# **📑 Project Documentation – Retail Sales ETL & BI Project**

## **1. Project Overview**

The objective of this project is to design and implement an **ETL pipeline** for a Retail Sales dataset and build a **Data Warehouse (DW)** to enable **Business Intelligence (BI) reporting**.

* **Source Data**: Raw CSV files containing sales transactions, products, customers, and store details.
* **Tools Used**:  
  + **Python (Pandas, SQLAlchemy)** → for data extraction, transformation, and loading.
  + **Oracle SQL** → for staging and data warehouse schema.
  + **Power BI**  → for dashboards and insights visualization.

## **2. ETL Workflow**

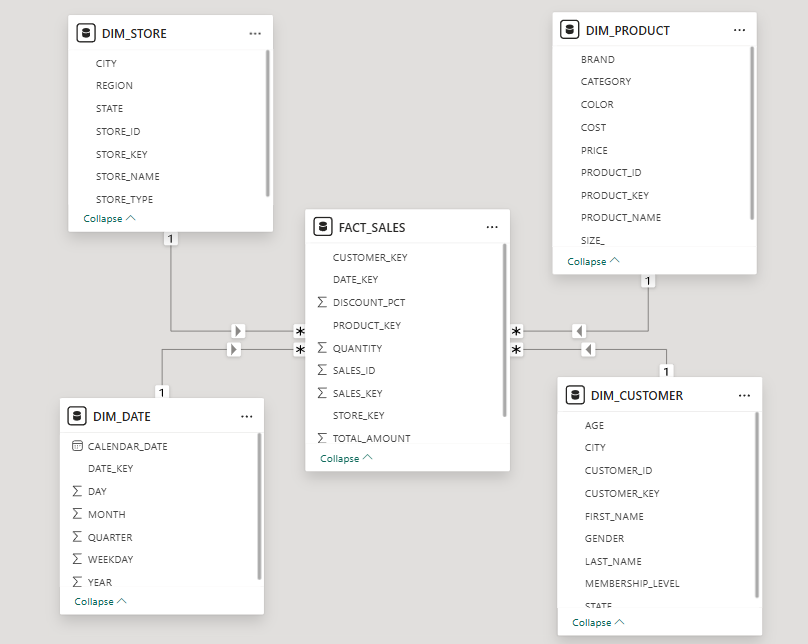
The project follows a 4-layered pipeline:

1. **Extraction**
   * Source CSV files (sales, products, customers, stores).
   * Loaded into Python Pandas DataFrames.
2. **Transformation**
   * Data cleaning: removed duplicates, handled missing values, corrected data types.
   * Business rules:  
     + Standardized product categories.
     + Converted dates into proper formats.
     + Calculated derived fields (e.g., revenue = quantity × unit\_price, profit = revenue – cost).
3. **Loading (Staging → DW)**
   * **Staging Layer**: Raw tables (same structure as source).
   * **Data Warehouse Layer**: Star schema with **FactSales** table and **Dimensions** (Product, Customer, Store, Date).
4. **BI & Reporting**
   * Connected DW tables to **Power BI**.
   * Built dashboards for Sales, Profit, Regional Performance, and Product Trends.

## **3. SQL Schema Design**

### **ER Diagram – Star Schema**

* **Fact Table**:  
  + FactSales
* **Dimension Tables**:  
  + Dim\_Date
  + Dim\_Product
  + Dim\_Customer
  + Dim\_Store



## 

## 

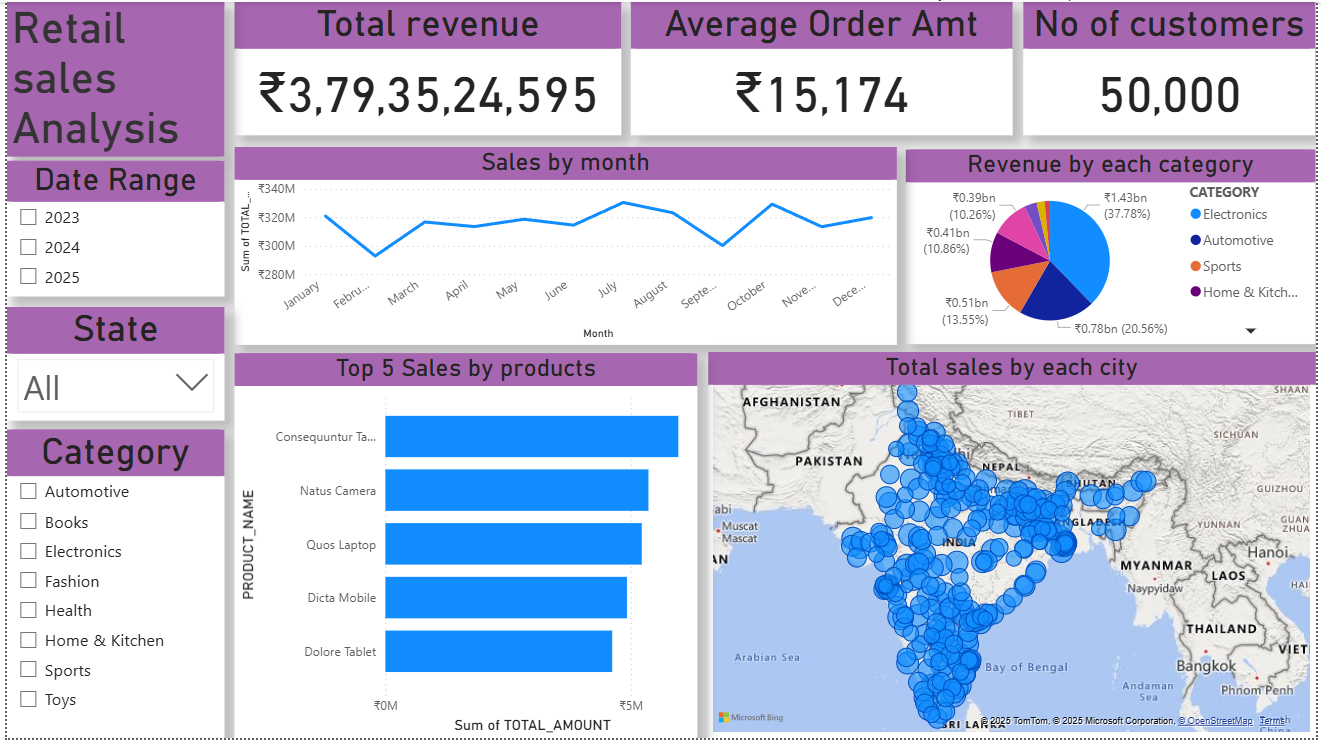
## 

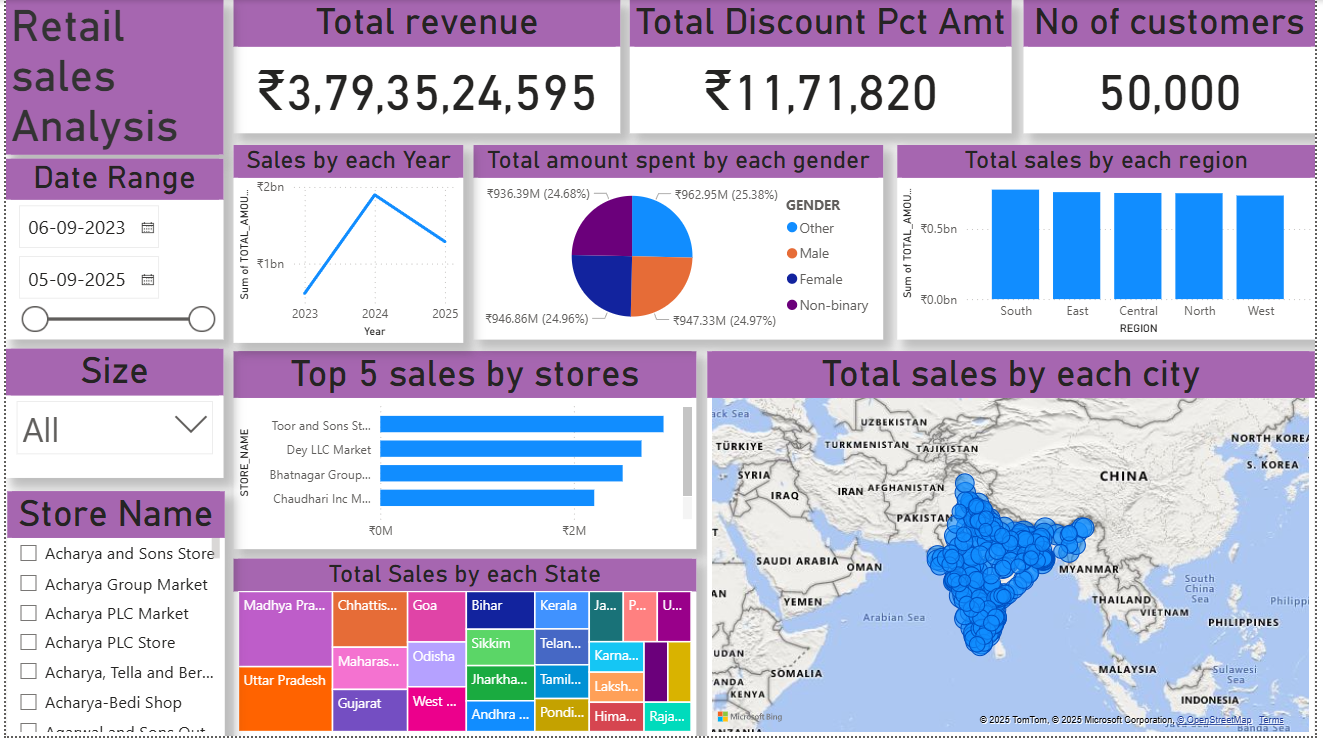
## 

## 

## 

## **4. Key SQL Queries & Insights**





## 

## **5. Dashboards & Findings**

### **Dashboards Created**

1. **Sales Overview**
   * **Total Revenue**: ₹3,79,35,24,595
   * **Average Order Amount**: ₹15,174
   * **Total Customers**: 50,000
2. **Sales by Month**
   * Trend chart shows consistent revenue across months with noticeable peaks in **July, August, and October**.
   * Slight dips in **February** and **September** indicating seasonal variations.
3. **Revenue by Category**
   * **Electronics** is the highest revenue contributor (37.78%, ₹1.43B).
   * Followed by **Home & Kitchen** (20.56%, ₹0.78B).
   * **Sports** and **Automotive** contribute around 10–13% each.
4. **Top 5 Products by Sales**
   * Consequuntur Tablet
   * Natus Camera
   * Quos Laptop
   * Dicta Mobile
   * Dolore Tablet
   * These products together represent the **highest-grossing items** in the portfolio.
5. **Sales by Geography**
   * A geographical heat map of India shows **widespread sales coverage across major cities**.
   * Strong concentration of sales in **urban clusters such as Delhi, Mumbai, Bengaluru, Hyderabad, and Chennai**.

### **Key Findings**

* **High Revenue**: Retail sales generated nearly **₹379 Crores**, with an **average ticket size of ~₹15K per customer**.
* **Customer Base**: The platform has **50,000 unique customers**, reflecting strong market penetration.
* **Category Insights**: Electronics and Home & Kitchen dominate the product mix, contributing more than **58% of total revenue**.
* **Product Leaders**: Tablets, laptops, and cameras are the **top-selling products**, indicating strong demand for technology-driven items.
* **Regional Insights**: Sales are distributed across the country, with **Tier-1 cities leading in contribution**.
* **Seasonal Trends**: Sales spikes in festive/holiday months (July–Aug, Oct–Dec), which aligns with consumer buying behavior.

## **6. Deliverables**

* **Code**: Python ETL scripts (etl.py).
* **Data**: Raw CSV files (Sales, Products, Customers, Stores).
* **SQL Scripts**: Table creation, ETL load queries, fact/dimension design.
* **Dashboards**: Power BI reports with insights.
* **Documentation**: This file + PPT (5–6 slides).