Sales Performance & Profitability Analytics Project

This document explains the complete end-to-end data analytics workflow followed in this project, from raw CSV ingestion to analytics-ready data modeling, exactly as done in an industry environment.

Step 1: Business Understanding

Objective: Analyze sales performance and profitability to support business decision-making. Key metrics include revenue, profit, profit margin, growth trends, top products, and underperforming regions.

Step 2: Raw Data Ingestion

1. CSV file imported into MySQL as a RAW table.
2. All columns stored as TEXT to preserve source integrity.
3. No transformations applied at this stage.

Step 3: Data Cleaning & Validation (ETL)

1. Created a clean table with correct data types (INT, DATE, DECIMAL).
2. Converted date formats using STR_TO_DATE.
3. Removed commas and invalid characters from numeric fields.
4. Used REGEXP validation to prevent ETL failures.
5. Invalid values converted to NULL instead of failing the pipeline.

Step 4: Data Quality Checks

1. Row count validation between raw and clean tables.
2. NULL analysis for sales, profit, and dates.
3. Negative profit analysis to identify loss-making orders.

Step 5: Exploratory Data Analysis (SQL)

1. Monthly and yearly sales trends.
2. Top and worst-performing products.
3. Regional and category-level profitability analysis.
4. Profit margin calculations.

Step 6: Data Modeling (Star Schema)

1. Created FACT table for sales transactions.
2. Created DIM tables: Customer, Product, Region, Date.
3. Prepared model for efficient JOINs and BI usage.

Step 7: BI & Dashboard Readiness

1. Data structured for Power BI import.
2. Star schema enables faster performance and simpler DAX.

3 Insights used to support business recommendations.

Outcome: This project demonstrates an industry-standard data analytics workflow including ETL, data validation, SQL-based analysis, dimensional modeling, and BI readiness.