# Supermarket Sales Data Analysis & Dashboard Integration

Author: Jahnavi Balla

Date: August 2025

## 1. Introduction

This project is an end-to-end sales data pipeline built using Streamlit, MySQL, and Power BI. It enables the collection, storage, cleaning, and visualization of supermarket sales data. The system allows the user to enter sales data directly through a web interface, store it in a database, and instantly update a Power BI dashboard for analysis. The main aim is to simplify sales entry while ensuring accurate and clean data for meaningful insights.

## 2. Objectives

- Create a user-friendly web interface for sales data entry.  
- Ensure data consistency and accuracy using automated cleaning and validation.  
- Store data securely in a MySQL database.  
- Provide actionable insights through an interactive Power BI dashboard.  
- Support real-time updates from new data entries.

## 3. Tools & Technologies

- Frontend/UI: Streamlit (Python)  
- Backend: MySQL  
- Data Processing: Pandas, SQLAlchemy  
- Visualization: Power BI  
- Database Connectivity: MySQL Connector, SQLAlchemy

## 4. Dataset

The dataset is adapted from the Kaggle Supermarket Sales Dataset, containing sales transactions from three branches:  
Branch A – Yangon  
Branch B – Mandalay  
Branch C – Naypyitaw  
The dataset includes details such as Invoice ID, Product Line, Quantity, Unit Price, Total, Tax, Payment Method, and Customer Ratings.

## 5. Methodology

### 5.1 Data Collection

The original dataset was imported from Kaggle. Data entry for new transactions is handled via the Streamlit web app, allowing multiple products under a single invoice.

### 5.2 Data Cleaning

Data cleaning is performed automatically during data entry:  
- Standardizing city, customer type, gender, and product names.  
- Mapping branch based on city.  
- Removing null and invalid values.  
- Recalculating derived metrics (Tax, Total, COGS, Gross Margin, Gross Income).

### 5.3 Database Integration

A MySQL database `supermarket\_db` stores all sales records. The Streamlit app inserts new data into the database via SQLAlchemy, ensuring secure and reliable data storage.

### 5.4 Dashboard Integration

Power BI connects directly to MySQL to create an interactive dashboard. The dashboard includes KPIs, branch-level sales, product category trends, and customer ratings. Users can refresh the dashboard to instantly include new sales data.

## 6. Features

- Multi-product invoices under a single Invoice ID.  
- Automatic calculation of Tax, Total, COGS, and Gross Margin.  
- Prevention of duplicate entries.  
- Option to delete mistakenly added products before submission.  
- Branch mapping from city names.  
- Power BI-ready data pipeline.  
- Real-time dashboard refresh to view updated insights.

## 7. Challenges

Several challenges were encountered during development:  
- Designing a database schema to handle multi-product invoices.  
- Ensuring data integrity when multiple products share the same invoice.  
- Avoiding duplicate data insertion.  
- Making the app intuitive for non-technical users.  
- Establishing seamless integration between MySQL and Power BI.

## 8. Results & Insights

The project successfully implements a functional sales data entry system with real-time analytics. Business users can now:  
- Quickly enter accurate sales data.  
- View real-time branch-level sales performance.  
- Identify top-performing product categories.  
- Monitor customer ratings and satisfaction trends.  
- Access updated sales metrics instantly via the Power BI dashboard.

## 9. Conclusion

This project demonstrates the implementation of a complete data pipeline from entry to visualization. The system ensures clean, validated data is always available for analysis, supporting informed decision-making. It can be extended to integrate predictive analytics, automate data cleaning for historical records, and expand to other retail domains.