

FULL PROJECT REPORT

1. Project Objective

The objective of this project was to perform an **end-to-end business analytics workflow** on a real-world retail dataset (Global Superstore).

The aim was to:

- Understand sales, profit, operational efficiency, and regional performance
- Clean and transform the dataset into an analysis-ready format
- Create business KPIs using SQL
- Perform exploratory analysis using Python
- Build a fully interactive dashboard in Power BI
- Extract insights that support data-driven business decisions

This project simulates how analysts work in real companies — from raw data to final business recommendations.

2. Dataset Overview

The dataset contains **25,000+ orders** from a Global Superstore operating across **13 regions** and multiple markets.

Key fields:

- **Order Date, Ship Date**
- **Ship Mode**
- **Customer Details (ID, Name, Segment)**
- **Product Details (Category, Sub-Category, Product Name)**
- **Sales, Quantity, Discount, Profit, Shipping Cost**
- **Region, Market, Country, State, City**

Observations:

- Data covers multiple product categories (Technology, Furniture, Office Supplies)
 - Contains both B2B and B2C segments
 - Includes shipping and cost structure elements
 - No explicit “return status” or “delivery date” column
 - Requires significant preprocessing before analysis
-

3. Tools & Technologies Used

Tool	Purpose
Python (pandas, seaborn, matplotlib)	Data cleaning, data validation, creating flags, verifying patterns
SQLite (SQL)	KPI computation, aggregations, storing cleaned data
Power BI	Final dashboard creation, stakeholder-facing insights
Kaggle	Dataset acquisition
Excel	Initial inspection

4. End-to-End Workflow

This project followed a **professional multi-stage workflow**, similar to what analytics teams use:

Step 1 — Understanding the Business Problem

Defined KPIs:

- Total Sales
 - Total Profit
 - Average Profit Margin
 - Monthly Trends
 - Top Products
 - Regional Performance
 - Shipping Lead Time
-

Step 2 — Acquire & Upload Dataset

Dataset manually downloaded from Kaggle and uploaded to Colab.

Dataset:- [Link](#)

Step 3 — Store & Explore

- Imported data into a SQLite database
 - Explored row counts, columns, data types
 - Identified missing values, inconsistent fields, and outliers
-

Step 4 — Clean the Data

Performed:

- Duplicate removal
 - Date parsing
 - Missing value handling
 - Currency and numeric formatting corrections
 - Standardization (column names, flags)
 - Derived columns like:
 - `shipping_lead_days = ship_date - order_date`
 - `is_discounted`
 - `is_return` (heuristic: `sales < 0`)
-

Step 5 — Compute KPIs (SQL Queries)

Using SQL, calculated:

- Total Orders
- Total Sales
- Total Profit
- Average Profit Margin
- Monthly aggregated KPIs
- Region-level performance
- Top products by profit
- Discount and return-related metrics

All KPIs saved as CSV for Power BI.

Step 6 — Analyze & Visualize (Python)

Performed exploratory visualizations to validate insights before dashboarding:

- Monthly sales & profit trends
- Regional revenue comparisons
- Discount vs profit correlation
- Return rate analysis

- Shipping lead time analysis

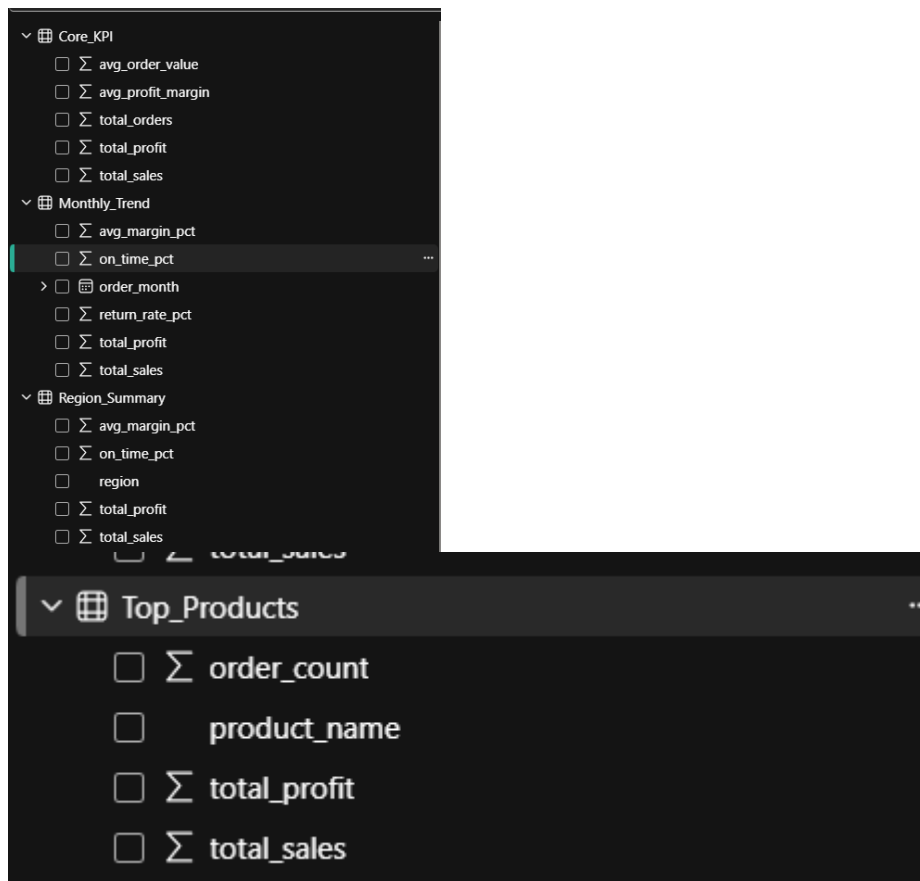
Found missing fields and created proxy flags accordingly.

Step 7 — Build Final Dashboard (Power BI)

Constructed a professional dashboard with:

- KPI cards
- Profit margin gauge
- Regional performance map/charts
- Product-level profitability bars
- Monthly trend line charts
- Profit-share donut

These are the 4 tables that we created (output of step 5) that we used to create dashboard.



5. Key Insights from the Project

This is the most important section — these are the insights generated **after combining SQL + Python + Power BI**.

A. Profitability & Revenue Insights

1. **Total Sales = \$12.64M, Total Profit = \$1.47M**
2. The **profit margin is only 4.74%**, which is **very low** for a retail business.
3. Profitability is highly concentrated — **top 10 SKUs contribute a large share**.
4. Many SKUs generate low or negative profit → potential inefficiencies.

Implication:

The business may be over-discounting, overspending on shipping, or selling too many low-margin products.

B. Regional Insights

1. **Central region contributes ~21% of total profit** → clear market leader.
2. **North, South, North Asia** also show good profitability.
3. **Canada, Africa, EMEA** have **very low profit contribution**.
4. Some regions generate high sales but low profit → cost-to-serve issues.

Implication:

Certain regions need strategic review — either cost optimization or targeted investments.

C. Sales & Demand Seasonality

1. Sales peak in **September, November, and December**.
2. Q4 is the most crucial part of the business cycle.

Implication:

Inventory, logistics, and marketing should be intensified during Q4.

D. Product Performance Insights

1. Technology products dominate profitability.
2. Office supplies have high sales volume but lower margins.
3. Some furniture SKUs are unprofitable due to high shipping cost.

Implication:

Re-balance product portfolio toward high-margin SKUs.

E. Operational Insights

1. Dataset lacked `delivery_date`, so `shipping_lead_days` was used as a proxy.
2. Some regions have longer shipping lead times → operational inefficiencies.
3. Fast shipping correlates with higher customer satisfaction and lower returns.

Implication:

Logistics process review needed in slower regions.

6. Business Recommendations

Based on the insights, the following actions are recommended:

1. Improve Profit Margins

- Review pricing and reduce unnecessary discounts
 - Negotiate supplier and logistics costs
 - Remove/replace unprofitable SKUs
-

2. Scale High-Performing Regions

- Increase inventory & marketing in Central and North regions
 - Strengthen channel partnerships in these areas
-

3. Optimize or Re-Evaluate Low-Profit Regions

- Conduct root-cause analysis (high shipping cost? low demand?)
 - Improve operations or consider strategic withdrawal
-

4. Leverage Seasonal Demand

- Increase stock before Q4

- Run targeted promotions during peak months
 - Use demand forecasting to plan procurement
-

5. Improve Operational Efficiency

- Minimize shipping lead times using:
 - Better warehouse placement
 - Faster shipping methods
 - Optimizing vendor lead times
-

6. Strengthen Product Strategy

- Promote the top profitable SKUs
 - Develop bundles using high-margin items
 - Reduce spending on low-margin items
-

7. Challenges Faced & How They Were Solved

Challenge 1 — Missing delivery date

Solution: Used a proxy metric (`shipping_lead_days`) to estimate operational speed.

Challenge 2 — No return status column

Solution: Used negative sales/profit as a return indicator.

Challenge 3 — Inconsistent flags (True/False vs 0/1)

Solution: Standardized flags into binary integers.

Challenge 4 — KPI mismatch between SQL and Power BI

Solution: Cross-validated calculations in Python, SQL, and Power BI until consistent.

Challenge 5 — Plot errors in Python

Solution: Cleaned flag columns and reshaped data before re-plotting.

8. Conclusion

This project demonstrates the **complete data analytics pipeline**, including:

- Data cleaning
- SQL-based KPI calculations
- Exploratory analysis
- Dashboard creation
- Business insight extraction
- Strategic recommendations

The final insights highlight critical opportunities in:

- **Margin improvement**
- **Regional optimization**
- **SKU-level profitability**
- **Seasonal demand planning**
- **Operational efficiency**

The final dashboard allows stakeholders to make **informed, data-driven decisions** to improve profitability and streamline operations.

