**Global Superstore SQL Analysis**

This project explores sales performance, profitability, and return-related losses using SQL and the Global Superstore dataset. The analysis focuses on identifying trends across countries, regions, and individual salespeople using structured queries.

**Project Objective**

The objective of this project is to extract meaningful insights from a retail dataset using SQL. It covers:

* Identifying high and low performing countries
* Analyzing annual and regional sales trends
* Calculating profitability per unit sold
* Quantifying lost profit from returned items
* Applying advanced SQL concepts like joins, aggregation, and conditional logic

**Tools and Technologies**

* SQLite 3 (executed via VSCode)
* CSV for data export
* Excel for data inspection and formatting
* Git & GitHub for version control

**Output File Description**

1. **Annual\_Sales\_Profit\_by\_Region\_and\_Person.csv**

* Summarizes annual sales performance by region and individual salesperson. Includes:
  + Quantity of sales
  + Total sales revenue
  + Units sold
  + Revenue per unit
  + Total profit
  + Profit per unit

1. **Countries\_Negative\_Profit\_or\_Margin.csv**

* Highlights countries with poor financial performance, based on:
  + Negative total profit
  + Negative profit per unit

1. **Lost\_Profit\_by\_Region.csv**

* Joins Orders, People, and Returned tables to calculate:
  + Number of returned items
  + Total profit lost
  + Lost profit per item Grouped by region and salesperson.

**How to Recreate the Analysis**

1. Open Global\_Superstore\_Code.sql in VSCode with SQLite 3 installed.
2. Run commands in order:
   1. Create and populate the database from the /data CSV files
   2. Execute each analysis query from parts F–M
   3. Export selected outputs using .output commands
3. Output CSVs will match those stored in the /output folder for validation.

**Key Insights**

* The U.S. leads in volume but also suffers significant profit loss from returns.
* Some countries consistently show negative profit margins — potential red flags for pricing or logistics.
* Salespeople in certain regions outperform others in profit per unit, revealing opportunities for benchmarking.

**Future Enhancements**

* Automate pipeline using Python (e.g., pandas + sqlite3)
* Visualize results with Tableau or Power BI dashboards
* Add parameterized queries for dynamic filtering (e.g., by region or year)
* Build a Streamlit app for interactive analysis