**Project Report: Item Sales Analysis**

**Project Overview**

This project demonstrates a complete **ETL (Extract, Transform, Load)** workflow using **MySQL Workbench**. The data source is a CSV file containing detailed item sales records from a convenience store. The primary objective is to clean, transform, and analyze the sales data to uncover insights such as revenue, profit, discount impact, and inventory levels.

**ETL Breakdown**

**1. Extract**

The raw data is stored in a CSV file. It is imported into a **staging table** named Staging\_ItemSales using MySQL’s CSV import tools.

**2. Transform**

The raw data contains currency symbols ($), percentage signs (%), and inconsistent data formats. These are cleaned using SQL functions like REPLACE() and converted into proper numeric types.

Key transformations:

* Remove $ and % from monetary and ratio fields.
* Convert string values to DECIMAL.
* Handle invalid or percentage-formatted profit values.
* Rename and normalize columns for consistency.

**3. Load**

The cleaned data is inserted into a structured final table Final\_ItemSales, which is used for analysis. Auto-incremented id is added for row identification.

**Database Schema**

**Staging Table: Staging\_ItemSales**

* Contains raw data with fields like Customer Name, # Sold, Price, Cost, Profit, etc.
* Data types are mostly VARCHAR to allow import without type conflicts.

**Final Table: Final\_ItemSales**

* Contains cleaned and converted data.
* Key fields: UnitsSold, Cost, Price, Profit, StockLevel, Discounts, RefundedAmount, etc.
* Types are defined as DECIMAL for financial accuracy.

**Key Analysis & Insights**

**Sales Insights**

* **Top 10 & 15 Selling Items**: Identifies best-performing products by quantity sold.
* **Total Revenue & Profit**: Provides the overall financial performance.
* **Revenue by Category**: Highlights product categories contributing most to sales.

**Inventory Insights**

* **Low Stock Alerts**: Finds items with stock levels under 10 for restocking.
* **Overstock**: Finds items with stock > average or > 300.
* **Stock Value Estimation**: Calculates the monetary value of current inventory.

**Discounts & Refunds**

* **Total Discounts**: Summarizes all discounts provided.
* **Category-wise Discounts**: Reveals which categories are most discounted.
* **Refunded Items**: Lists products with refunds and their refunded amounts.

**Profitability**

* **Category-Wise Profit**: Evaluates which categories are most profitable.
* **Margin Distribution**: Groups records by profit margins to assess pricing strategy.
* **Highest Margin Items**: Lists products with the highest individual margins.

**Supplier Analysis**

* Compares suppliers by total revenue and profit to evaluate vendor performance.

**Post-ETL Cleanup**

* The Staging\_ItemSales table is dropped after the transformation and loading to free up space and avoid duplication.

**Conclusion**

This ETL project:

* Successfully imported raw CSV data.
* Transformed it into a clean and analysis-ready format.
* Produced meaningful business insights on inventory, sales, revenue, discounts, and supplier performance.