

E-Commerce Case Study Outline

Prepared by Mckenly Chau – Aspiring IT & Data Analytics Professional with 9+ Years of Leadership Experience in Security, Risk Management, and Customer Relations

Overview

This SQL-based project delivers a complete analytics solution for an e-commerce retail dataset, transforming over 500,000 raw transactional records into structured, insight-ready tables using MySQL.

Objective

Analyze e-commerce transaction data to uncover customer purchasing behavior, identify key trends, and generate actionable insights to support strategic decision-making in marketing, retention, and product planning.

License

This project is for educational and portfolio purposes only. The original e-commerce dataset is credited to the UCI Machine Learning Repository and Kaggle contributors, with all rights retained by the original data owners.

Additional public data sources include the [UN M49 Country and Area Classifications](#) (UN Statistics Division, public domain). All external data has been transformed for analysis, with proper attribution provided. No proprietary data is distributed, and all licensing terms are observed.

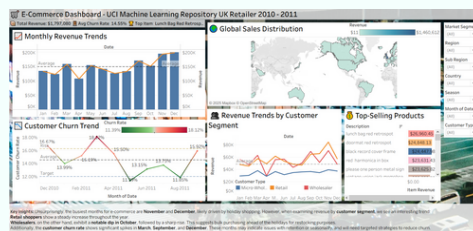
Features

Cleaned Database Schema

Structured and cleaned transactional database featuring Recency, Frequency, and Monetary (RFM) segmentation for customer classification.

Tableau Preview

Interactive Tableau dashboard with end-user usability in mind.



Executive Summary

Strategic insights backed by actionable recommendations for marketing, retention, and product planning.



Key highlights:

- End-to-End Data Pipeline: Imported, cleaned, and transformed real-world e-commerce data, addressing issues like duplicates, missing values, zero-price items, and refund anomalies.
- Robust Data Cleaning & Modeling: Applied surrogate keys, index optimization, and data type normalization to create a high-performance, relational database schema.
- Classified customers into Retail, Micro-Wholesaler, and Wholesaler segments.
- RFM-Ready Customer Segmentation: Engineered customer-level features—recency, frequency, and monetary value—to support CLV estimation and advanced segmentation, laying the groundwork for targeted marketing or retention strategies.
- Calculated Customer Lifetime Value (CLV) using historical and estimated models.
- Temporal & Seasonal Analysis: Enriched the dataset with monthly, seasonal, and region-level breakdowns to uncover trends in sales volume and customer behavior.
- Refund Matching Algorithm: Built logic to detect and classify full and partial refunds, improving accuracy in revenue and retention analysis.
- Insight-Ready Summary Tables: Delivered clean, well-indexed tables for invoices, customers, products, countries, and dates, ready for visualization in tools like Tableau or Excel.