

Project Title

Retail Sales and Profit Analysis — Superstore Dataset

Introduction

This project focuses on analyzing sales data from a fictional US-based retail company, known as *Superstore*. The dataset includes detailed records of orders such as sales, profit, discounts, shipping, product categories, customer segments, and regional information.

The goal of this analysis is to extract valuable business insights by identifying sales trends, understanding profitability across different categories and regions, and evaluating the impact of discount strategies. These insights aim to support data-driven decision-making in areas like inventory planning, pricing, and marketing strategies.

Objectives

- Understand overall sales and profit performance.
 - Identify high-performing product categories and sub-categories.
 - Evaluate regional contribution to sales.
 - Examine the effect of discounts on profitability.
 - Recommend actionable strategies based on patterns observed.
-

Tools Used

- **Python:** Data handling and analysis
- **Pandas:** Data manipulation
- **Matplotlib & Seaborn:** Data visualization

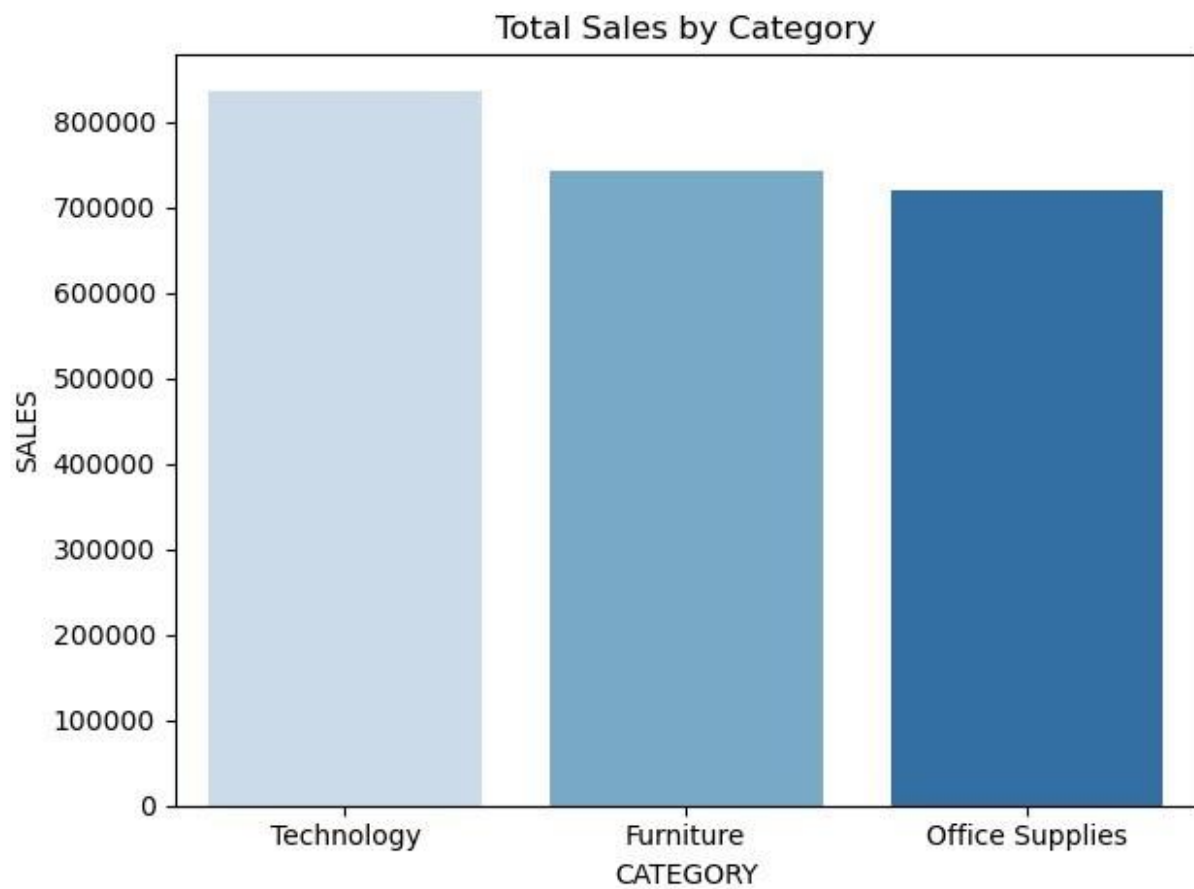
- **Jupyter Notebook:** Interactive analysis

- This following section loads the dataset and processes it to calculate total sales across product categories. The grouped data is sorted and prepared for visualization to identify which categories drive the most revenue.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_excel("Sample - Superstore.xls")
category_sales=df.groupby('Category') ['Sales'].sum().sort_values(ascending=False)
category_sales=category_sales.reset_index()
```

1.Total Sales by category

```
sns.barplot(x='Category',y='Sales',hue='Category',data=category_sales,  
palette='Blues',legend=False) plt.title('Total Sales by Category')  
plt.xlabel('CATEGORY') plt.ylabel('SALES') plt.tight_layout()  
plt.show()
```



➤ **Insight:**

In the above presented dataset Technology exceeds \$8,30,000 and leaves the Furniture with \$7,50,000 and Office supplies with \$7,25,000. It seems that customers are spending on more tech related products due to more demand and value.

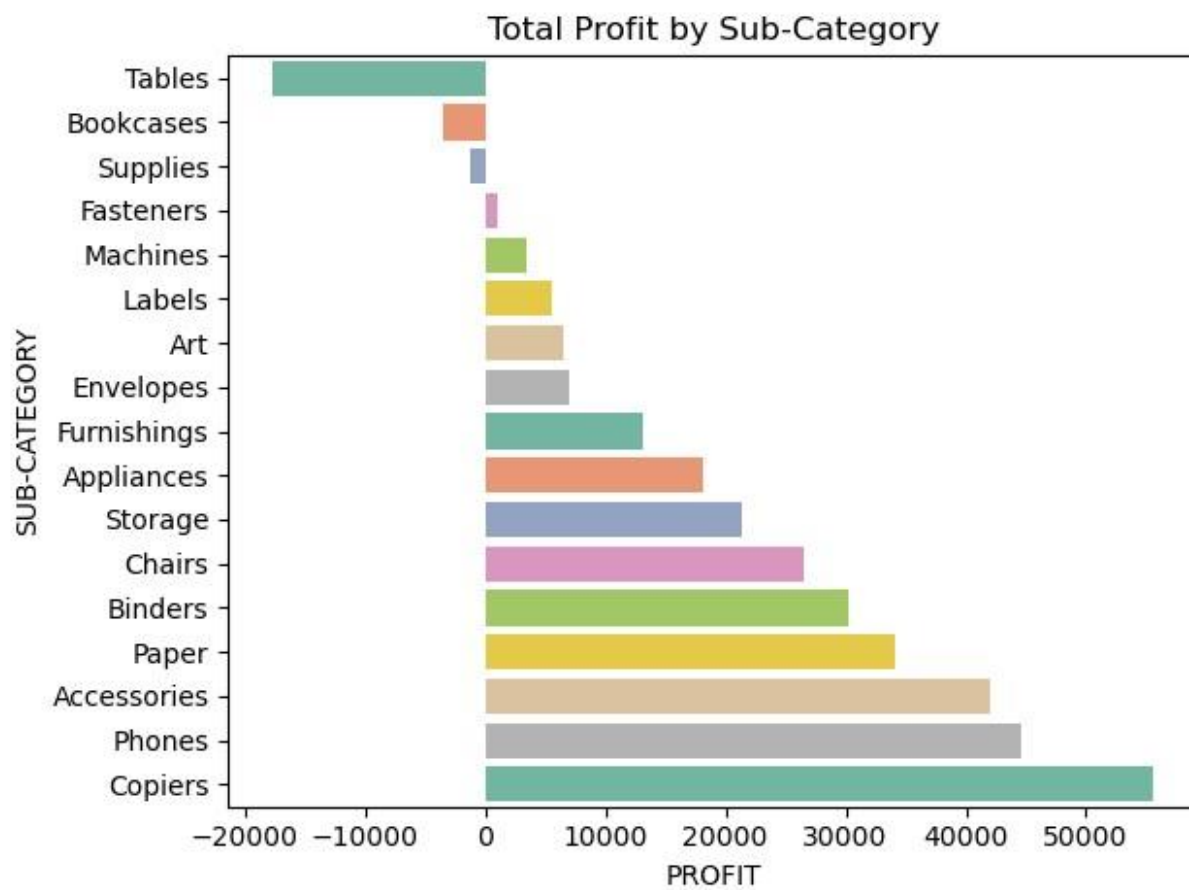
➤ **Recommendation:**

Focusing on more marketing and inventory ideas on Technology as it yields more revenue.

2.Total Profit by Sub-Category

```
sub_category_sales=df.groupby('Sub-Category')
['Profit'].sum().sort_values(ascending=True)
sub_category_sales=sub_category_sales.reset_index()

sns.barplot(y='Sub-Category',x='Profit',hue='Sub-Category',data=sub_category_sales,palette='Set2',legend=False)
plt.title('Total Profit by Sub-Category') plt.xlabel('PROFIT')
plt.ylabel('SUB-CATEGORY') plt.tight_layout() plt.show()
```



➤ **Insight:**

According to the given data the copiers sub-category has higher demand and it has a higher profit as compared to other sub-categories while phones and accessories are also in the high profit gaining race. "Tables generate losses — indicating a need to review pricing or reduce costs."

➤ **Recommendation:**

1). The profit margin on the sub-categories like Tables and Bookcases should be increased in order to gain more profit and to cover the loss.

2). Proper analysis should be done behind the reason why Tables and Bookcases have such a low profit and according to it proper business ideas should be applied to them.

3). A little bit focus on the other categories like storage, chairs, art, envelopes furnishing, binders and paper should be considered and profit margin should be increased in order to gain more profit.

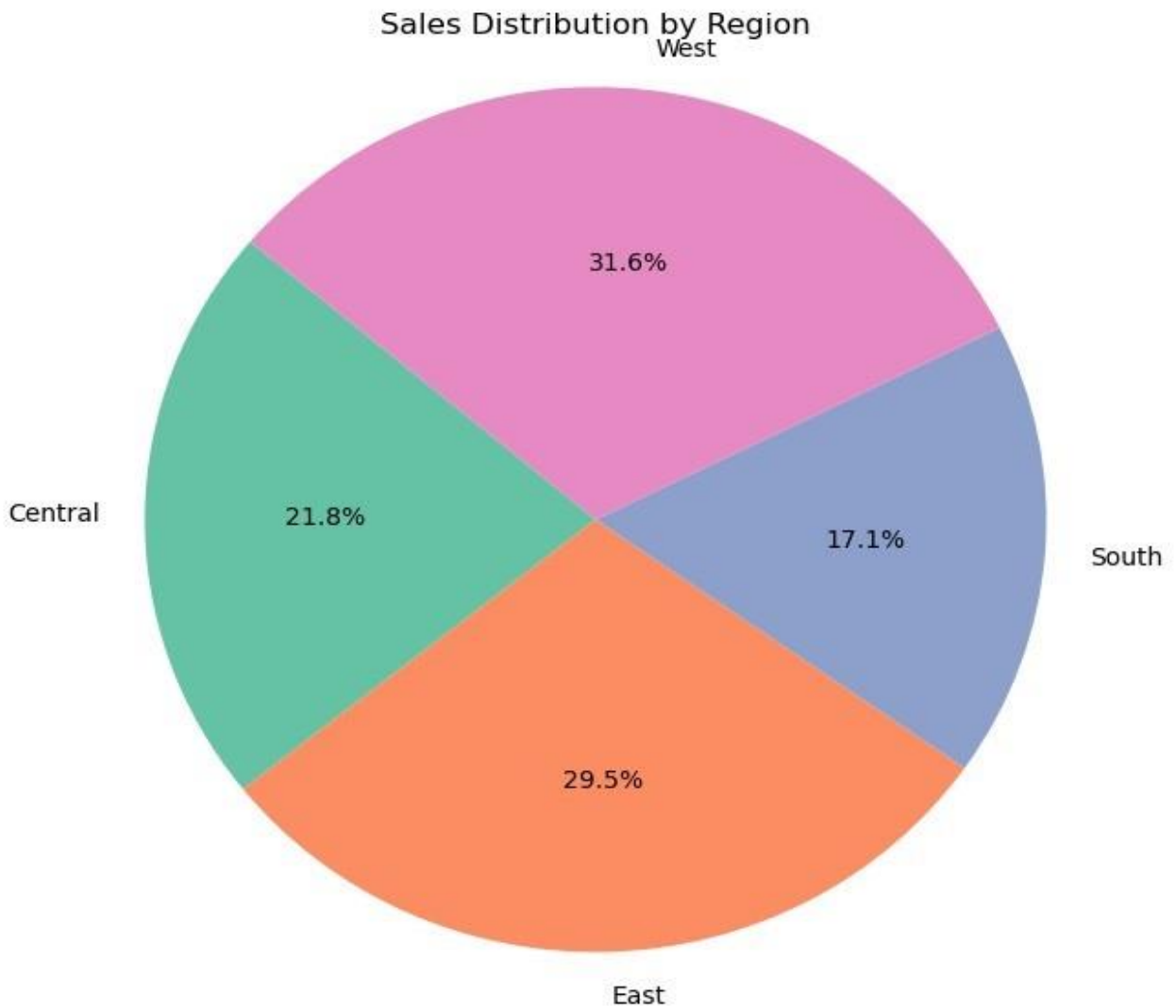
3.Sales Distribution by Region

```
region_sales=df.groupby('Region')['Sales'].sum().sort_values(ascending=False)
region_sales=region_sales.reset_index()
import matplotlib.pyplot as plt
# Group by Region
region_sales = df.groupby('Region')['Sales'].sum()

# Data for pie chart labels =
region_sales.index sizes =
region_sales.values

# Optional: create a color list
manually or use a colormap
safely colors =
plt.cm.Set2(range(len(labels)))
# This safely samples colors

# Plot pie chart plt.figure(figsize=(7, 7))
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140, colors=colors)
plt.title('Sales Distribution by Region')
plt.axis('equal') # Equal aspect ratio ensures the pie chart is circular
plt.show()
```



➤ **Insight:**

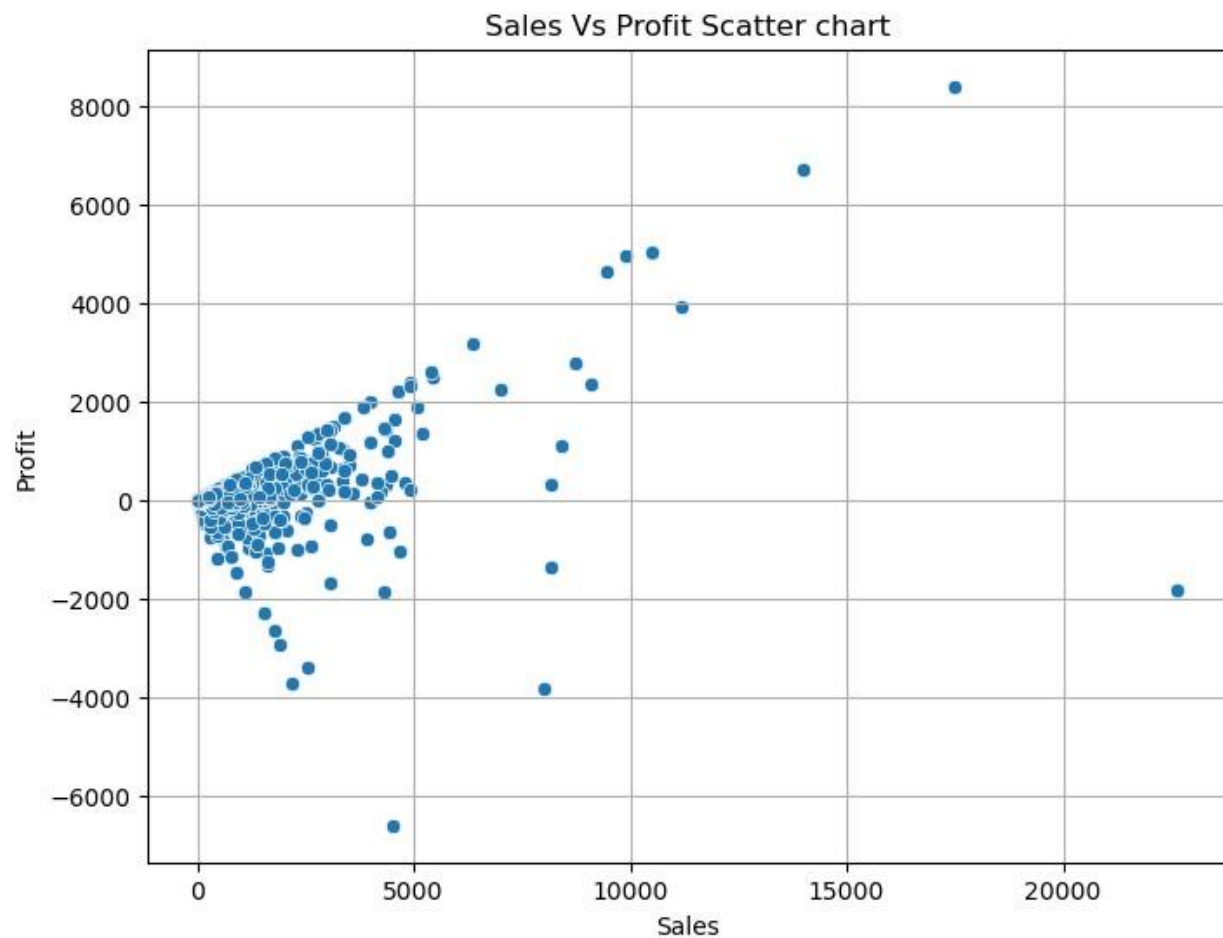
1). Dominance of the west: According to the represented data west has a great dominance in sales with 31.6% which actually shows more sales in the west part.

2). East as key: The east part with 29.5% is at the second position for the higher sales and this area should be focused in order to get a good percentage of sales.

3). Meanwhile the central and south parts have little bit of low sales percentage with 21.8% and 17.1% respectively which should be considered in order for more sales result.

4. Sales Vs Profit Chart

```
plt.figure(figsize=(8,6))
sns.scatterplot(data=df,x='Sales',y='Profit') plt.title('Sales Vs
Profit Scatter chart') plt.xlabel('Sales') plt.ylabel('Profit')
plt.grid(True) plt.show()
```



➤ Insight:

1). Positive Correlation Between Sales and Profit - Most data points show that as sales increase, profit tends to rise, indicating a general trend of profitability scaling with higher sales.

2). Clustered Low Sales & Profit - A significant number of points are concentrated in the lower-left quadrant, suggesting that many transactions involve lower sales and profit margins.

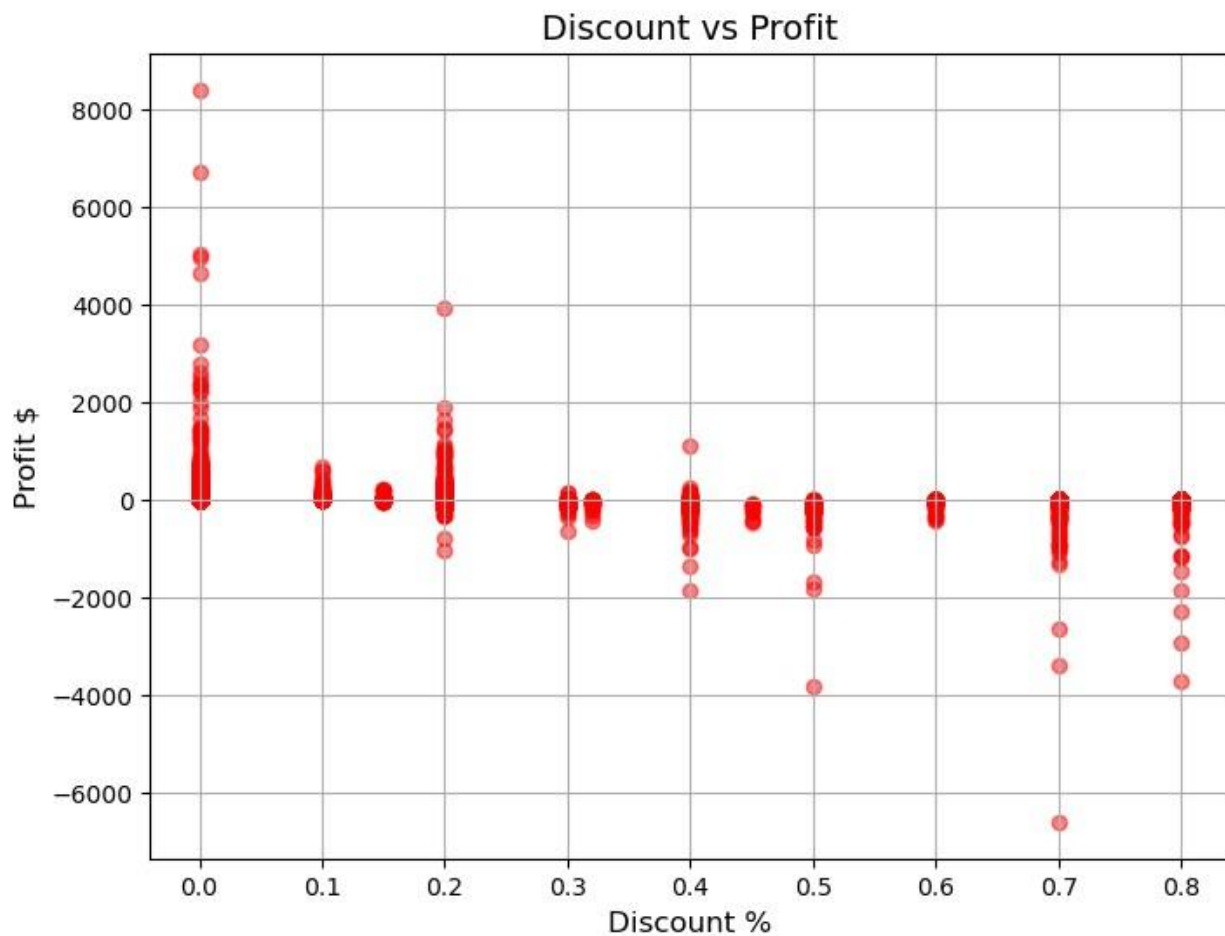
3). High-Profit Outliers - Some points extend upwards and to the right, showing instances where high sales resulted in substantial profit. These could represent premium products or successful marketing strategies.

4). Negative Profit Cases - A few points indicate losses despite sales, meaning certain products or transactions might have high costs or inefficiencies leading to negative profit.

5). Potential for Optimization - Understanding why some sales lead to losses could help refine pricing strategies, cost management, or product selection.

5. Discount Vs Profit

```
plt.figure(figsize=(8,6))
plt.scatter(df['Discount'], df['Profit'], color='red',alpha=0.5)
plt.title('Discount vs Profit', fontsize=14) plt.xlabel('Discount %',
fontsize=12) plt.ylabel('Profit $', fontsize=12)
plt.grid(True) plt.show()
```



➤ Insight:

1). Inverse Relationship Between Discount and Profit - If the trend shows that higher discounts generally lead to lower profits (or losses), it suggests that excessive discounts are reducing profitability.

2). Threshold for Optimal Discounts - Identifying a range where discounts improve sales without significantly cutting into profits can help optimize pricing strategies.

3). Negative Profit Areas - If there are several instances where high discounts result in negative profit, it indicates that certain products or categories may not be suited for aggressive discounting.

4). Profitable Discounts - Some discount levels may still allow for positive profit margins, meaning there could be an ideal percentage that encourages sales without harming profitability.

5). Cluster Analysis - If most data points are concentrated within a specific discount range, businesses could analyze this section to determine a safe discounting strategy that maintains profit levels.

Summary of Insights

Briefly restate the **main findings**:

- **Technology** category leads in sales and profit.
 - **High discounts** negatively impact profit margins.
 - **West and East regions** contribute the most to sales.
 - **Sub-categories like Tables and Bookcases** consistently generate losses.
-

Final Recommendations

Give **business or data-driven suggestions**:

- Reduce or optimize discounts to protect profit.
- Focus on top-performing regions for targeted campaigns.
- Investigate why certain sub-categories underperform (e.g., cost, demand, pricing).
- Invest more in inventory and promotion of high-profit products like Copiers and Phones.