

STUDENT DETAILS



Name- Gaurav Gupta

Skillsbuild ID- gaurav.2024cse1083@kiet.edu

College- Kiet Group Of Institutions Muradnagar
Ghaziabad Up

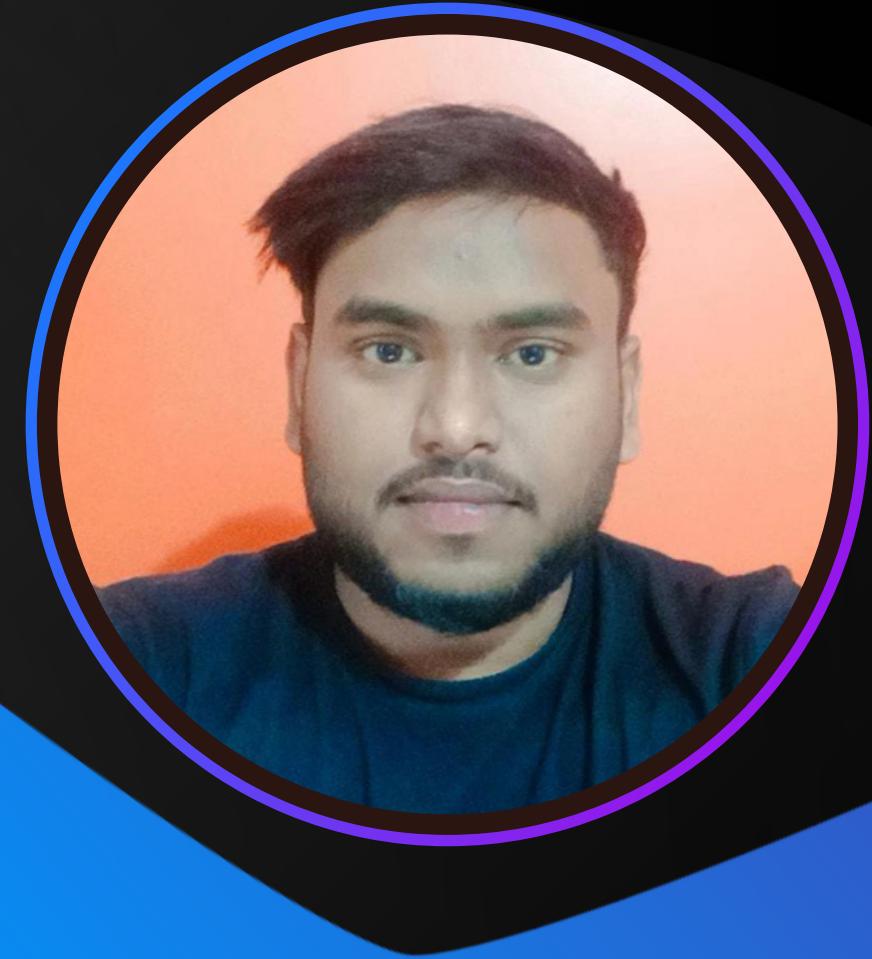
State - Uttar Pradesh

Domain- DATA ANALYTICS

Duration- 12-06-2023 to 24-07-2023

AICTE Student ID: STU61bb43610bd841639662433

**IBM SKILLSBUILD
EDUNET FOUNDATION**



PROJECT TITLE/Problem Statement

- *Analysis of Sample Superstore Dataset*
- *The management of Sample Superstore, a retail company, wants to gain insights into their sales data to optimize their operations and improve profitability. They have provided a dataset containing information about their customers, products, and transactions over a certain period. The objective of this project is to analyze the dataset and extract meaningful insights by performing various data analytics techniques and statistical analyses. The analysis should focus on identifying patterns, trends, and relationships within the data, as well as identifying factors that may impact sales and customer behavior.*



AGENDA

1. Introduction to the Sample Superstore Dataset:

- Overview of the dataset structure and variables.
- Understanding the business context and objectives of the project.

2. Data Exploration and Cleaning:

- Exploring the dataset to identify missing values, outliers, and inconsistencies.
- Cleaning the dataset by addressing missing values, outliers, and inconsistencies.

3. Descriptive Analytics:

- Performing basic statistical analysis to summarize and describe the dataset.
- Generating descriptive statistics, such as mean, median, mode, standard deviation, etc., for relevant variables.
- Visualizing data using charts, graphs, and histograms to gain initial insights.

4. Customer Segmentation:

- Conducting customer segmentation based on various customer attributes (e.g., demographics, purchase behavior, etc.).
- Applying clustering algorithms (e.g., K-means clustering) to identify distinct customer segments.
- Analyzing the characteristics and preferences of each segment.

5. Sales Analysis:

- Analyzing sales trends over time (e.g., monthly, quarterly, yearly) to identify patterns and seasonality.
- Examining the performance of different product categories and sub-categories.
- Investigating the correlation between sales and other factors (e.g., region, customer segment, etc.).

6. Profitability Analysis:

- Calculating profit margins for different products and product categories.
- Identifying the most profitable and least profitable products.
- Analyzing the impact of discounts, shipping costs, and other factors on profitability.

7. Customer Behavior Analysis:

- Examining customer purchasing patterns, such as frequency, recency, and monetary value.
- Analyzing customer loyalty and retention rates.
- Investigating factors influencing customer churn and identifying potential strategies for customer retention.

8. Recommendations and Insights:

- Summarizing the key findings and insights from the analysis.
- Providing actionable recommendations to improve sales, profitability, and customer satisfaction.
- Presenting the results in a clear and concise manner using visualizations and data-driven insights.

9. Conclusion:

- Concluding the analysis project and summarizing the key takeaways.
- Reflecting on the limitations of the analysis and potential areas for further exploration.

Project Overview



Purpose

This project focuses on analyzing the SuperStore dataset, which contains Sales and Profit data from a retail store in the United States. The goal of this project is to conduct a comprehensive analysis of the Sample Superstore dataset to gain valuable insights into Sales trends and Profitability of the store and identify areas for improvement.



Scope

The findings and recommendations derived from the analysis can support strategic planning, informed decision-making, and the overall success of the Superstore.



Objectives

Performance Evaluation, Customer Insights, Inventory Management, Cost Analysis, Operational Efficiency, Operational Efficiency

WHO ARE THE END USERS OF THIS PROJECT?

The end users of the Analysis of Sample Superstore Dataset project can include various stakeholders within the Sample Superstore company. These stakeholders may include:

- 1. Management Team:** *The project's analysis and insights are primarily aimed at the management team of Sample Superstore. They are the key decision-makers who will utilize the findings to make strategic decisions and implement changes to improve the company's operations, sales, and profitability.*
- 2. Business Analysts:** *Business analysts within Sample Superstore may use the project results to gain a deeper understanding of customer behavior, product performance, and sales trends. They can further analyze the insights and incorporate them into their reports and recommendations.*
- 3. Marketing Team:** *The marketing team can benefit from the project's findings to develop targeted marketing strategies based on customer segmentation and preferences. The analysis can help them identify potential customer segments to focus on and tailor marketing campaigns accordingly.*
- 4. Sales Team:** *The sales team can leverage the project's insights to optimize their sales approach. By understanding sales trends, product performance, and customer behavior, they can adapt their sales strategies to better meet customer demands and increase sales revenue.*

- 
- 5. Operations Team:** The operations team can utilize the project's recommendations to streamline processes, improve inventory management, and optimize supply chain operations. They can identify areas of improvement based on the analysis of product categories, profitability, and customer satisfaction.
 - 6. Finance Team:** The finance team can benefit from the project's analysis of profitability and cost factors. The insights can help them identify cost-saving opportunities, evaluate the impact of discounts and shipping costs, and optimize pricing strategies.
 - 7. Data Analysts/Scientists:** Data analysts or data scientists within the organization can use the project as a reference for similar analyses and as a benchmark for future data-driven projects. They can also contribute their expertise in analyzing the dataset and extracting insights.

YOUR SOLUTION AND ITS VALUE PROPOSITION

Solution:

- Our solution for the Analysis of Sample Superstore Dataset project involves applying advanced data analytics techniques to gain valuable insights from the provided dataset. By conducting comprehensive data exploration, cleaning, and analysis, we aim to uncover patterns, trends, and relationships within the data. The project includes customer segmentation, sales analysis, profitability analysis, and customer behavior analysis, among other key components.

Value Proposition:

- **Actionable Insights:** Our analysis provides actionable insights to the management team of Sample Superstore, enabling them to make data-driven decisions. The findings help identify areas of improvement, optimize operations, and enhance profitability.
- **Improved Sales and Customer Satisfaction:** By understanding customer behavior, sales trends, and product performance, our analysis enables the development of targeted marketing strategies and sales approaches, leading to increased sales revenue and improved customer satisfaction.
- **Cost Optimization:** Our profitability analysis highlights the most and least profitable products and identifies factors impacting profitability. This empowers the management team to optimize pricing, manage discounts, and streamline operations for cost savings.

- **Enhanced Decision-Making:** Our project equips business analysts and stakeholders with a comprehensive understanding of the dataset, enabling them to make informed decisions. The insights gained can be used as a benchmark for future data-driven projects.
- **Competitive Advantage:** By leveraging data analytics techniques, Sample Superstore gains a competitive edge in the retail industry. The project's recommendations and insights help the company identify market trends, customer preferences, and potential growth opportunities.
- **Streamlined Operations:** The analysis provides valuable insights for the operations team, facilitating optimized inventory management, supply chain operations, and overall operational efficiency.
- **Scalability and Adaptability:** Our solution can be scaled and adapted to accommodate additional datasets, variables, and analysis requirements, allowing Sample Superstore to leverage data analytics for ongoing improvement and growth.

Summary of the data

In [] :

```
df.value_counts()
```

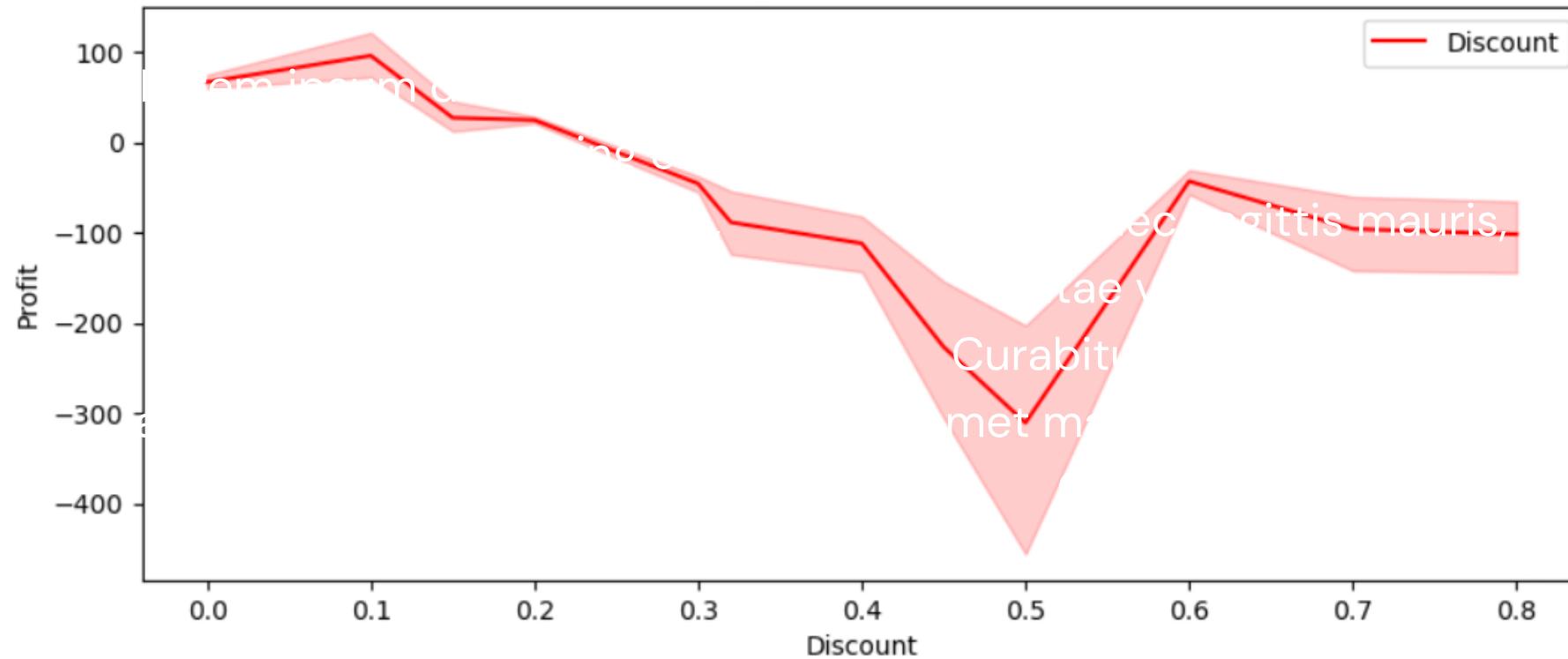
```
Out[ ]: Ship Mode      Segment      Country      City      State      Region      Category      Sub-Category      Sales
Quantity  Discount  Profit
Standard  Class  Consumer      United States  Los Angeles  California  West  Office Supplies  Paper      19.440
3          0.0    9.3312      4
                  Corporate      United States  Houston     Texas      Central  Office Supplies  Paper      15.552
3          0.2    5.4432      3
                  Consumer      United States  San Francisco  California  West  Office Supplies  Paper      25.920
4          0.0    12.4416      2
                  Detroit        Michigan     Central  Furniture      Chairs      389.970
3          0.0    35.0973      2
                  Philadelphia  Pennsylvania  East  Office Supplies  Paper      20.736
4          0.2    7.2576      2
.
.
.
Second  Class  Corporate      United States  Little Rock  Arkansas  South  Office Supplies  Paper      44.960
2          0.0    20.6816      1
                  Storage        62.040
4          0.0    17.3712      1
                  367.960
4          0.0    14.7184      1
                  Technology      Phones      494.970
3          0.0    148.4910      1
Standard  Class  Home Office  United States  Yuma        Arizona  West  Technology      Machines      599.985
5          0.7    -479.9880      1
Length: 9944, dtype: int64
```

Outputs

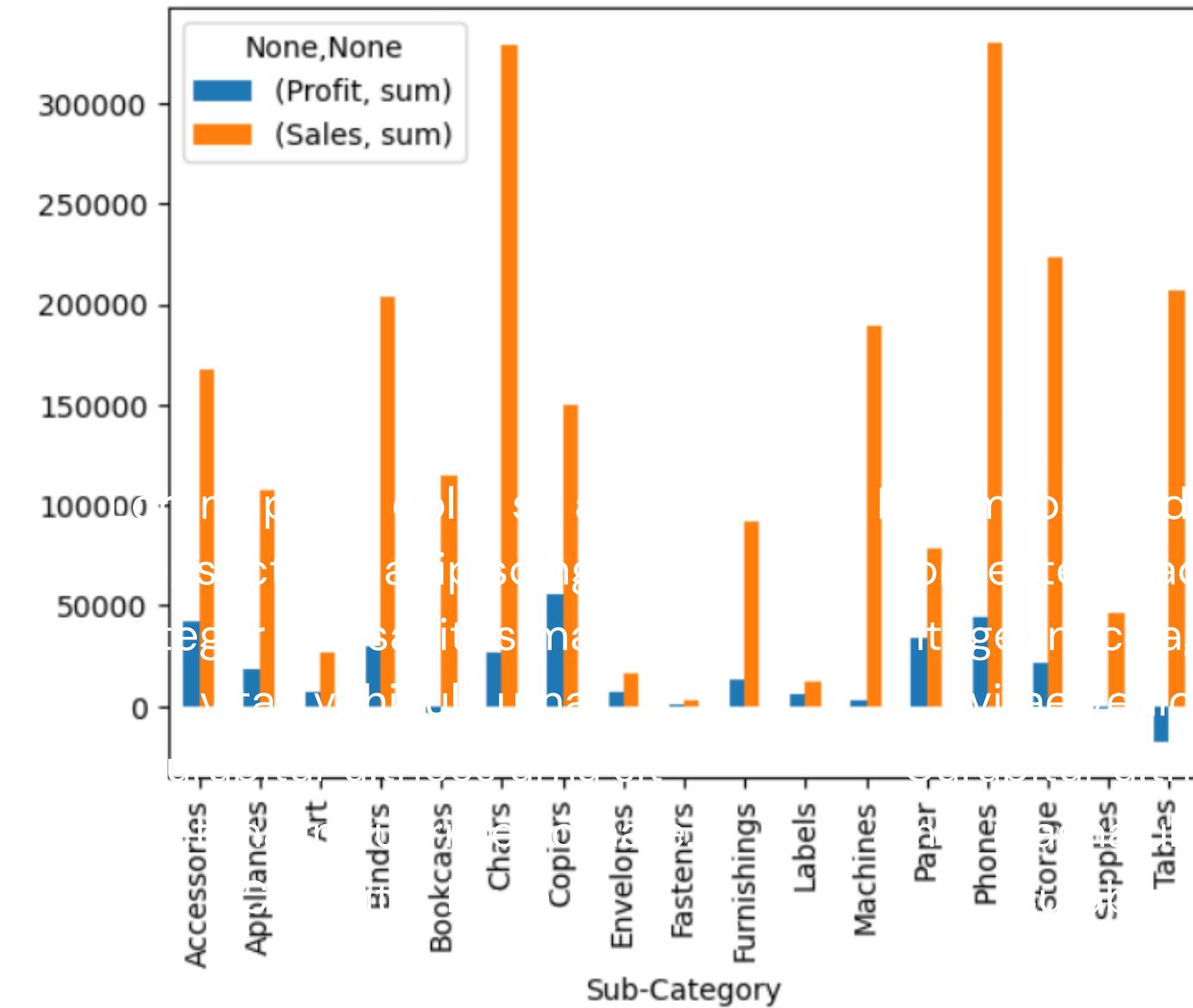
Line plot of Discount Vs Profit

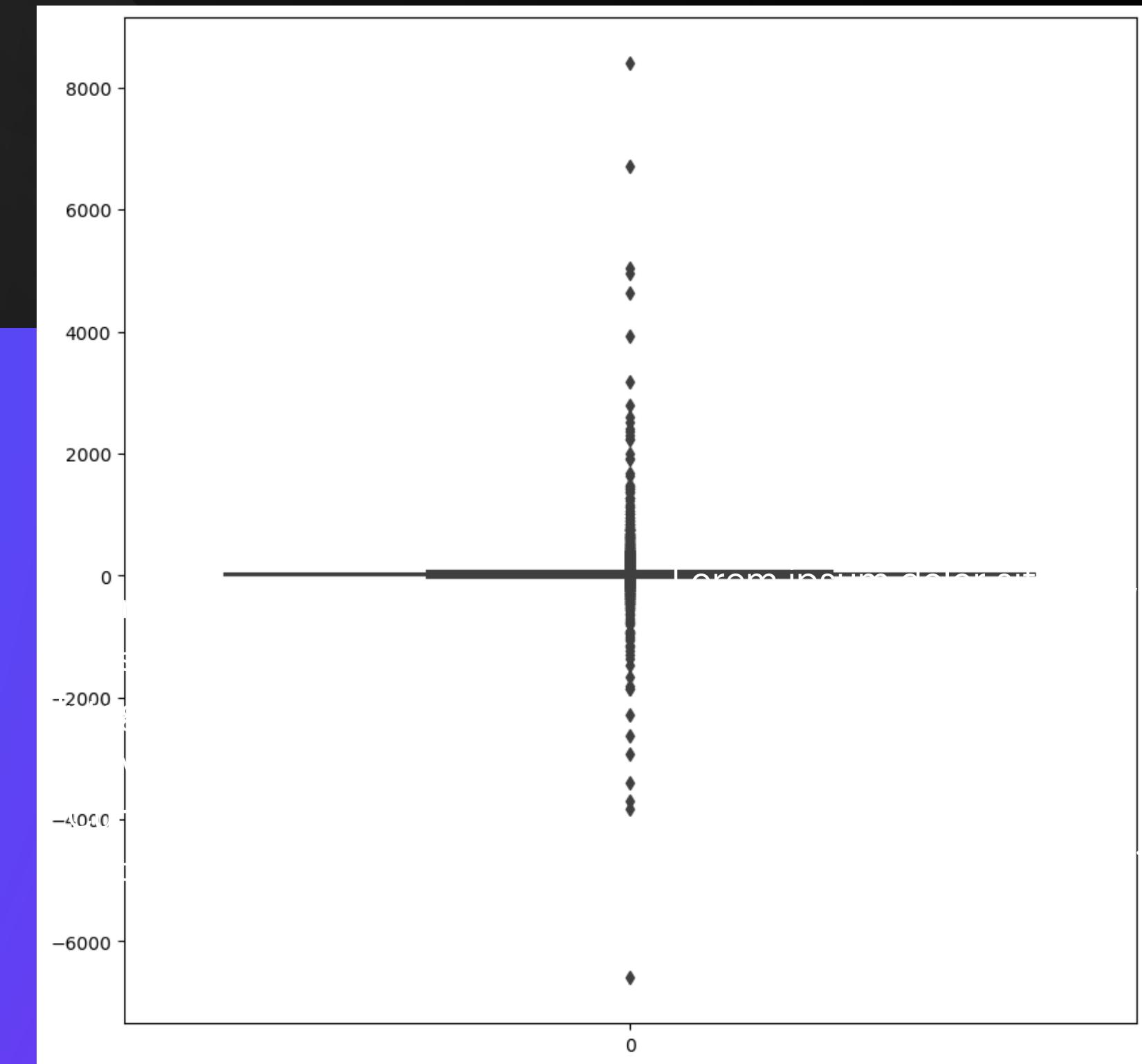
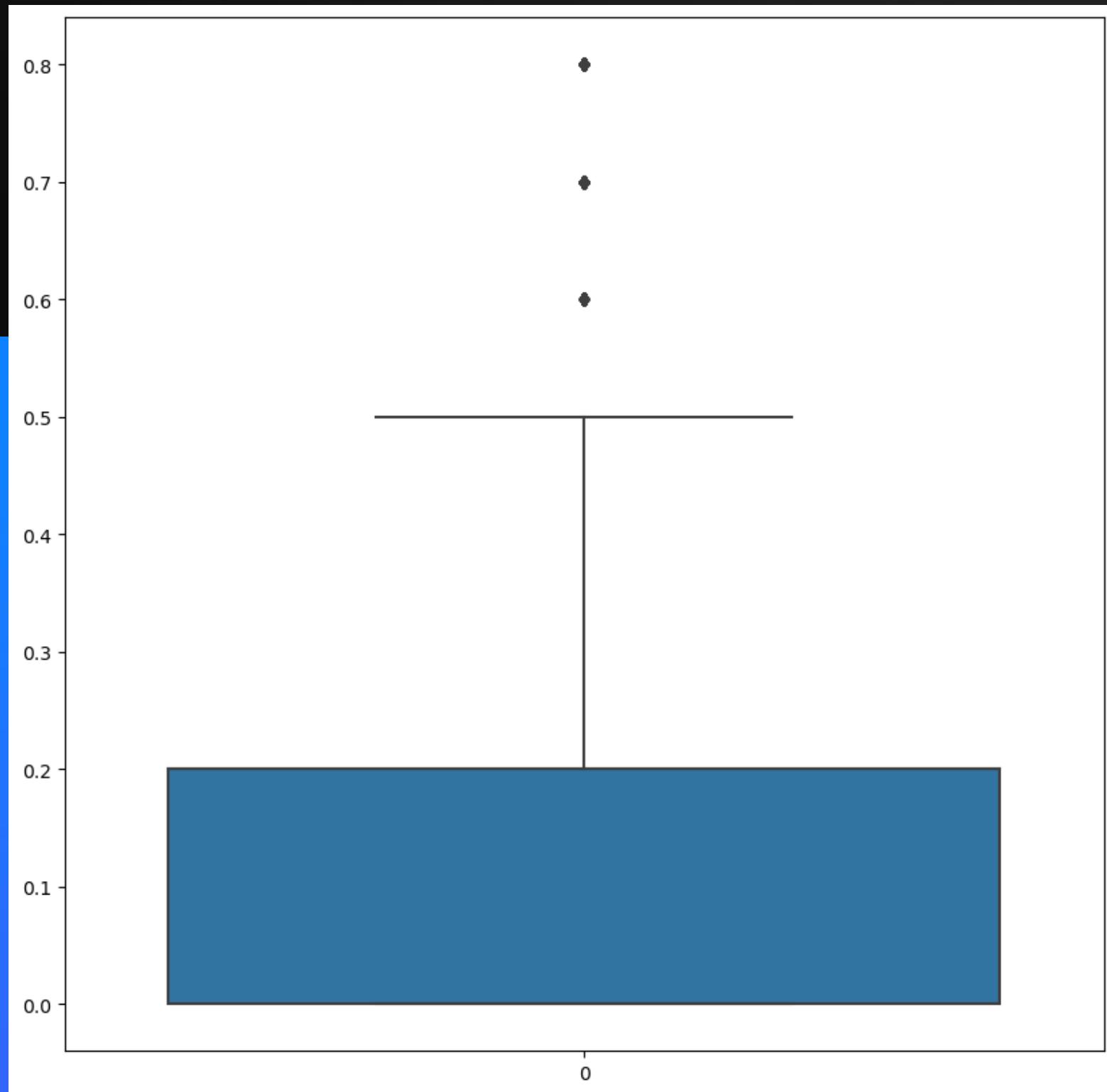
```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(10, 4))
sns.lineplot(x='Discount', y='Profit', data=df, color='r', label='Discount')
plt.legend()
# plt.show()
```

<matplotlib.legend.Legend at 0x267afbe1960>

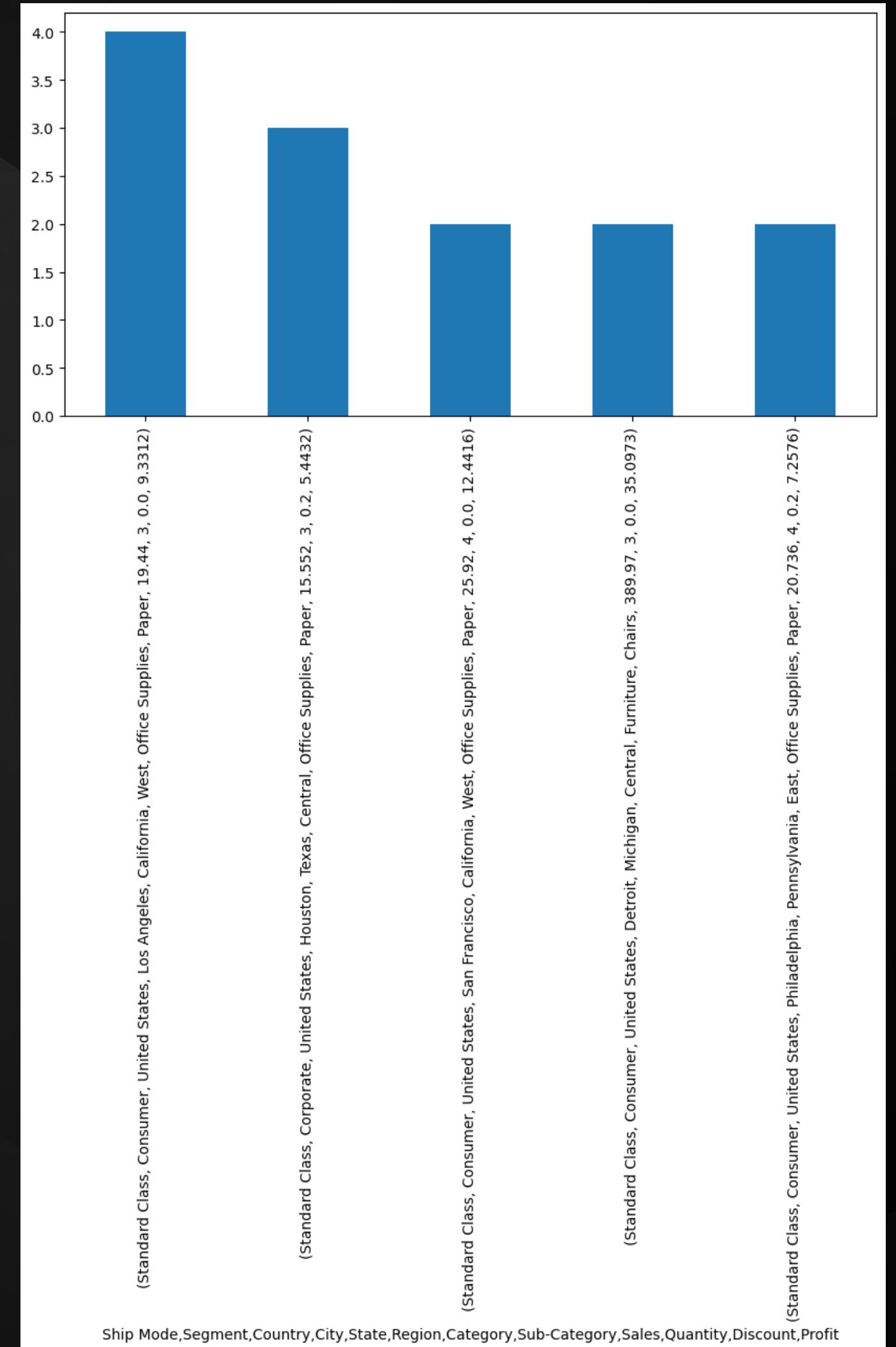


Total profit and Sales per Sub-Category





1. Profit in south and central is less.
2. Profit In east and west region is better than south and central.
3. The Highest profit is earned in copiers while the selling of phones and chairs are extremely high compared to other products.
4. Another interesting fact-peoples don't prefer to buy tables and Bookcases from superstore as sales is medium but they are facing loss.
5. The store has wide variety of office supplies especially in Binders and Papers.
6. Negative correlation between profit and Discount.
7. Total sum of profit in sale of tables is negative.
8. Profit is more in sale of copiers.
9. No or very less profit in sales of supplies.
10. Technology segment is more profitable.





HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

Steps to be performed for analysis

- Drawing from my domain knowledge or prior experience, I provided context-specific insights that go beyond the basic analysis

1. Define the problems

2. Importing Libraries

3. Importing the dataset

4. Cleaning of data-Pre-Processing of data

5. summarizing data set to extract insights

6. Visualization Of Data

7. Conclusion

MODELLING

Importing Libraries

```
[ ]: import pandas as pd  
      import numpy as np  
      import matplotlib.pyplot as plt  
      import seaborn as sns  
      from sklearn.cluster import KMeans      MODELLING  
      from pandas.plotting import scatter_matrix
```

Importing the DataSet

```
[ ]: ds = pd.read_csv("/content/SampleSuperstore.csv")
```

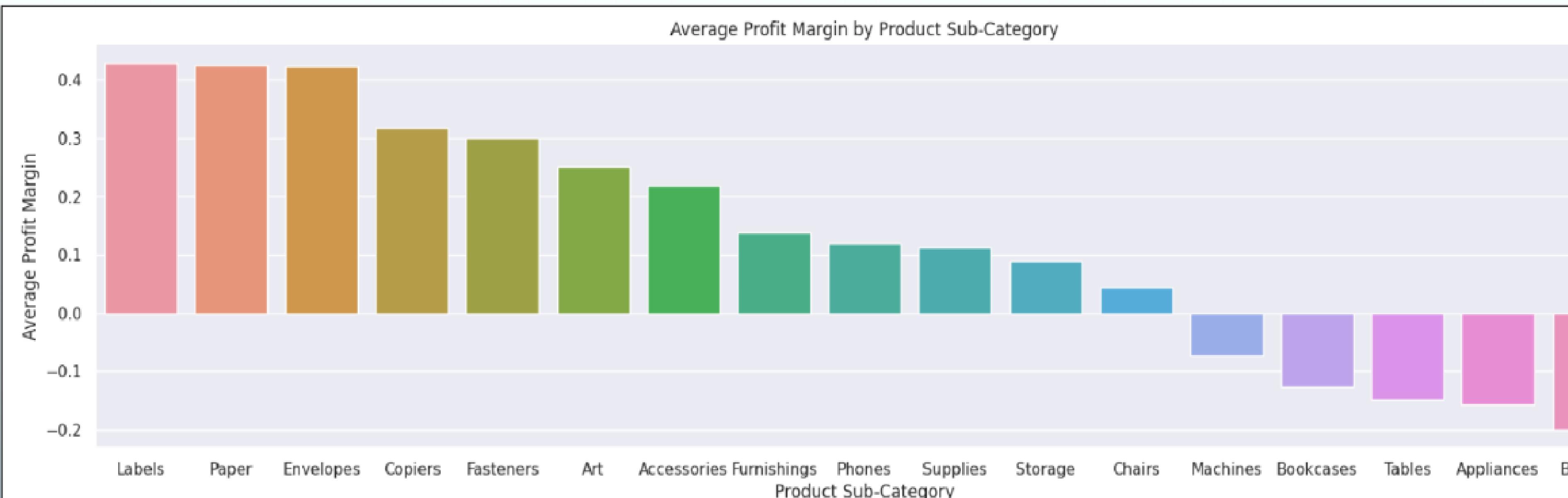
- Your Data cleaning and preprocessing is performed to ensure accurate and reliable analysis results.
- Exploratory data analysis (EDA) is conducted to reveal patterns, trends, and relationships in the data.
- Key performance metrics such as sales revenue, profit, and customer segments are investigated.
- Geographical sales distribution and identified potential target markets are examined.

paragraph text

MODELLING AND INSIGHTS

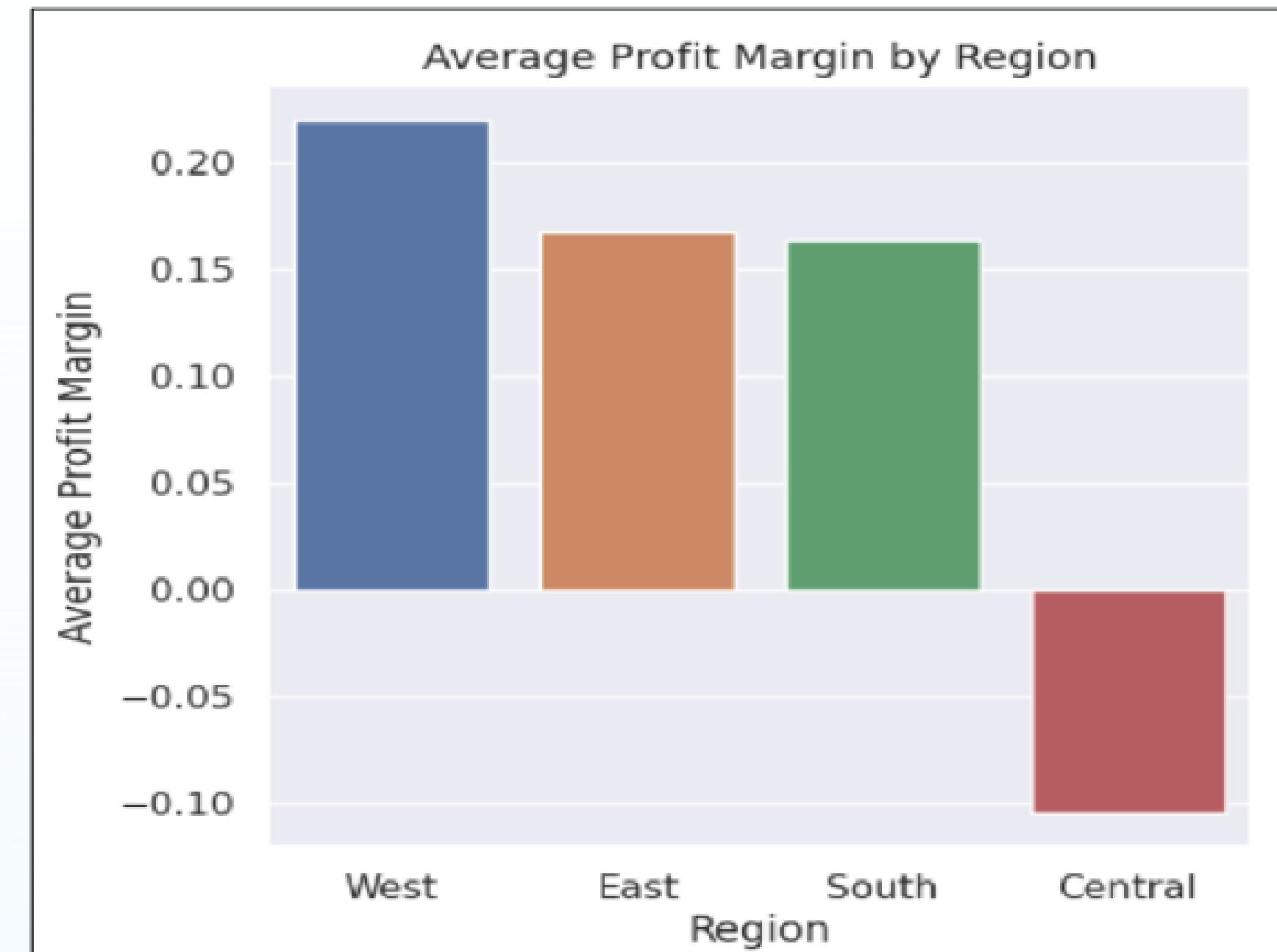
➤ Data Visualization:

Advanced data visualization techniques using tools like Python libraries (e.g., Matplotlib, Seaborn) were used to create visually appealing and informative charts and graphs. These visualizations facilitated the effective interpretation of the analysis results and provided a clear understanding of key findings.



MODELLING AND INSIGHTS

These modelling techniques formed the pillars of the Project – **Analysis of Superstore Dataset** for Data Analytics, ensuring a systematic and data-driven approach to extract valuable insights from the dataset.



```
df['Profit Margin'] = df['Profit'] / df['Sales']
# Group the data by region and calculate the average profit margin for each region
avg_profit_margin_by_category = df.groupby(['Region'],as_index=False)[['Profit Margin']].mean().sort_values(by='Profit Margin',ascending=False).head(50)
sns.set(rc={'figure.figsize':(5,5)})
sns.barplot(data = avg_profit_margin_by_category, x = 'Region',y= 'Profit Margin')
plt.title("Average Profit Margin by Region")
plt.xlabel("Region")
plt.ylabel("Average Profit Margin")
plt.show()
```



RESULTS

1. Profit in south and central is less.
2. Profit In east and west region is better than south and central.
3. The Highest profit is earned in copiers while the selling of phones and chairs are extremely high compared to other products.
4. Another interesting fact-peoples don't prefer to buy tables and Bookcases from superstore as sales is medium but they are facing loss.
5. The store has wide variety of office supplies especially in Binders and Papers.
6. Negative correlation between profit and Discount.
7. Total sum of profit in sale of tables is negative.
8. Profit is more in sale of copiers.
9. No or very less profit in sales of supplies.
10. Technology segment is more profitable.



LINKS

Github Link-

<https://github.com/codinggaurav7/IBM-Skillsbuild-Edunet-Foundation-Project->

Dataset

<https://github.com/codinggaurav7/IBM-Skillsbuild-Edunet-Foundation-Project-/blob/main/Analysis%20of%20Super%20Store%20-%20DA.csv>