

Specialization: *[Project focus]*

Business Focus: *[what industry]*

Tool: *[Tool]*

Capstone Project: Sales Intelligence Platform for Strategic Retail Decisions

Project Learning Opportunities

[A brief description of the case study- 2 line summary of case]

Tools and Technology to be Used

[Icons of Technologies used in the case study]

Case Study Overview

Introduction to the Business

A nationwide retail company wants to understand customer behavior, product performance, and regional sales patterns across multiple years. They aim to build a centralized analytics platform that can scale for insights, reporting, and optimization.



Case Study Overview

Problem Statement

Despite having large volumes of sales data, the retail company faces several operational and strategic challenges:

- Sales performance insights are fragmented across spreadsheets and reports, making them hard to reconcile.
- There's no unified data warehouse to support consistent and repeatable analysis across departments.
- Profitability tracking by product, region, or customer is difficult due to a lack of integrated models.
- Manual reporting processes are time-consuming, error-prone, and not scalable.
- Business leaders are unable to quickly identify high-performing categories, profit leakages, or customer behavior patterns.

Project Components

✓ 1. Database and Data Warehouse Modeling

- Business Requirements Gathered:
 - Track profit and loss by category, city, and customer.
 - Measure order frequency, high-profit sub-categories, and peak months.
 - Enable dimensional modeling for easier BI integration.
- Schema Design:
 - Operational DB (Normalized): Orders, Customers, Products, Payments, Locations
 - Data Warehouse (Denormalized):
 - FactSales: holds metrics like Amount, Profit, Quantity.
 - DimCustomer, DimProduct, DimDate, DimLocation, DimPayment.
- ERD & Transformations:
 - Built using foreign keys (Order ID, Customer Name, City) and surrogate keys (CustomerID, ProductID)
 - Cardinalities captured: 1-Many for Customers → Orders, Products → Orders
 - Export ERD to GitHub with SQL schema
 -

Project Components

2. Data Cleaning & Analysis with Pandas

- Key Operations:
 - Read and clean:
- Transformations:
 - Aggregated monthly profits, top customers, quantity sold per category
 - Filtered cities with low profitability
 - Wrote cleaned tables to CSV:
- Created tables:
 - `customer_df`, `product_df`, `location_df`, `sales_fact_df`, `date_dim_df`

Project Components

3. SQL + Python + Pandas (End-to-End Analysis)

- **Joins:** Join FactSales with DimLocation and DimProduct to extract insights.
- **Ranking & Window Functions**
- **CASE**
- **Aggregations & Subqueries:**
 - Avg. monthly sales per product
 - Cities with revenue > 95th percentile

Project Components

- . Database Access via Python (Serverless Option)
 - Use PostgreSQL
 - Supported:
 - Data insertion (`to_sql()` via SQLAlchemy)
 - Update/Remove old records
 - Auto-refresh table from cleaned CSV



Deliverables

Artifact Description

`ERD.png`

Entity Relationship Diagram

`schema.sql`

SQL scripts for schema creation

`data_cleaning.ipynb`

Jupyter notebook for data cleaning

`load_to_db.py`

Python script to insert data into PostgreSQL

`advanced_sql_queries.sql`. Collection of CTEs, window functions, rankings

Data Description

The dataset captures multi-year retail sales transactions with key attributes such as order details, customer information, product categories, financial metrics (amount and profit), and payment methods. It includes 12 columns and supports insights into product performance, regional sales trends, customer segmentation, and profitability analysis. The dataset is well-suited for building a complete data pipeline involving modeling, transformation, and analytics using SQL and Python.

READY TO DELVE IN?

