Global Superstore

Exploratory Data Analysis & Business Insights

Author: Sagnik Bhattacharya

Tools: Python, SQL (MySQL), Excel

Date: December 27, 2025

Abstract

This project presents an exploratory data analysis of the Global Superstore dataset to identify key sales, profit, customer, and operational insights. Using Python, SQL, and Excel, the analysis uncovers profitability drivers, loss-making segments, discount impacts, and regional performance patterns. The findings are translated into actionable business recommendations supported by an interactive dashboard.

Introduction

Retail businesses generate vast volumes of transactional data across products, customers, and regions. Analyzing this data enables organizations to optimize pricing strategies, improve profitability, and enhance customer targeting. This project analyzes the Global Superstore dataset to understand sales trends, profit drivers, and operational efficiency across global markets.

Dataset Description

The dataset was sourced from Kaggle:

- **Source:** https://www.kaggle.com/datasets/fatihilhan/global-superstore-dataset

- Over 51,000 transactional records

- Covers multiple countries, regions, markets, and customer segments

Key variables include sales, profit, discounts, shipping costs, order priorities, customer segments, and product categories.

Data Cleaning and Feature Engineering

The raw dataset was cleaned and transformed to ensure analytical accuracy. Key steps included:

- Standardizing order and shipping date formats

- Creating derived features such as Order Year, Month, Week Number

- Calculating Delivery Days and Profit Margin

- Flagging high-discount orders

- Removing unnecessary or redundant columns

These transformations enabled effective time-series analysis and profitability evaluation.

Exploratory Data Analysis

Sales and Profit Trends

Sales and profit were analyzed across yearly and monthly timelines to identify growth patterns and volatility.
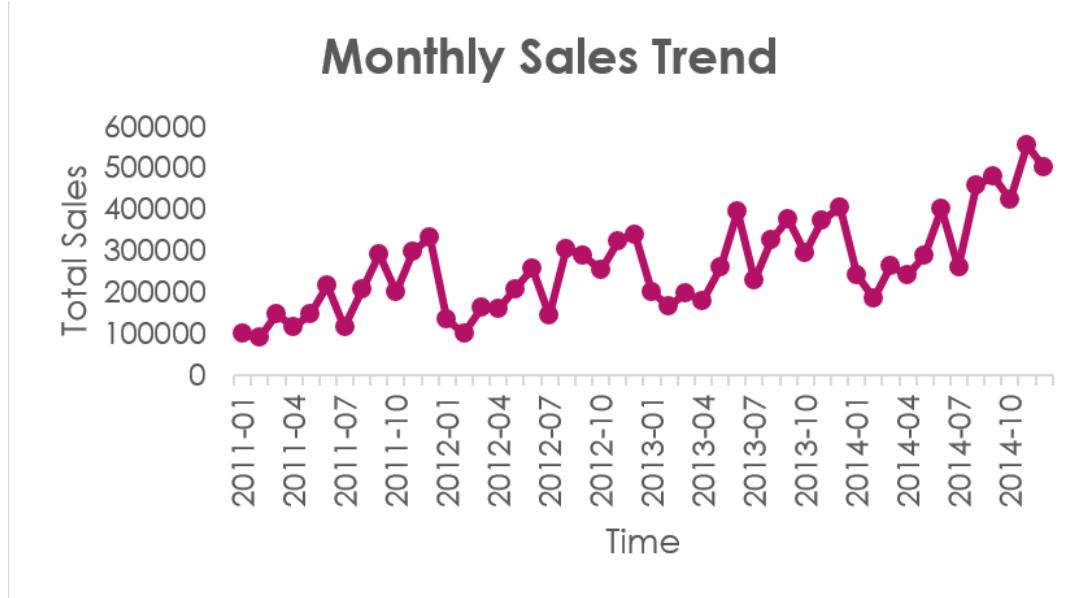

Figure 1: Monthly Sales Trend

Insight: Sales exhibit steady growth over time, while profit shows higher volatility due to discounts and shipping costs.

Category and Sub-Category Performance

Product categories and sub-categories were evaluated to identify high-performing and loss-making areas.
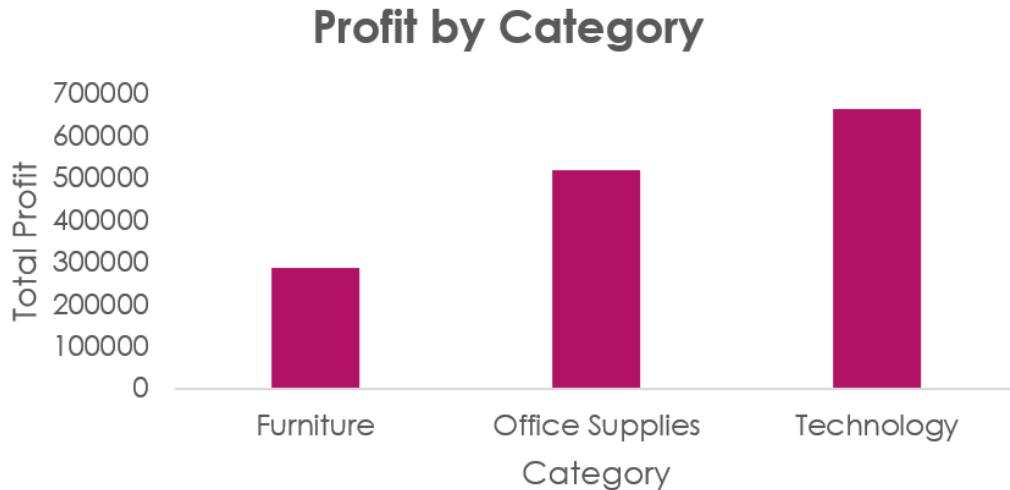
Figure 2: Profit by Product Category

Insight: The Furniture category consistently underperforms in profitability compared to Technology and Office Supplies.

Regional and Market Analysis

Sales and profit were compared across regions and markets to assess geographic performance.

Insight: Certain regions generate high sales volumes but low profit margins, indicating pricing or cost inefficiencies.

Customer and Segment Analysis

Customer segmentation analysis reveals revenue concentration among a small group of high-value customers.

Insight: A minority of customers contribute disproportionately to total revenue, supporting targeted retention strategies.

Discount and Shipping Impact

Discount levels and shipping modes were analyzed to assess margin erosion.

Insight: High discounts and Same Day shipping significantly reduce profit margins.

Dashboard Overview

An interactive Excel dashboard was created to visualize KPIs and trends.
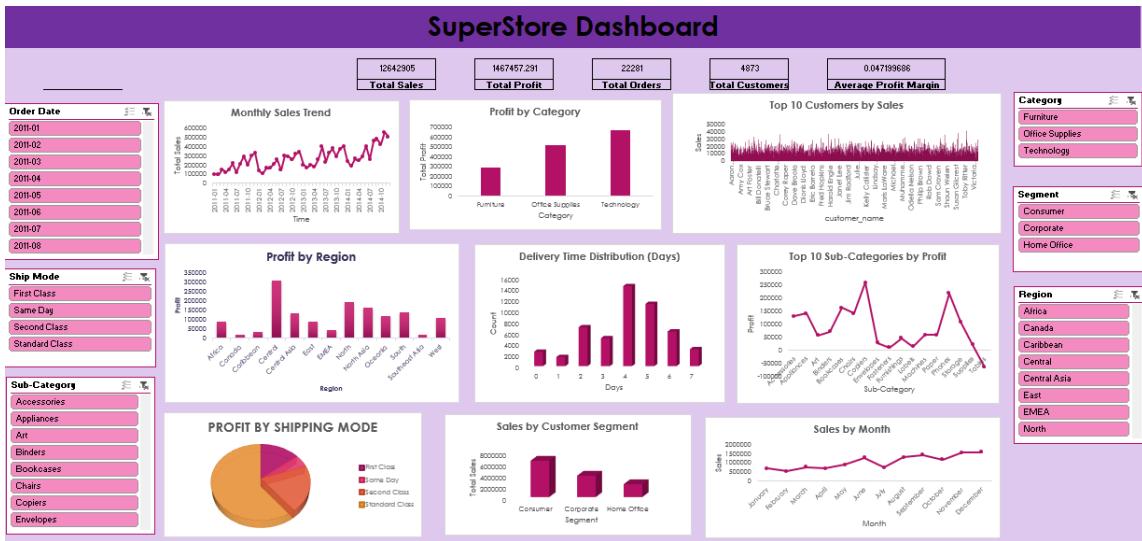
Figure 3: Global Superstore Excel Dashboard

The dashboard includes KPI cards, trend charts, and slicers for year, region, category, segment, and shipping mode.

Key Business Insights

- Sales growth does not always translate into profit growth

- High discount levels negatively impact profitability

- Furniture category consistently underperforms

- Same Day shipping reduces margins

- Revenue is concentrated among a small customer segment

Business Recommendations

1. Optimize discount strategies to protect margins

2. Re-evaluate consistently loss-making sub-categories

3. Focus marketing efforts on high-value customer segments

4. Reduce reliance on high-cost shipping modes

5. Apply region-specific pricing strategies

Conclusion

This analysis demonstrates how transactional data can be transformed into actionable business insights. The project highlights end-to-end data analytics skills including data cleaning, EDA, SQL-based KPIs, and dashboard development.