**Superstore Sales Performance Analysis**

Data Analytics Report

Prepared by: Laiba

Date: May 28, 2025

# Table of Contents

1. 1. Business Problem
2. 2. Tools & Methodology
3. 3. Data Exploration
4. 4. Key Insights (EDA)
5. 4.1 Region Analysis
6. 4.2 Top Profitable Products
7. 4.3 Loss-Making Sub-Categories
8. 4.4 Discount Impact
9. 4.5 Profit by Category & Sub-Category
10. 4.6 Monthly Sales Trends
11. 4.7 Segment-wise Profit
12. 4.8 Region vs. Sub-Category Heatmap
13. 5. Recommendations
14. 6. Reflections
15. 7. Appendix: Code Snippets

# 1. Business Problem

The objective of this project is to explore retail sales data from a Superstore to uncover patterns in sales, profit, discount impact, and regional performance. Insights will help drive better decisions on pricing, promotions, and inventory strategy.

# 2. Tools & Methodology

 Tools Used: Python, Pandas, Seaborn, Matplotlib, Google Colab

 Methodology: CRISP-DM (Data understanding → Preparation → EDA → Insights)

# 3. Data Exploration

#### 3.1 Dataset Info

* Rows: 9,994
* Columns: 21
* Source: Kaggle Superstore Dataset

#### 3.2 Missing Values

* No missing values found

#### 3.3 Data Types

* Dates converted to datetime
* All columns correctly typed (no corrections needed)

#### 3.4 Removed Columns

* Row ID and Postal Code were dropped due to irrelevance

# 4. Key Insights (EDA)

#### 4.1 Region Analysis

* 📍 West is the most profitable region

#### 4.2 Top Profitable Products

* Copiers, Phones, and high-end office equipment (e.g., Canon Copiers)

#### 4.3 Loss-Making Sub-Categories

* Bookcases, Tables, and Supplies — consistently negative profit

#### 4.4 Discount Impact

* High discounts (> 40%) → lead to **negative profit**
* Office Supplies suffer the most from discounting

#### 4.5 Profit by Category & Sub-Category

* 📈 Copiers & Phones = high profit
* 📉 Bookcases & Tables = losses despite high sales

#### 4.6 Monthly Sales Trends (if completed)

* Peaks during year-end months
* Low performance in mid-year months

#### 4.7 Segment-wise Profit

* Consumer segment shows higher profit than Home Office or Corporate

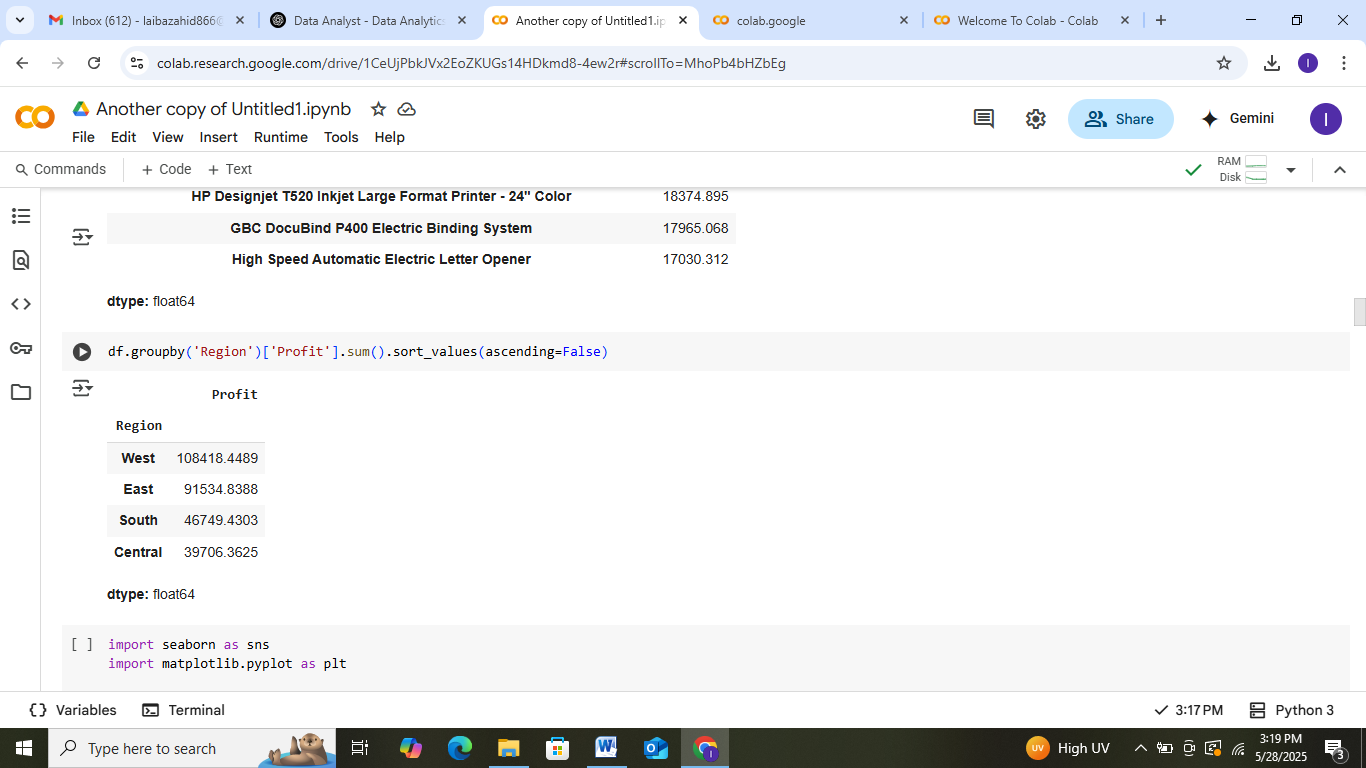
#### 4.8 Heatmap: Region vs. Sub-Category

* Example: Bookcases in the East = consistent loss
* Phones and Chairs in the West = strong profit

## 4.1 Region Analysis

#### 4.1 Region Analysis

* 📍 West is the most profitable region



## 4.2 Top Profitable Products

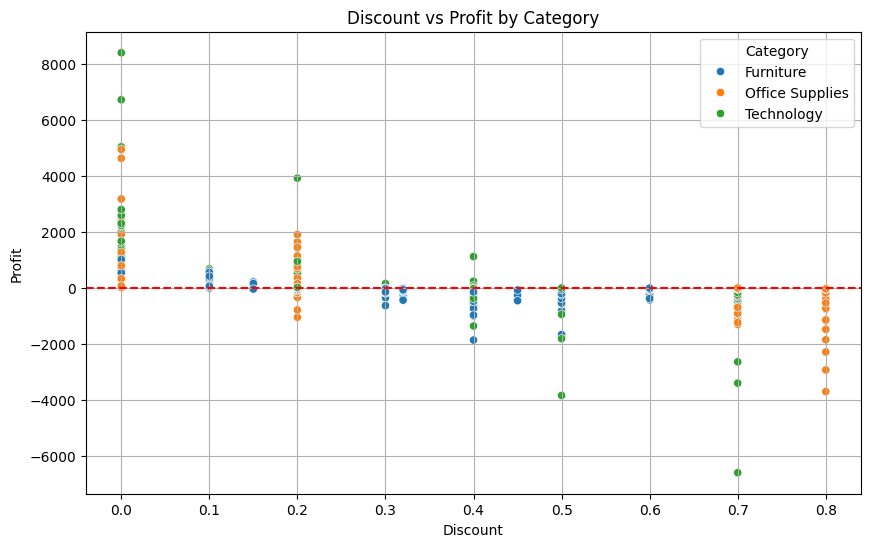
* Copiers, Phones, and high-end office equipment (e.g., Canon Copiers)

## 4.3 Loss-Making Sub-Categories

* Bookcases, Tables, and Supplies — consistently negative profit

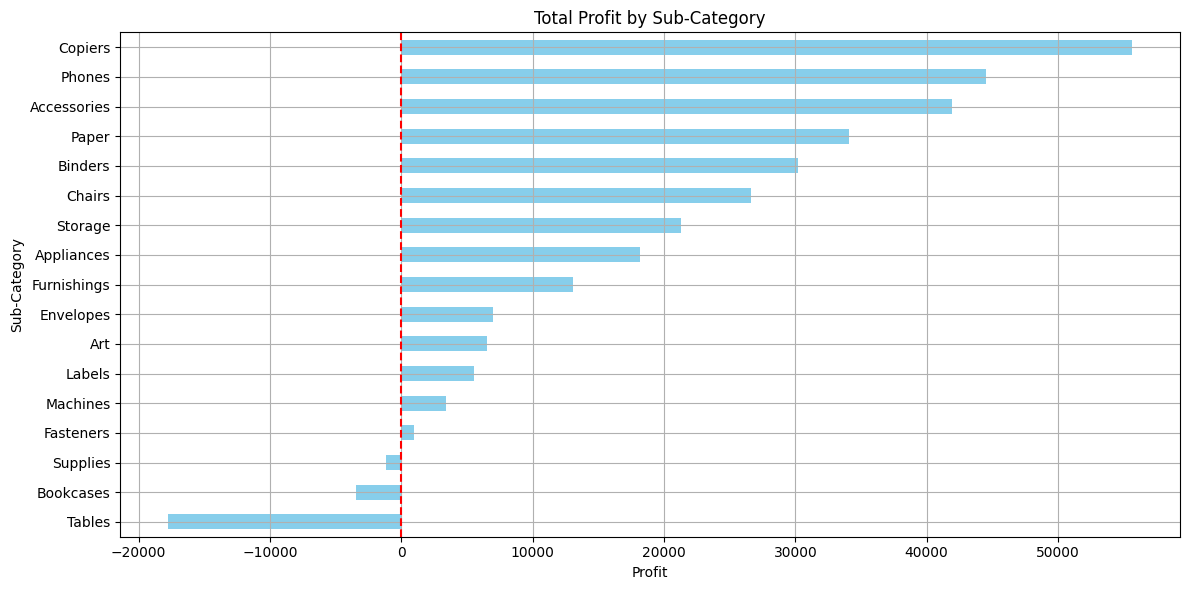
## 4.4 Discount Impact

* High discounts (> 40%) → lead to **negative profit**
* Office Supplies suffer the most from discounting



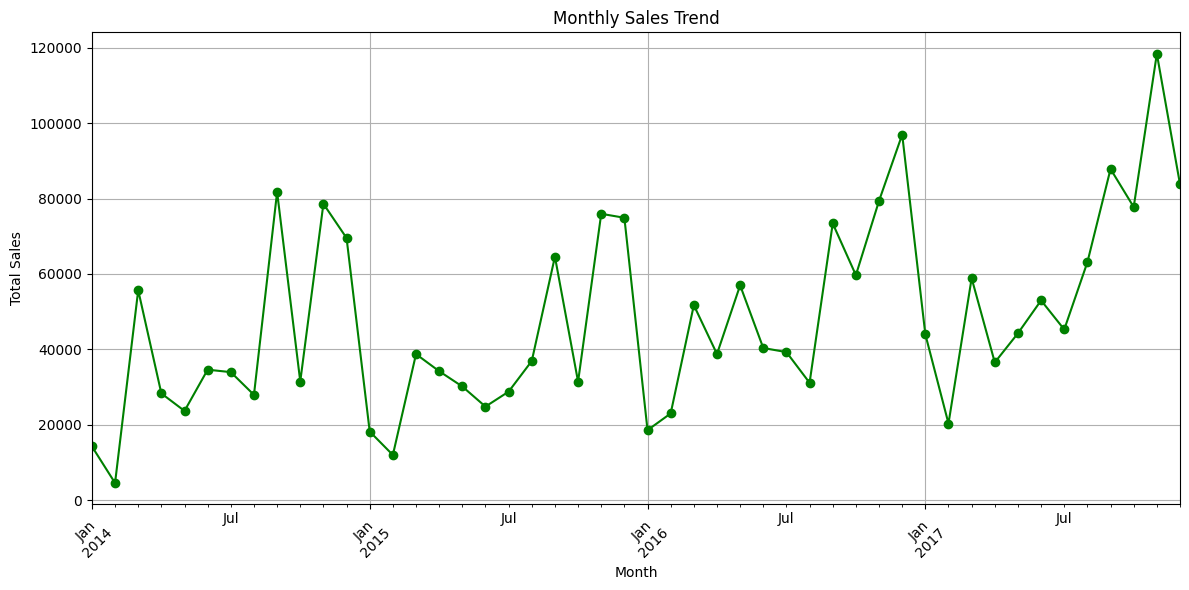
## 4.5 Profit by Category & Sub-Category

* 📈 Copiers & Phones = high profit
* 📉 Bookcases & Tables = losses despite high sales



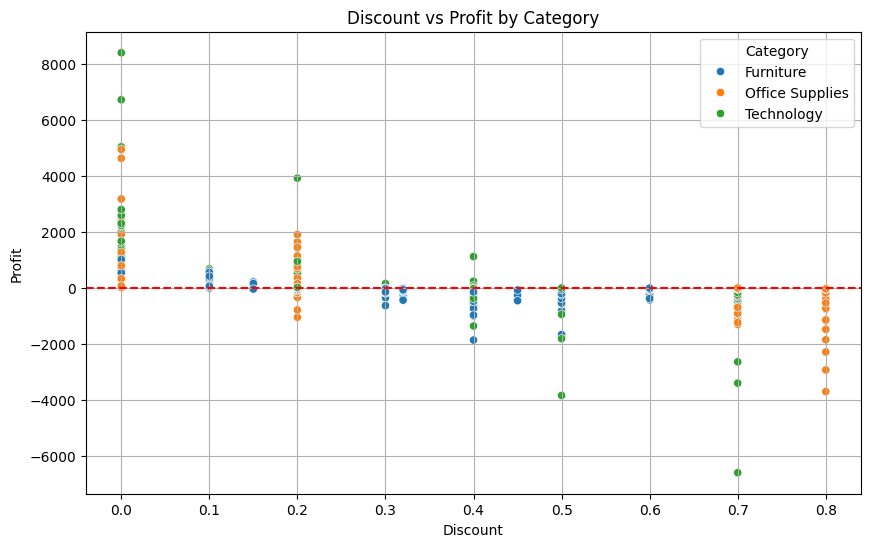
## 4.6 Monthly Sales Trends

* Peaks during year-end months
* Low performance in mid-year months



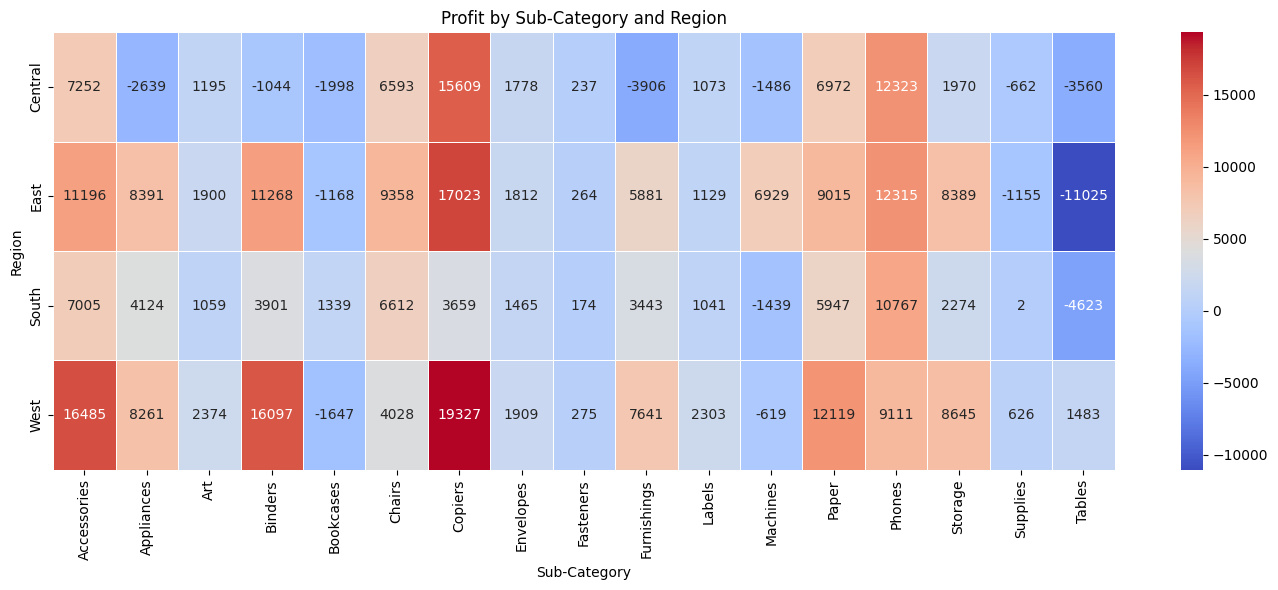
## 4.7 Segment-wise Profit

* Consumer segment shows higher profit than Home Office or Corporate



## 4.8 Region vs. Sub-Category Heatmap

* Example: Bookcases in the East = consistent loss
* Phones and Chairs in the West = strong profit



# 5. Recommendations

 Limit high discounts, especially in low-margin sub-categories like Bookcases and Tables

 Focus promotions on profitable categories (Copiers, Phones)

 Tailor strategy per region and sub-category use heatmap.

 Monitor discount thresholds to prevent margin loss

# 6. Reflections

This project helped me apply real data cleaning, exploration, and visualization techniques to a business problem. I learned how to interpret patterns in profit and sales, and how to build a clear analytical story using Python.

# 7. Appendix: Code Snippets

**For loading the data set**

## import pandas as pd

## df = pd.read\_csv('/content/drive/My Drive/Sample - Superstore.csv', encoding='latin1')

**For cleaning and data prepration**

## df.drop(columns=[col for col in ['Row ID', 'Postal Code'] if col in df.columns], inplace=True)

## df['Order Date'] = pd.to\_datetime(df['Order Date'], errors='coerce')

## df['Ship Date'] = pd.to\_datetime(df['Ship Date'], errors='coerce')

## df['Month'] = df['Order Date'].dt.to\_period('M')

**Discount vs profit graph**

## import seaborn as sns

## import matplotlib.pyplot as plt

## plt.figure(figsize=(10,6))

## sns.scatterplot(data=df, x='Discount', y='Profit', hue='Category')

## plt.axhline(0, color='red', linestyle='--')

## plt.title('Discount vs Profit by Category')

plt.savefig("discount\_vs\_profit.png")**Profit vs sub category**

## df.groupby('Sub-Category')['Profit'].sum().sort\_values().plot(kind='barh', color='skyblue')

## plt.title('Total Profit by Sub-Category')

## plt.axvline(0, color='red', linestyle='--')

## plt.savefig("profit\_by\_subcategory.png")

**Monthly sale trend**

## monthly\_sales = df.groupby('Month')['Sales'].sum()

## monthly\_sales.plot(kind='line', marker='o', color='green')

## plt.title('Monthly Sales Trend')

## plt.savefig("monthly\_sales\_trend.png")