Innovation & Entrepreneurship Hub for Educated Rural Youth (SURE Trust – IERY)

GLOBAL SUPERSTORE DATASET

The domain of the Project:

SQL AND PowerBI

Team Mentors (and their designation):

SRAVAN NEMANA(Co-Founder of stealth AI Startup)

Team Members:

Ms.VINEELA

Period of the project

July 2025 to December 2025

Declaration

The project titled “Global SuperStore Dataset” has been mentored by N.Sravan, organised by SURE Trust, from July 2025 to December 2025, for the benefit of the educated unemployed rural youth for gaining hands-on experience in working on industry relevant projects that would take them closer to the prospective employer. I declare that to the best of my knowledge the members of the team mentioned below, have worked on it successfully and enhanced their practical knowledge in the domain.

Team Members:

Ms. Vineela
Signature

Mentor’s Name: N.Sravan
Co-Founder—Stealth AI Startup

Prof. Radhakumari
Executive Director & Founder
SURE Trust

Table of contents

Executive Summary

The Global Superstore Dataset Project aims to analyse retail sales data to identify key drivers of sales, profit, and customer performance across regions and product categories. The primary objective is to support data-driven business decisions by uncovering trends, inefficiencies, and growth opportunities within the organization.

The project utilizes data cleaning, exploratory data analysis, and visualization techniques to evaluate sales performance, profitability, customer segments, shipping modes, and discount strategies. Analytical methods were applied to compare regional performance, identify top- and bottom-performing products, and assess the impact of discounts and shipping on profitability.

Key findings reveal that while certain regions and product categories generate high sales, they do not always yield high profits due to excessive discounting and shipping costs. Consumer segments and specific product sub-categories were identified as major contributors to overall profitability, whereas others require strategic improvement. Additionally, shipping mode efficiency varies by region, influencing delivery costs and customer satisfaction.

Based on these insights, the project recommends optimizing discount strategies, focusing on high-margin products and profitable regions, and improving logistics planning to enhance operational efficiency. Overall, the analysis provides actionable insights that enable executives and stakeholders to make informed decisions to improve revenue growth and profitability.

Background and Context

The Global Superstore Dataset represents a multinational retail business operating across multiple regions, product categories, and customer segments. The dataset contains detailed transactional records, including sales, profit, discounts, shipping modes, and customer information. In today's competitive retail environment, organizations rely heavily on data analytics to understand customer behaviour, optimize operations, and improve profitability. This project uses the Global Superstore dataset as a real-world case to simulate business analysis and decision-making in a global retail context.

Problem Statement / Project Goals

The primary goal of this project is to analyse the Global Superstore dataset to uncover meaningful insights related to sales performance, profitability, customer segments, and regional trends. The project aims to identify key factors affecting profit, evaluate the effectiveness of discount and shipping strategies, and determine high- and low-performing products and regions. Ultimately, the project seeks to support data-driven decision-making by providing actionable recommendations to improve business performance.

Scope and Limitations

Scope:
The scope of this project includes data cleaning, exploratory data analysis, and visualization of historical sales data from the Global Superstore dataset. Analysis is focused on sales, profit, discounts, customer segments, regions, and shipping modes. The project evaluates trends and patterns within the available dataset to derive business insights.

Limitations:
The analysis is limited to the data provided in the Global Superstore dataset and does not include real-time or external data such as market conditions, competitor analysis, or customer feedback.

Innovation Component

The innovative aspect of this project lies in the integration of business-oriented analytics with data visualization to translate complex datasets into clear, actionable insights. Instead of focusing solely on descriptive statistics, the project emphasizes insight-driven analysis by correlating discounts, shipping modes, and regional performance with profitability.

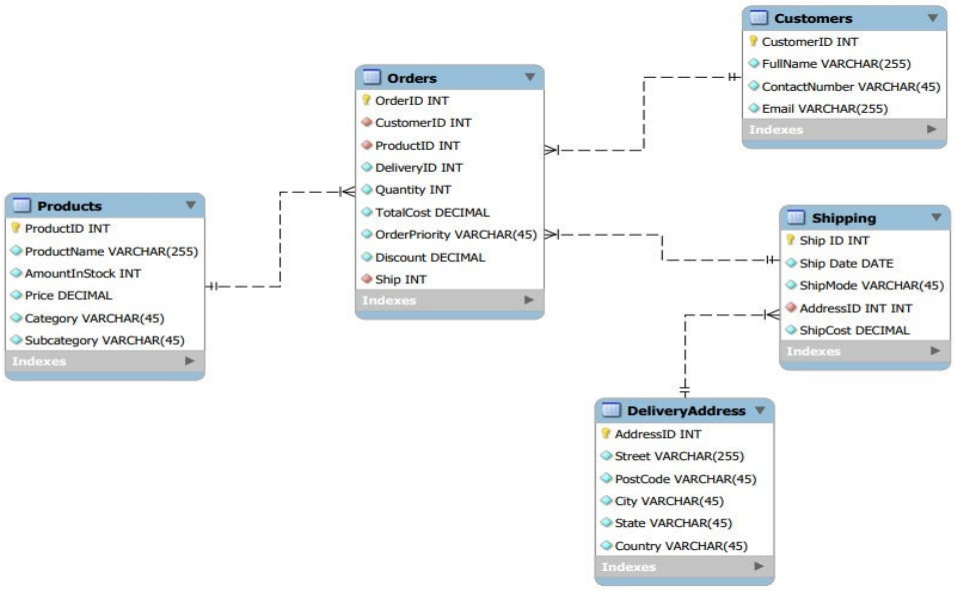
Project Objectives

The Global Superstore Dataset Project is designed to analyse retail transaction data with the objective of deriving meaningful insights that support data-driven business decision-making. The project focuses on understanding sales and profit trends across different regions, product categories, and customer segments. Key goals include cleaning and pre-processing the dataset to ensure accuracy, identifying high-performing and underperforming products and regions, evaluating the impact of discounts and shipping modes on profitability, and presenting insights through clear and effective visualizations.

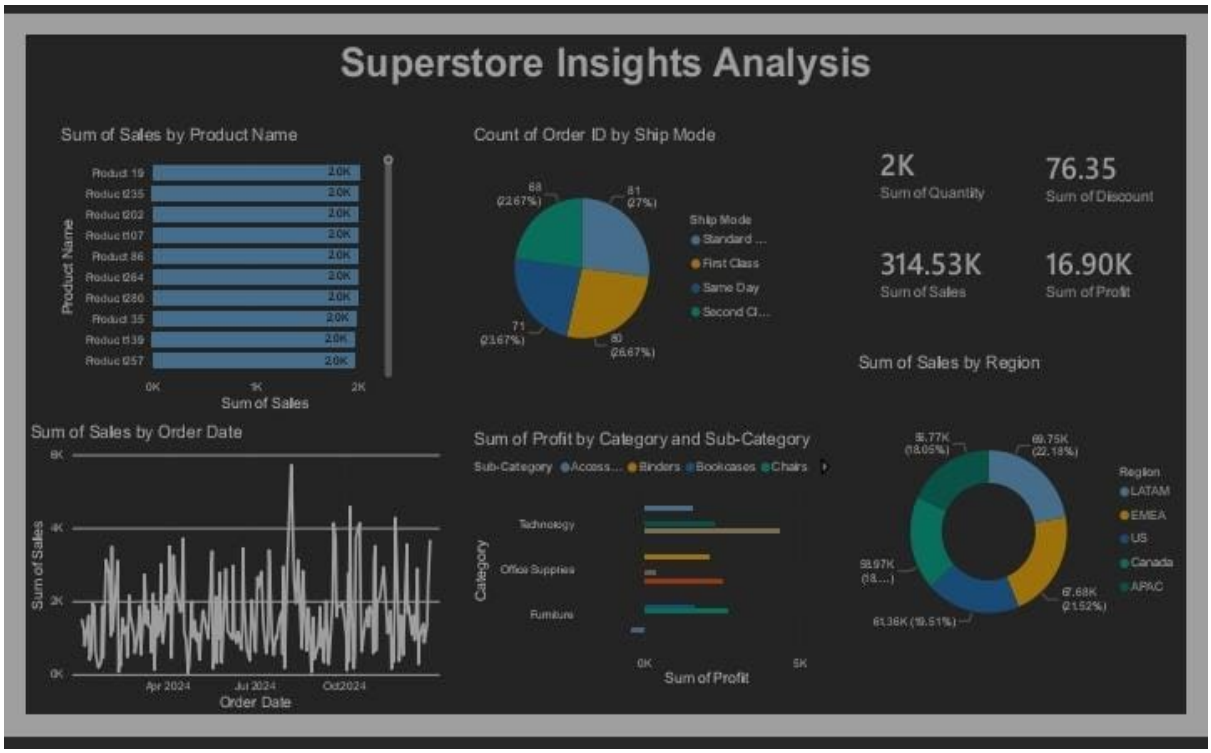
The expected outcomes of the project include a comprehensive understanding of the factors influencing sales and profitability, identification of profitable regions and customer segments, and insights into operational inefficiencies caused by discounting and shipping strategies. The main deliverables of the project consist of a cleaned and well-structured dataset, analytical findings supported by visualizations or dashboards, actionable business recommendations, and a final project report summarizing the analysis and results.

Methodology and Results

- **Methods / Technology Used:**
  The project employed structured data analysis methods including data cleaning, data pre-processing, exploratory data analysis, and descriptive analytics. SQL was used for querying and aggregating data, Python was used for data analysis and insight generation, and Excel supported data validation and basic analysis. Visualization techniques were applied to interpret trends and patterns effectively.
- **Tools / Software Used:**
  SQL, Microsoft Excel, Python, and Power BI were used for data extraction, analysis, visualization, and dashboard creation.
- **PROJECT ARCHITECHTURE:**



The project architecture follows a structured data analytics workflow designed to transform raw retail data into meaningful business insights. The Global Superstore dataset serves as the data source and is initially handled using Microsoft Excel for basic inspection and validation. SQL is then used to extract, filter, and aggregate data from structured tables, enabling efficient querying and data transformation.



The dashboard provides a consolidated overview of Global Superstore performance by presenting key sales and profitability metrics. It highlights top-selling products, overall sales, profit, quantity, and discount values for quick performance assessment. The visuals show order distribution by shipping mode, sales trends over time, profit contribution by product category and sub-category, and sales performance across regions. Overall, the dashboard enables stakeholders to quickly understand business performance, identify profitable areas, and support data-driven decision-making.

GITHUB LINK :

https://github.com/vineelavadlamuri-24/Global-Superstore-Dataset-Analysis

Learning and Reflection

New Learnings:
Through this project, I gained practical experience in working with real-world retail data and applying end-to-end data analytics techniques. I improved my technical skills in SQL for data extraction and querying, Python for data cleaning and exploratory data analysis, Excel for data validation, and Power BI for building interactive dashboards. I also learned how to integrate multiple tools into a single workflow and how to interpret analytical results from a business perspective. In addition, the project enhanced my understanding of data-driven decision-making, time management, and structured problem-solving.

Overall Experience:
The overall experience of working on the Global Superstore Dataset Project was highly valuable and enriching. It provided hands-on exposure to the complete data analysis lifecycle, from raw data to actionable insights. The project helped me strengthen both my technical and analytical thinking skills while improving my confidence in presenting insights visually. Collaborating on this project and managing tasks effectively contributed to a positive learning experience and reinforced the importance of analytical tools in solving real business problems.

Conclusion and Future Scope

The Global Superstore Dataset Project successfully achieved its objective of analysing retail data to gain insights into sales performance, profitability, customer segments, and regional trends. Through effective use of SQL, Excel, Python, and Power BI, the project transformed raw transactional data into meaningful insights and visual dashboards. Key achievements include identifying high- and low-performing products and regions, understanding the impact of discounts and shipping modes on profit, and presenting results in a clear, decision-focused manner.

The future scope of this project can be extended by incorporating real-time or updated datasets to enable continuous monitoring of business performance. Advanced analytics such as predictive modeling and sales forecasting can be applied to anticipate future demand and profitability. The project can also be enhanced by integrating external data sources like market trends or customer feedback, and by adding more interactive and automated dashboards to support deeper analysis and strategic planning.