

# **SQL Data Analysis Project — Grocery Sales Database**

## **1. Project Overview**

This project focuses on analyzing a grocery sales dataset using SQL. The goal is to clean the data, load it into MySQL, perform transformations, and generate meaningful business insights that can help improve company revenue and performance.

---

## **2. Objectives**

- Clean and prepare raw CSV files for database use
  - Design a relational database structure
  - Import large datasets using efficient SQL commands
  - Transform date/time fields into proper SQL formats
  - Perform analytical queries on sales, customers, products, and categories
  - Generate insights for business decision-making
- 

## **3. Tools & Technologies**

- **MySQL Workbench**
  - **SQL (DDL, DML, Joins, Window Functions)**
  - **Excel (Cleaning & Splitting Columns)**
- 

## **4. Database Tasks Performed**

### **✓ Data Cleaning**

- Removed empty rows in Excel
- Split combined Date-Time into separate columns
- Saved clean file as CSV

### **✓ Database Setup**

- Created Sales, Products, Employees, Customers, Categories, Cities, and Countries tables
- Added primary keys and foreign keys
- Defined proper data types for all columns

### **✓ Bulk Data Loading**

Used:

LOAD DATA LOCAL INFILE ...

to load thousands of rows quickly and handle data at scale.

### ✓ Data Transformation

- Converted SalesDate\_PART + SalesTime\_PART into a real DATETIME field
  - Deleted temporary columns after conversion
- 

## 5. Analytical SQL Work

### A. Basic Metrics

- Total Sales Revenue
- Total Number of Orders
- Average Order Value
- Top Products by Revenue & Quantity

### B. Customer Insights

- Top 10 Customers by Spending
- One-time vs Repeat Customers
- Customer Lifetime Value (CLV)
- Highest Revenue Cities & Countries

### C. Product & Category Insights

- Category-wise Revenue and Quantity
- High-revenue but low-quantity products
- Low-revenue but high-quantity products
- Best-performing product in each category

### D. Salesperson Performance

- Top & Bottom Salespersons
- Salesperson repeat customer handling
- Peak performance time for each salesperson

### E. Time-based Insights

- Best time range for sales (Morning / Afternoon / Evening / Night)
  - Sales trends based on customer behavior
- 

## 6. Key Business Insights (Summary)

- A few **top customers** generate the majority of revenue → focus on loyalty programs

- Certain **products and categories** consistently drive high revenue → promote and stock them
  - Some salespersons perform better during specific time ranges → optimize shift planning
  - Discounted orders often show higher quantity → smart discount strategy increases sales
  - Peak sales happen at certain hours → run marketing campaigns during those times
- 

## 7. Final Outcome

This project demonstrates:

- Strong SQL data cleaning skills
- Ability to design and manage relational databases
- Skill in writing analytical queries
- Understanding of business intelligence
- Capability to convert raw data into actionable insights