 2401010298**
- **Hardik Shreyas - 2401010178**
- **Ravichandra shinde - 2401010377**
- **Arina Ali - 2401020114**
- **Utkarsh - 2401010484**
- **Shivansh Upadhayay - 2401020109**

Institute:
Newton School of Technology

Faculty / Guide:
Satyaki Das

Date of Submission:
18th February, 2026

EXECUTIVE SUMMARY

Problem

The retail industry operates in a highly competitive and margin-sensitive environment. Store managers must make strategic decisions regarding product assortment, pricing strategies, inventory planning, and promotional campaigns. However, transactional sales data often contains inconsistencies, missing values, and operational recording errors, which can distort business insights and lead to inefficient decision-making.

This project addresses the key business question:

> ***"Which product categories and discount strategies drive revenue performance, and where do operational inefficiencies impact overall sales optimisation?"***

Approach

The project utilized a retail sales dataset sourced from Kaggle, intentionally containing inconsistencies for data cleaning and analytical practice.

Using a retail sales dataset (12,575 transactions), we:

- Cleaned and standardized data in Google Sheets
- Derived missing price and quantity values where possible
- Standardized missing categories and discount entries
- Created new fields: Transaction Month and Day of Week
- Flagged zero-value transactions for transparency
- Developed KPI-driven analysis using pivot tables
- Built an interactive dashboard with slicers

The raw dataset contained 12,575 rows and 11 columns. After cleaning and feature engineering, a structured analytical dataset was developed for KPI calculation and dashboard implementation.

Key Insights

- Revenue is concentrated in a small number of high-performing categories.
- Discounted transactions do not show a significant increase in transaction frequency or total revenue, suggesting that the discount strategy may not be effectively influencing customer purchasing behavior.
- Revenue varies across months and days, indicating seasonality.
- A small proportion of transactions required data correction, highlighting operational recording gaps.

Key Recommendations

- Prioritize inventory allocation and marketing efforts toward top-performing categories to stabilize and grow revenue.
- Reassess discount strategy and implement targeted promotions instead of broad discounting.
- Use seasonal and day-level trends to optimize staffing and promotional timing.
- Strengthen transaction recording processes to improve reporting accuracy and forecasting reliability.

SECTOR & BUSINESS CONTEXT

Sector Overview

The retail sector plays a critical role in consumer-driven economies. It involves direct sales of goods to customers across multiple product categories. Retail performance depends on demand forecasting, product assortment strategy, and pricing optimization.

Revenue stability in retail is influenced by:

- Product mix

- Promotional strategy
- Seasonal demand patterns
- Transaction accuracy

Current Challenges

Retail businesses face:

- Thin profit margins
- Revenue concentration risk
- Ineffective discount utilization
- Data recording inconsistencies
- Seasonal revenue volatility

Accurate and clean transactional data is essential for sustainable decision-making.

Why This Problem Was Chosen

This problem was selected because:

- Retail analytics has high real-world relevance
- Dirty datasets simulate operational challenges
- Revenue optimisation directly impacts business sustainability
- Data cleaning and dashboard creation reflect real-world analytics workflows

---

PROBLEM STATEMENT & OBJECTIVES

Formal Problem Definition

To analyze retail transaction data in order to identify revenue drivers, evaluate discount effectiveness, and uncover operational inefficiencies affecting sales performance.

Project Scope

The project includes:

- Cleaning and preparation of transactional data
- KPI development
- Trend and comparison analysis
- Dashboard design with interactive filters
- Business recommendation formulation

Excludes:

- Profit analysis (no cost data available)
- Customer demographic segmentation
- Supplier or logistics analysis

Success Criteria

The project is successful if:

- KPIs are clearly defined and measurable
- Revenue concentration is identified
- Discount effectiveness is evaluated
- Dashboard supports interactive decision-making
- Actionable business recommendations are provided

DATA DESCRIPTION

Dataset Source

Kaggle Dataset:
Retail Store Sales (Dirty Dataset for Cleaning Practice)
https://www.kaggle.com/datasets/ahmedmohamed2003/retail-store-sales-dirty-for-data-cleaning/data

---

Data Structure

Raw dataset:
12,575 rows × 11 columns

Columns include:

- Transaction ID
- Customer ID
- Category
- Item
- Price Per Unit
- Quantity
- Total Spent
- Payment Method
- Location
- Transaction Date
- Discount Applied

Data Limitations

- Missing price and quantity values
- Missing discount values
- Zero-value transactions
- No cost or margin data (profit analysis not possible)
- Dataset does not include customer demographics

DATA CLEANING & PREPARATION

All cleaning was performed in Google Sheets.

Missing Values Handling

- Missing Item → Replaced with “Unknown”
- Missing Discount → Assumed as “FALSE”
- Missing Price → Derived using Total Spent ÷ Quantity
- Missing Quantity → Derived where possible
- If derivation not possible → Set to 0

Data Issue Flag

A “Data Issue Flag” column was created.

- Transactions with Total Spent = 0 were flagged.
- These rows were retained for transparency but monitored to avoid KPI distortion.

Feature Engineering

Added:

- Transaction Month
- Day of Week

These enabled seasonality and operational trend analysis.

Assumptions

- Missing discount implies no discount applied
- Zero-value transactions indicate recording issues
- No inflation or market adjustment applied

KPI & METRIC FRAMEWORK

KPI	Definition	Formula	Business Relevance
Total Revenue	Total sales value	SUM(Total Spent)	Measures overall performance
Total Transactions	Count of transactions	COUNT(Transaction ID)	Measures sales volume
Avg Spend (Discounted)	Avg sales when discount applied	AVERAGEIF	Evaluates discount behavior
Top Category Share	Revenue concentration	Top Category Revenue ÷ Total Revenue	Identifies risk & dependency
Discount Impact Rate	Impact of discount on spend	(Avg Discount – Avg No Discount) ÷ Avg No Discount	Measures promotion effectiveness

---

EXPLORATORY DATA ANALYSIS

Trend Analysis

- Monthly revenue patterns show peak and low seasons.
- Line chart analysis reveals revenue variability across months.
- Day-of-week analysis reveals variability in demand.

Comparison Analysis

- Category comparison shows revenue concentration.
- Discount vs non-discount comparison reveals limited transaction growth.

Distribution Analysis

- Revenue is skewed toward top-performing categories.
- The majority of categories contribute moderate revenue levels.

Correlation Insights

- Weak relationship between discount and transaction frequency.
- Strong relationship between category and revenue.

ADVANCED ANALYSIS

Revenue Concentration

Few categories contribute disproportionately to total revenue.

Discount Strategy Evaluation

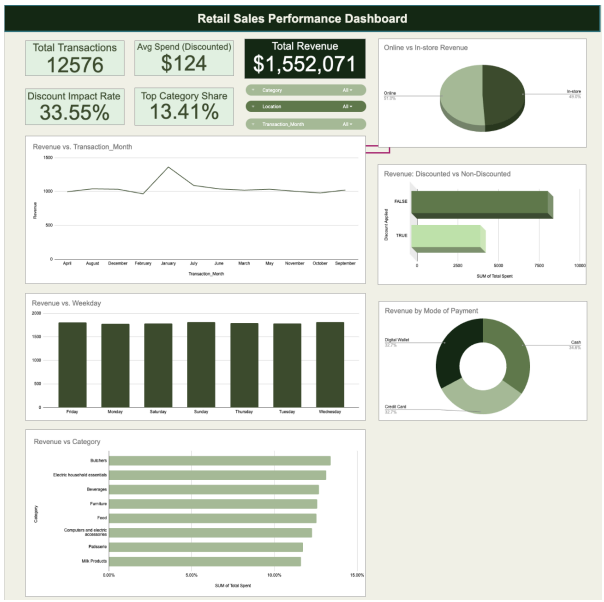
Discounts do not significantly increase transaction frequency or revenue.

Operational Insight

Zero-value transactions indicate data recording inefficiencies.

DASHBOARD DESIGN

Objective



Structure

Top Row:
KPI Cards

First Column:
Monthly Trend
Day-of-Week Performance
Revenue by Category

Second Column:
Online and In-store Comparison
Discount Comparison
Mode of Payment Trend

Filters & Drilldowns

Slicers added for:

- Category
- Month
- Location

Dashboard implemented in Google Sheets using pivots and slicers.

INSIGHTS SUMMARY

- Revenue concentrated in a few top categories.
- Discounts do not increase transaction frequency.
- Discount strategy does not significantly increase revenue.
- Monthly revenue fluctuations indicate seasonality.
- Certain days show higher demand patterns.
- Data inconsistencies affect reporting reliability.
- Revenue dependency on limited categories increases risk.
- Promotion timing can be optimized.

RECOMMENDATIONS

Insight	Recommendation	Business Impact	Feasibility
Revenue concentration	Focus inventory on top categories	Revenue stability	High
Discount inefficiency	Implement targeted promotions	Reduce margin loss	High
Seasonal fluctuations	Plan campaigns strategically	Smooth revenue	Medium
Data issues	Improve transaction recording controls	Better forecasting	High

IMPACT ESTIMATION

- 10–15% optimization in discount strategy may improve revenue efficiency.
- Better inventory planning reduces stockouts and overstock risk.
- Improved data quality enhances forecasting accuracy.

LIMITATIONS

- No cost data → Profit analysis unavailable
- Assumptions in derived values
- No customer behavior segmentation
- Dataset represents simulated retail environment

FUTURE SCOPE

- Incorporate profit margin analysis
- Add forecasting models
- Perform category clustering
- Include customer segmentation

CONCLUSION

This project demonstrates how structured data cleaning and analytical modeling can transform inconsistent retail transaction data into actionable business intelligence. The findings highlight revenue concentration risk, limited discount effectiveness, and operational inefficiencies. The interactive dashboard supports evidence-based decision-making, enabling retail managers to optimize revenue performance and operational efficiency.

CONTRIBUTION MATRIX

This section must clearly document the contribution of each team member across all project stages. Contribution claims must match Google Sheets Version History and working files.

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Niharika Choudhary	✓		✓	✓			**Dashboard Lead**
Hardik Shreyas	✓				✓	✓	**PPT**

Ravichandra shinde		✓	✓			✓	**Analysis Lead**
Arina Ali	✓			✓	✓		**Strategy Lead & Quality Lead**
Utkarsh			✓	✓		✓	**Project Lead**
Shivansh Upadhayay		✓	✓				**Data Lead**

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Signature Block: _____Section B - Group 10_____