

Sales Data Analysis Using SQL

1. Introduction

This project focuses on analyzing sales data using MySQL to extract meaningful and actionable business insights. The goal is to demonstrate end-to-end SQL data analysis skills, starting from database creation and data loading to advanced querying and reporting. The project simulates a real-world sales analytics scenario commonly used in business intelligence and data analysis roles.

2. Objectives

The main objectives of this project are:

- To design a structured sales database using fact and dimension tables
 - To clean and transform raw sales data into analyzable formats
 - To perform sales, customer, product, and regional analysis using SQL
 - To apply advanced SQL concepts such as CTEs, window functions, and stored procedures
 - To generate insights that support data-driven business decisions
-

3. Database Design

The database follows a star-schema-like design, which is widely used in analytical systems.

3.1 Fact Table

- **fact_sales**: Stores transactional sales data such as order dates, product keys, customer keys, territory keys, and order quantities.

3.2 Dimension Tables

- **product**: Product details including product name, price, and cost
- **product_subcategory**: Subcategory information for products
- **product_category**: Category-level product information
- **customer_lookup**: Customer details such as name and birth date
- **territory**: Sales territories and regions

- **calendar**: Date-related information for time-based analysis
 - **returns_data**: Data related to returned products
-

4. Data Loading

Dimension tables were imported using the MySQL Table Data Import Wizard. The large fact table (fact_sales) was loaded using the LOAD DATA INFILE method to enable fast and efficient bulk data insertion from CSV files.

5. Data Cleaning and Transformation

To ensure data accuracy and consistency:

- Date columns were converted into proper DATE format using STR_TO_DATE()
 - Numeric columns such as product price and product cost were converted to appropriate data types (INT, DECIMAL)
 - SQL safe update mode was temporarily disabled where necessary to allow bulk updates
-

6. Data Analysis and Queries

Various SQL queries were written to answer real business questions:

- Total sales quantity per product
 - Total and regional sales revenue
 - Revenue by product category
 - Top-spending and repeat customers
 - Monthly sales comparison using LAG()
 - Ranking customers and products using RANK() and ROW_NUMBER()
 - Identifying customers who ordered from multiple territories
-

7. Advanced SQL Techniques Used

This project makes extensive use of advanced SQL concepts, including:

- JOINs to combine fact and dimension tables
 - CTEs (Common Table Expressions) for readable and modular queries
 - Window functions such as LAG, RANK, and ROW_NUMBER
 - Stored procedures for reusable and parameterized analysis
 - Aggregate functions such as SUM, COUNT, and AVG
-

8. Key Insights

Some of the important insights generated from the analysis include:

- Identification of top-selling products and categories
 - Recognition of high-value and repeat customers
 - Regional revenue performance comparison
 - Month-to-month sales trend analysis
 - Product return percentage analysis
-

9. Conclusion

This project demonstrates strong practical skills in SQL-based data analysis, including database design, data cleaning, advanced querying, and analytical thinking. The project reflects real-world sales analysis use cases and provides a solid foundation for roles related to data analysis, business intelligence, and database management.

10. Author

Al Fahim Fuyad

BSc in Computer Science & Engineering

East West University