# Data Analyst Internship Project: Retail Sales Performance Analysis

## 📌 Objective

This project is designed to simulate a real-world Data Analyst role. By completing it, you will gain hands-on experience with SQL, Python, and Power BI. The goal is to analyze retail sales data, discover insights, and build a professional dashboard.

## 📂 Datasets

1. Superstore Sales Dataset  
 - Source: Kaggle (https://www.kaggle.com/datasets/vivek468/superstore-dataset-final)  
 - Columns: Order ID, Order Date, Ship Date, Customer ID, Product ID, Category, Sub-Category, Sales, Quantity, Discount, Profit, Region  
  
2. Customer Personality Analysis Dataset  
 - Source: Kaggle (https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis)  
 - Columns: CustomerID, Age, Education, Marital Status, Income, Spending Score, Region

## 📝 Tasks

### Part 1 – SQL: Data Storage, Cleaning & Querying

- Load both datasets into a SQL database (MySQL/PostgreSQL).  
- Clean the data: remove duplicates, handle missing values, format dates.  
- Create SQL queries to:  
 • Calculate monthly sales and profit by region  
 • Find top 10 products by sales  
 • Count number of orders per customer  
 • Join both datasets on Region or CustomerID to analyze demographic-based trends

### Part 2 – Python: Data Analysis & EDA

- Use Python (pandas, numpy, matplotlib/seaborn) to:  
 • Connect to the SQL database and pull cleaned tables  
 • Perform Exploratory Data Analysis (EDA)  
 • Detect outliers and anomalies in sales/profit  
 • Calculate KPIs (average sales per customer, profit margin, monthly growth)  
 • Plot: Monthly Sales Trend, Profit Distribution, Sales vs Discount, Region-wise performance

### Part 3 – Power BI: Dashboard Building

- Build an interactive dashboard showing:  
 • Sales by Region (map/bar chart)  
 • Monthly Sales Trend (line chart)  
 • Top Products by Sales (bar chart)  
 • Profit vs Discount scatter plot  
 • Filters for Year, Region, and Category

## 📁 Final Deliverables

Submit the following items:  
1. SQL Scripts (.sql)  
2. Python Jupyter Notebook (.ipynb)  
3. Power BI Dashboard (.pbix)  
4. Summary Report (.pdf/.docx) explaining insights and recommendations

## ✅ Evaluation Criteria

- Data cleaning and preparation accuracy  
- Quality of SQL queries and joins  
- Depth of EDA and statistical insights in Python  
- Visual design and interactivity of Power BI dashboard  
- Ability to communicate insights clearly and concisely

## 💼 Industry Relevance

By completing this project, you will develop industry-grade skills in data wrangling, analysis, and visualization. These are core skills tested in most entry-level data analyst interviews and will strengthen your portfolio for real-world jobs.