

# **Global Superstore Sales Analysis (EDA Project)**

**By: Monami Banerjee**

Institute of Engineering and Management (Newtown)

Department of Computer Science

**June 19, 2025**

# Table of Content

|                                       |                |
|---------------------------------------|----------------|
| 1. Abstract .....                     | Page No. 3     |
| 2. Introduction .....                 | Page No. 4     |
| 3. Objective .....                    | Page No. 5     |
| 4. Dataset .....                      | Page No.6      |
| 5. Technologies Used .....            | Page No. 7     |
| 6. Methodology .....                  | Page No. 8-9   |
| 7. Analysis & Results .....           | Page No. 10-12 |
| 8. Conclusion and Recommendations ... | Page No. 13    |

# **Abstract**

This project focuses on conducting an Exploratory Data Analysis (EDA) of the Global Superstore sales dataset. The primary objective was to uncover patterns, trends, and actionable business insights related to sales, profit, and customer behaviour across various regions, product categories, and shipping modes. Utilizing Python libraries such as Pandas for data manipulation, and Matplotlib, Seaborn, and Plotly for visualization, the analysis identifies key areas for business improvement, including regional profitability challenges, the impact of discounts, and product performance. The insights derived aim to support strategic decision-making to optimize sales and profit margins.

# Introduction

In today's competitive global market, understanding sales data is paramount for business success. The Global Superstore dataset provides a rich source of information spanning international sales operations. This project undertakes an in-depth exploratory data analysis to identify underlying patterns, assess performance metrics like sales and profit, and pinpoint factors influencing these metrics. The insights gained will serve as a foundation for developing data-driven strategies to enhance operational efficiency and increase profitability for the Global Superstore.

## **Objective**

The main objective of this project was to explore, visualize, and extract actionable business insights from the Global Superstore dataset using Python (Pandas, Seaborn, Plotly) in a Jupyter Notebook environment.

# Dataset

The dataset used for this analysis is "Global\_Superstore2.csv". It comprises over 10,000 records (specifically, 51,290 entries and 24 columns) of international sales data. The dataset includes comprehensive information such as:

- Transaction Details: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Sales, Quantity, Discount, Profit, Shipping Cost, Order Priority.
- Customer Information: Customer ID, Customer Name, Segment.
- Geographical Data: City, State, Country, Postal Code, Market, Region.
- Product Information: Product ID, Category, Sub-Category, Product Name.

# Technologies Used

- Programming Language: Python
- Libraries:
  - Pandas (for data manipulation and analysis)
  - Matplotlib (for static visualizations)
  - Seaborn (for enhanced statistical visualizations)
  - Plotly Express (for interactive visualizations)
  - Environment: Jupyter Notebook

# Methodology

The project followed a standard EDA methodology:

1. Data Loading: The Global\_Superstore2.csv dataset was loaded into a Pandas DataFrame.
2. Basic Exploration: Initial checks were performed to understand the dataset's shape, column names, data types, and descriptive statistics.
3. Data Cleaning:
  - Order Date and Ship Date columns were converted to datetime objects to facilitate time-based analysis.
  - Missing values were identified. It was noted that the 'Postal Code' column had a significant number of null values (41,296 out of 51,290 records). For the purpose of this analysis, missing postal codes were not critical to the primary insights and thus were not imputed or dropped.
4. Feature Engineering (Implicit): The difference between Order Date and Ship Date was calculated to derive 'Ship days', providing insight into delivery efficiency.
5. Exploratory Data Analysis (EDA): Various analyses were conducted:
  - Regional Performance: Aggregation of Sales and Profit by Region to identify top and underperforming regions.
  - Category and Sub-Category Analysis: Examination of Sales and Profit across different Category and Sub-Category to identify best-selling and most profitable product types.



- Market Analysis: Assessment of sales and profit contributions by Market.
- Discount Impact: Analysis of the relationship between Discount and Profit to understand discounting strategies.
- Time-Series Analysis: Monthly trends of Sales and Profit were visualized to identify seasonality or cyclical patterns.
- Shipping Mode Efficiency: Evaluation of Sales, Profit, and Ship days across different Ship Mode categories.
- Customer Segment Performance: Analysis of Sales and Profit by Segment (Consumer, Corporate, Home Office).
- Product Deep Dive: Identification and analysis of the top 10 products by sales.

6. Visualization: Matplotlib, Seaborn, and Plotly Express were extensively used to create various plots (e.g., bar charts, line plots, scatter plots) to visualize the data and present insights clearly.

# Analysis & Results

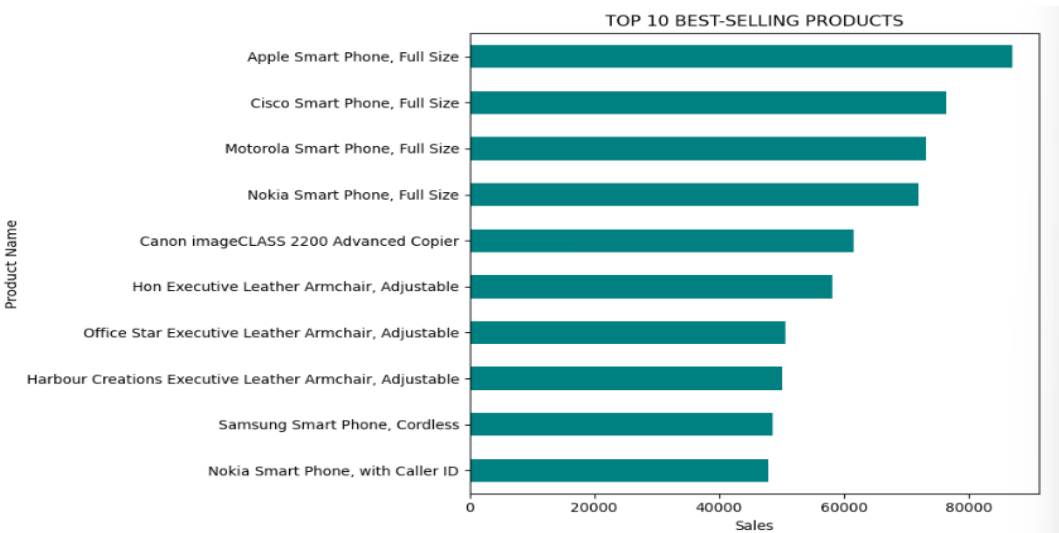
The EDA revealed several key findings:

- Regional Performance:** While some regions showed high sales, their profitability was relatively low. Notably, the **Central region exhibited high sales but surprisingly low profits**, suggesting potential issues with pricing or operational costs.

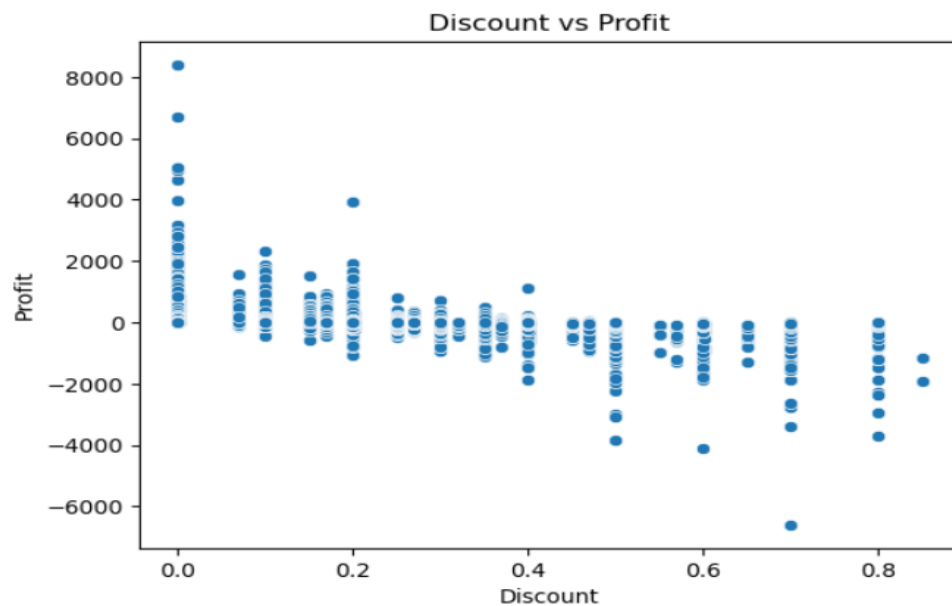
Sales by Region



- Product Contribution:** The analysis of product categories and sub-categories highlighted that a small number of **top 10 products contribute significantly to overall revenue**. This indicates a concentrated revenue stream.



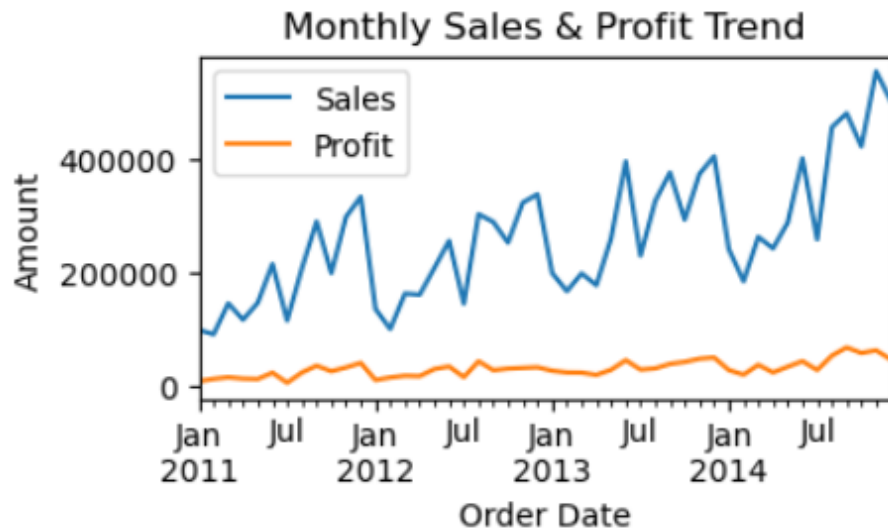
- **Impact of Discounts:** A critical finding was the **sharp decrease in profitability when discounts exceeded 20%**. This suggests that aggressive discounting negatively impacts profit margins.



- **Shipping Mode Efficiency vs. Ship Days:** The "Same Day" shipping mode resulted in the shortest delivery times but came with significantly higher shipping costs. This trade-off needs careful consideration.



- **Sales Seasonality:** The monthly trend analysis clearly demonstrated **seasonality in sales**, with certain months experiencing peak performance.



# Conclusion and Recommendations

Based on the detailed exploratory data analysis of the Global Superstore dataset, the following conclusions and actionable recommendations are put forth:

- **Profitability Review in Central Region:** The Central region, despite its high sales volume, suffers from low profitability. It is **recommended to thoroughly review pricing strategies and delivery expenses** in this region to identify and address cost inefficiencies or adjust pricing models to improve profit margins.
- **Focus on Top Products:** Given that the top 10 products are major revenue drivers, it is advisable to **focus marketing efforts and inventory management on these high-performing products** to drive more predictable and sustained sales.
- **Strategic Discounting:** The inverse relationship between high discounts and profitability necessitates a revised discount strategy. It is **suggested to implement a discount cap or a more targeted discount strategy** to protect profit margins and prevent losses from excessive price reductions.
- **Optimized Shipping Offerings:** The "Same Day" shipping mode, while fast, is costly. It is **recommended to offer this expedited shipping option only for high-priority or high-value customers** to ensure that the higher shipping costs are justified by the revenue generated or customer lifetime value.
- **Leverage Sales Seasonality:** The identified monthly sales seasonality provides an opportunity for strategic planning. It is **suggested to align promotional campaigns and inventory stocking with peak sales months** to maximize returns and capitalize on consumer demand.