SUPER STORE DATA ANALYSIS

This assignment is part of a business analytics internship offered by TTBTE as a component of my second interview round.

Made by: Sunita Bisht
Email Id: bishtsunita2002@gmail.com

Problem Statement

The primary objective of this analysis is to identify the key factors impacting sales and profitability in the Superstore business. We aim to uncover patterns, trends, and opportunities for improvement to optimize business strategies and maximize profitability.

Project Overview

Super Store Data Analysis

01

Objective 1

Identify top-performing product categories, regions, or customer segments and analyze their sales performance and profitability.

02

Objective 2

Uncover trends, patterns, and opportunities for improvement to enhance business success and optimize strategies.

03

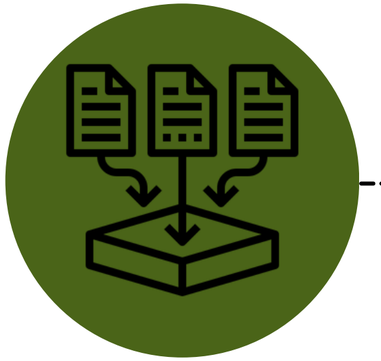
Objective 3

Provide actionable insights to maximize profitability by leveraging the identified top-performing areas and addressing factors impacting business success.

DataSet Description

The analysis utilized a dataset sourced from the Superstore database, encompassing comprehensive information regarding sales transactions. This dataset includes details such as sales category, states, city, segments, regions, profit, product categories, and sales figures. In total, the dataset comprises 9994 rows and 21 columns.

MODELLING

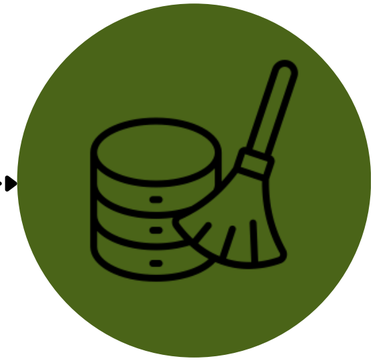


Data Extraction

Obtained the Sample Superstore dataset as the primary data source for the analysis.

Understand Data

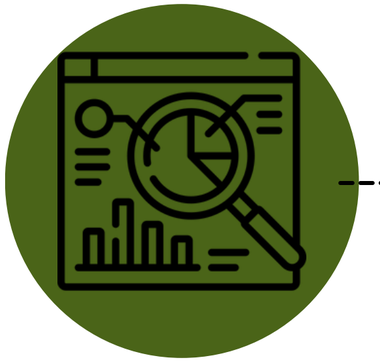
Explored the dataset to gain a comprehensive understanding of its structure and variables .
Reviewed data types, distributions, Columns, rows and any initial observations or patterns.



Data Cleaning

Conducted data cleaning procedures to ensure data quality .
Handled missing values, removed duplicates, and performed necessary data transformations.

MODELLING



Statistical Analysis

Descriptive Analysis : Utilized descriptive statistical techniques, such as calculating mean, median, and standard deviation, to summarize and explore key variables.
Generated descriptive statistics to understand the central tendency, variability, and distribution of the data.

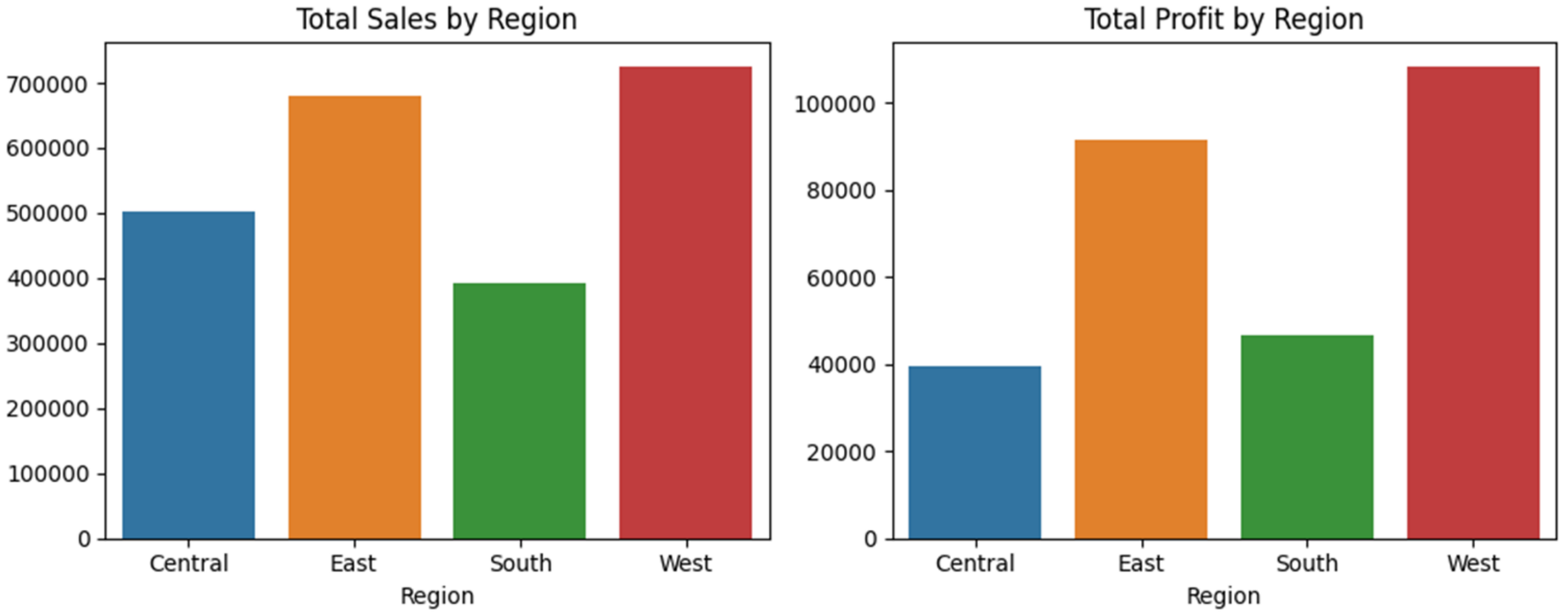
Data Visualization

Utilized various visualization techniques, including bar plots, pie charts, line plots, and scatter plots, to effectively communicate the findings of the analysis.
Created visually appealing and informative graphs and charts to present the insights to stakeholders.

Result and Outputs

Profit Anaylsis

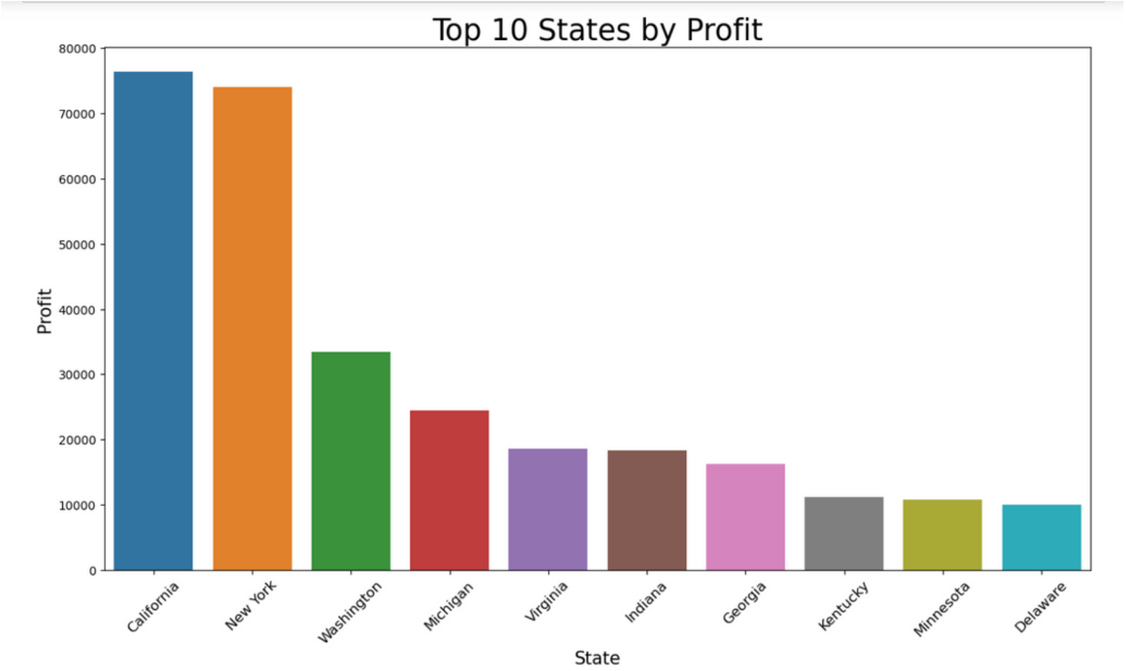
Analyzed the sales and profit distribution by region. The left chart displays the total sales by region, highlighting regions with high sales volume. The right chart shows the total profit by region, identifying regions with strong profit margins. These insights inform strategic decision-making, resource allocation, and revenue optimization.

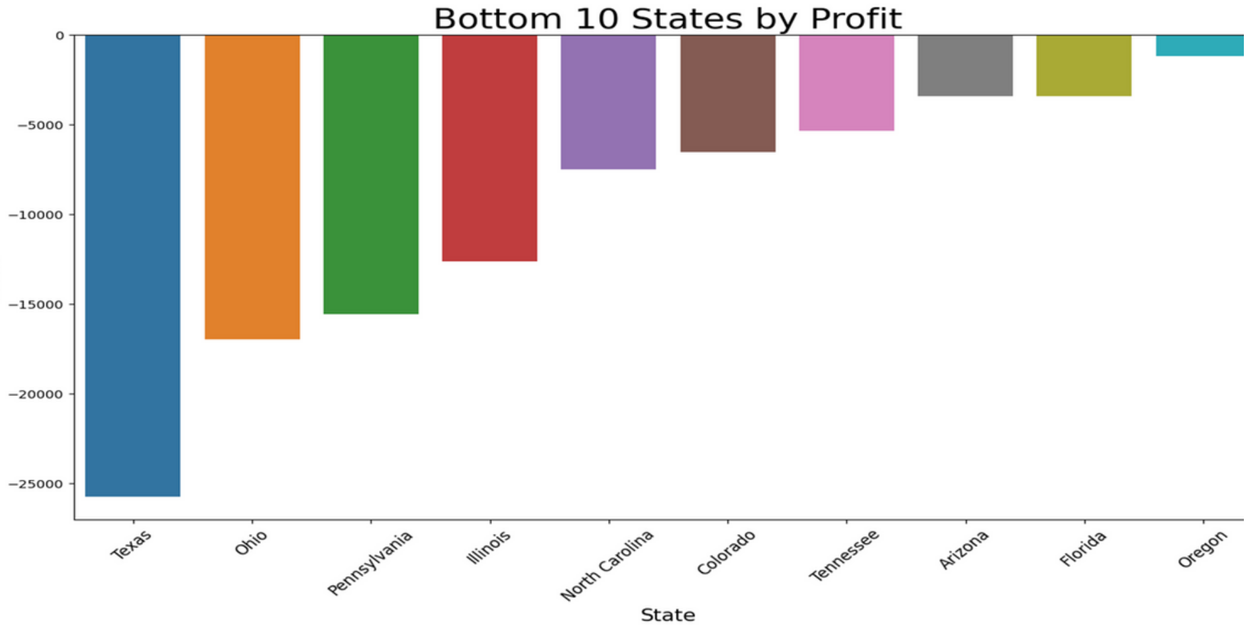


Result and Outputs

BOTTOM AND TOP 10 STATES BY PROFIT

I have analyzed the profitability of different states and identified the bottom 10 states based on profit generated.The analysis of the bottom 10 states by profit provides valuable insights into areas where the company can focus on improving profitability

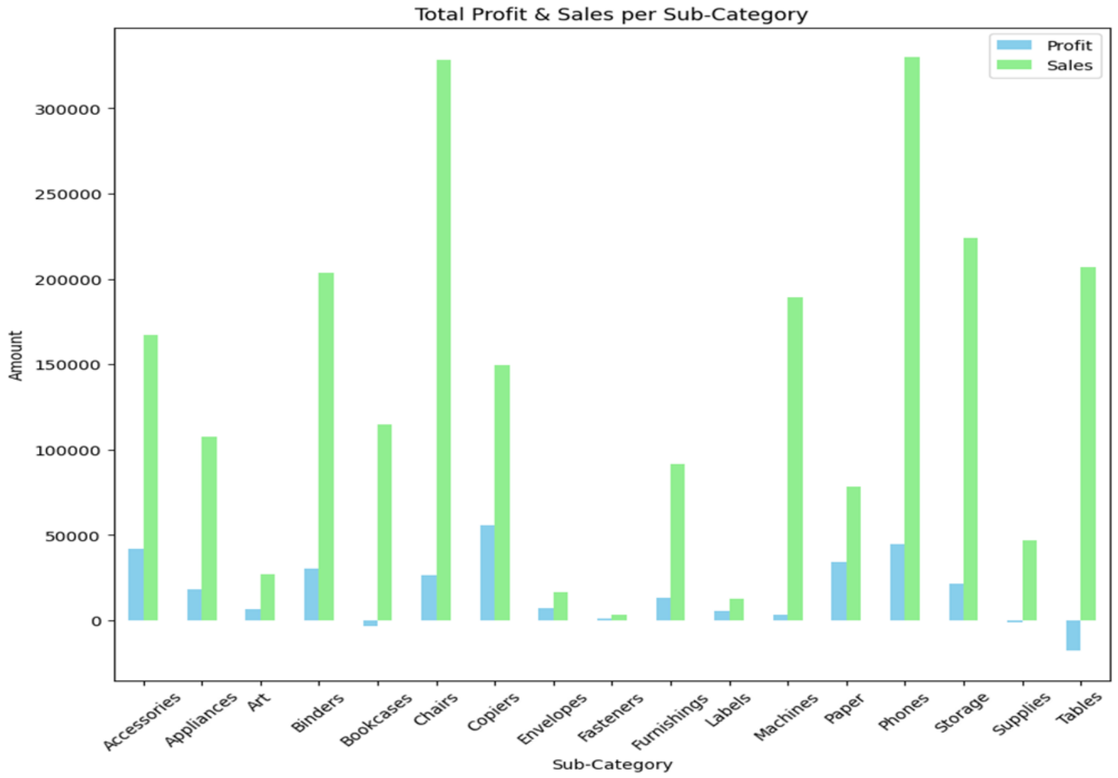


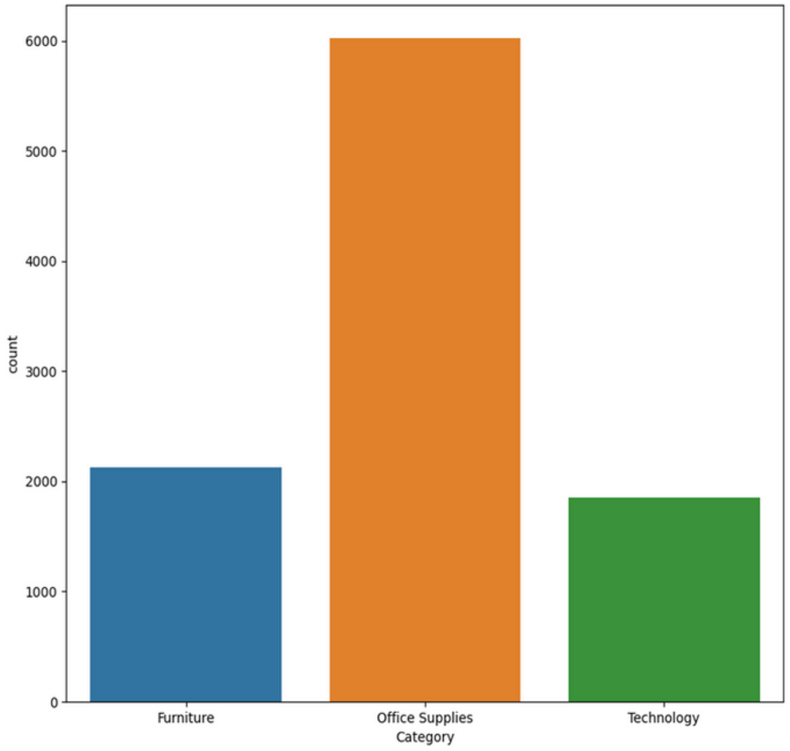


Result and Outputs

Product Category Analysis

In the product category analysis, we examined the distribution of products across different categories. The bar plot showcases the count of each product category, providing insights into the composition and prevalence of products in the dataset.





Result and Outputs

Distribution Analysis

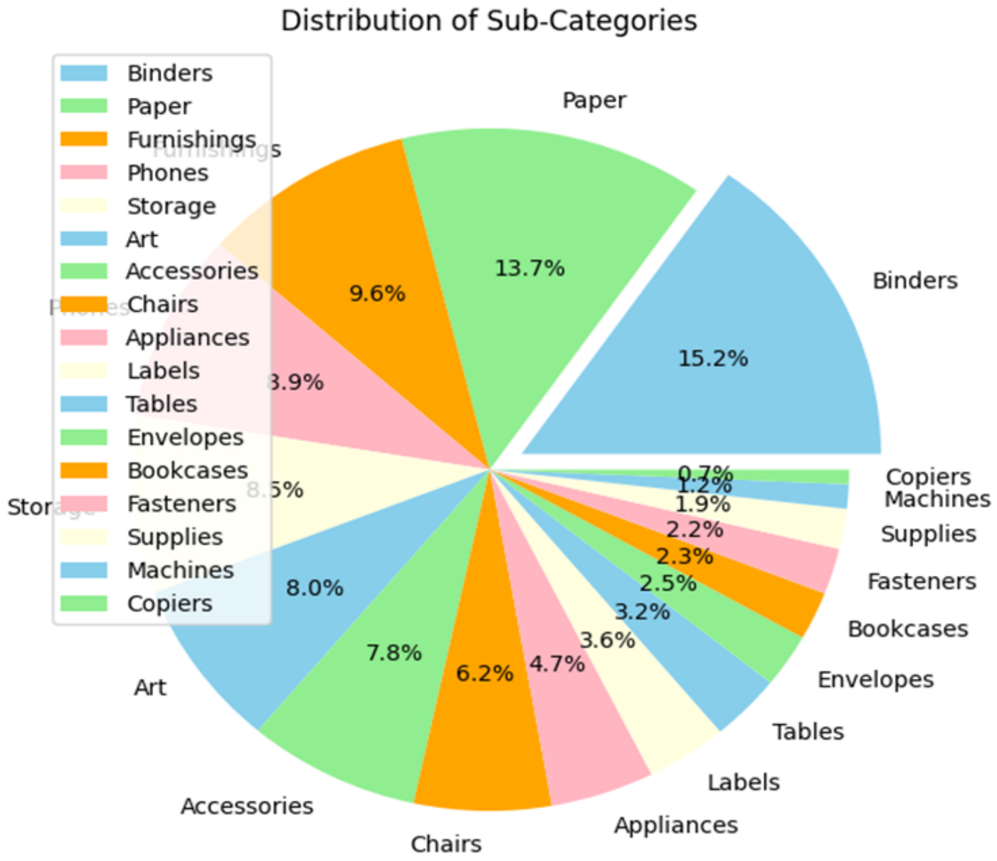


Figure: Distribution of Sub-Categories

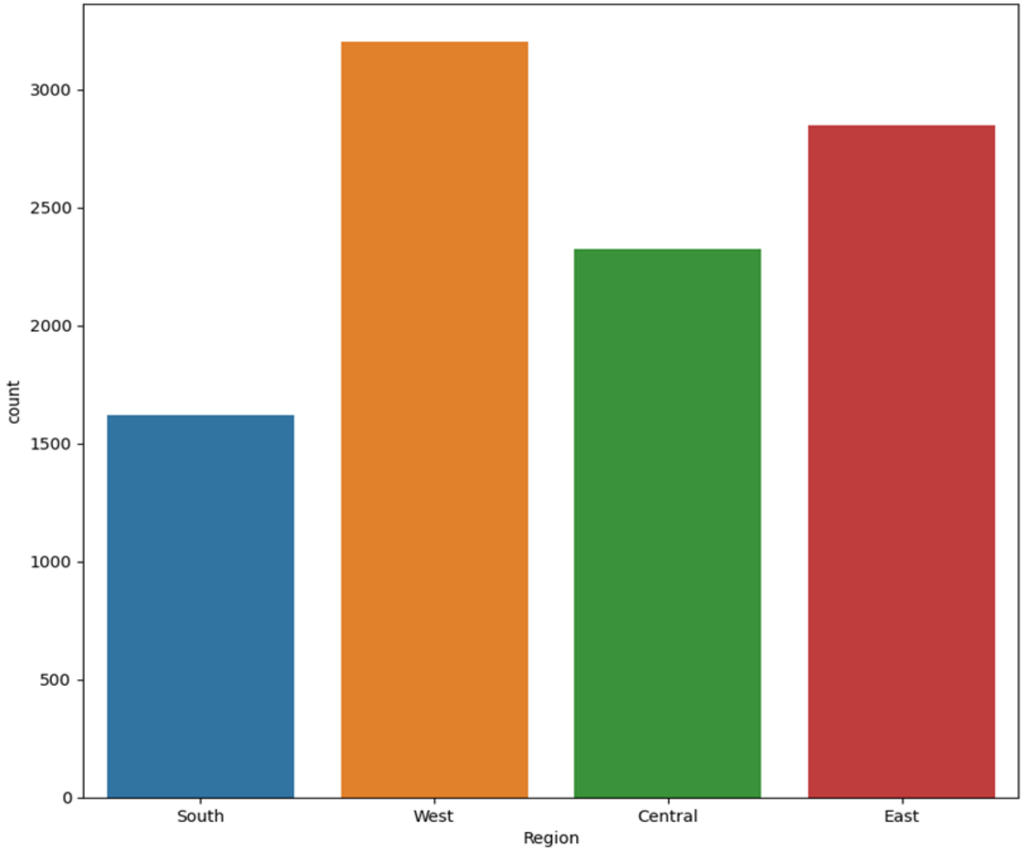


Figure: Distribution of Sales by Region

Result and Outputs

Additional Visualization

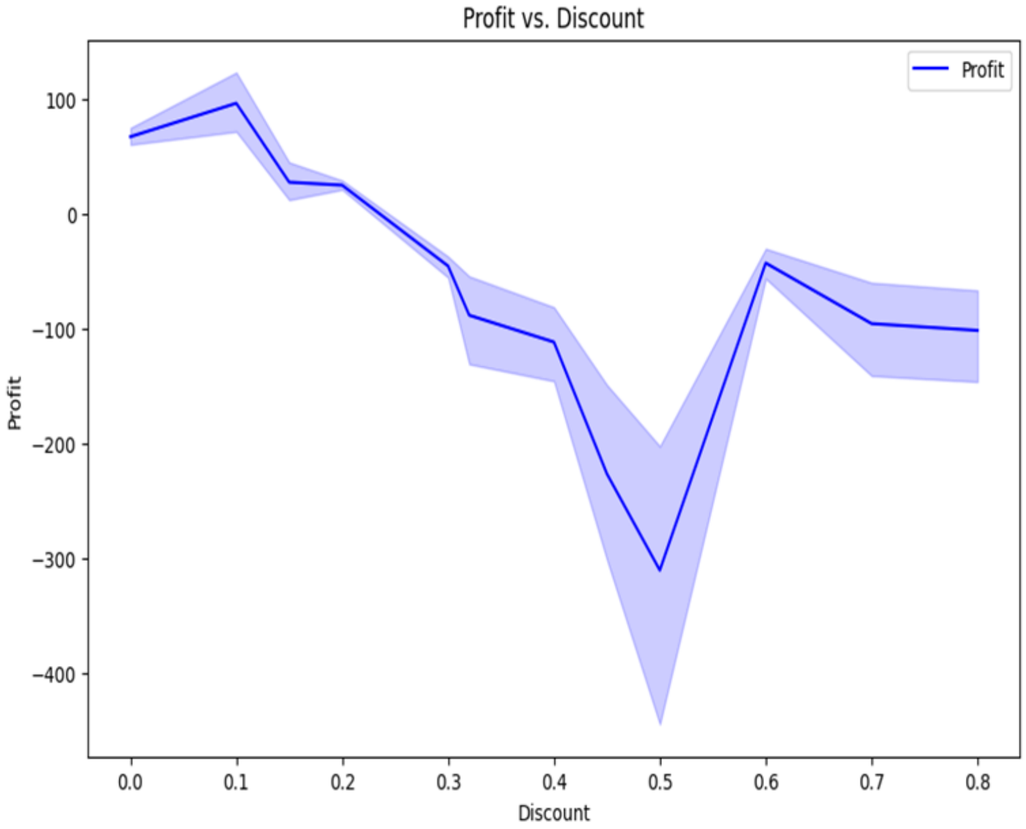


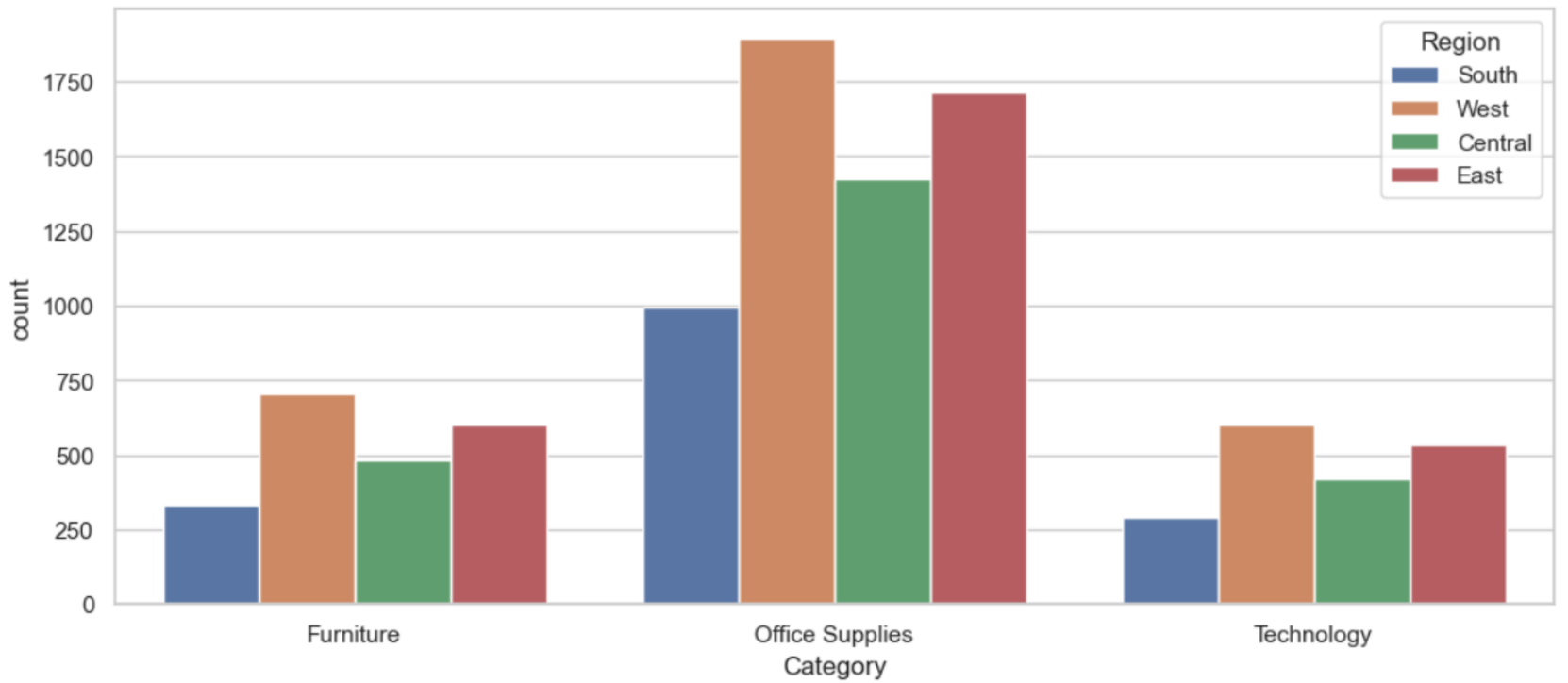
Figure: Profit vs. Discount

Result and Outputs

Count of Sub-Category Region-Wise

COUNT OF SUB-CATEGORY REGION-WISE

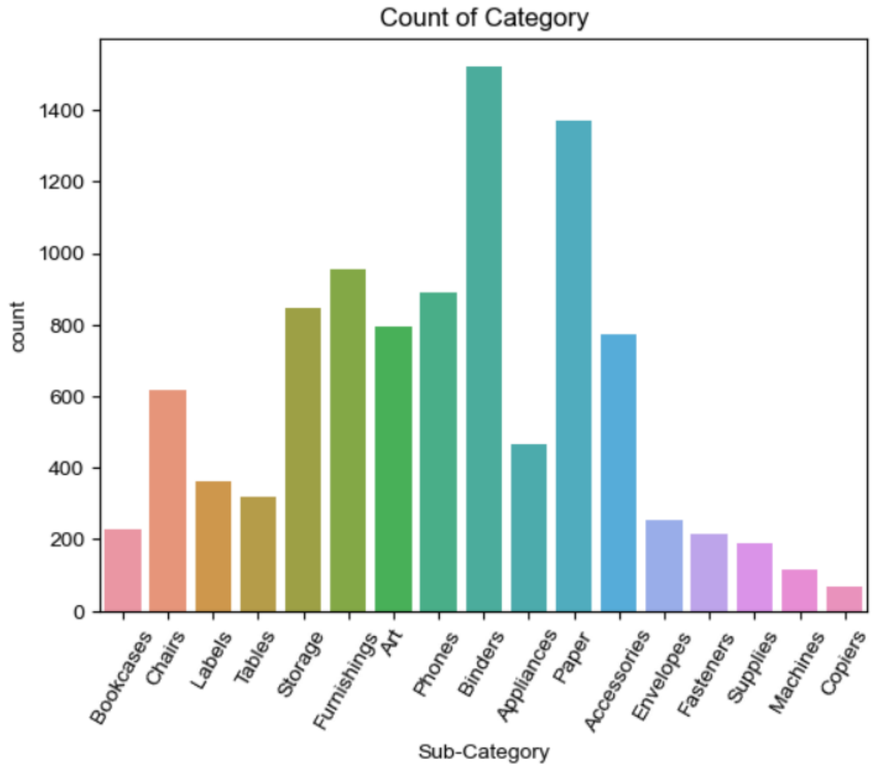
```
In [67]: plt.figure(figsize=(12,5))
         sns.countplot(x="Category", hue="Region", data=df)
         plt.show()
```



It is observed in the above line of code that people residing in Western part of US tend to order more from superstore.

Result and Outputs

Count of Category



Customer Analysis

```
In [69]: #Top 10 customers who order frequently
         df_top10= df['Customer Name'].value_counts().head(10)
         df_top10

Out[69]: William Brown            37
         John Lee                 34
         Matt Abelman             34
         Paul Prost               34
         Chloris Kastensmidt      32
         Seth Vernon              32
         Jonathan Doherty         32
         Edward Hooks             32
         Zuschuss Carroll         31
         Emily Phan               31
         Name: Customer Name, dtype: int64
```

Figure: Top 10 customers who order frequently

Result and Outputs

Best Performing Product

In [63]:
```
fur_entr=df[df['Category']=='Furniture']
```

In [65]:
```
fur_group=fur_entr.groupby(['Category', 'Product Name'])['Quantity'].sum().reset_index().sort_values('Quantity',
                                                                                                     ascending = False)
fur_group=fur_group[['Product Name','Quantity']].set_index('Product Name')
fur_group
```

Out[65]:

Product Name	Quantity
KI Adjustable-Height Table	74
Situations Contoured Folding Chairs, 4/Set	64
Staple-based wall hangings	62
Chromcraft Round Conference Tables	61
Eldon Wave Desk Accessories	61
...	...
Atlantic Metals Mobile 2-Shelf Bookcases, Custom Colors	3
Ultra Commercial Grade Dual Valve Door Closer	2
Barricks Non-Folding Utility Table with Steel Legs, Laminate Tops	2
Bush Saratoga Collection 5-Shelf Bookcase, Hanover Cherry, *Special Order	1
Global Enterprise Series Seating Low-Back Swivel/Tilt Chairs	1

CONCLUSION

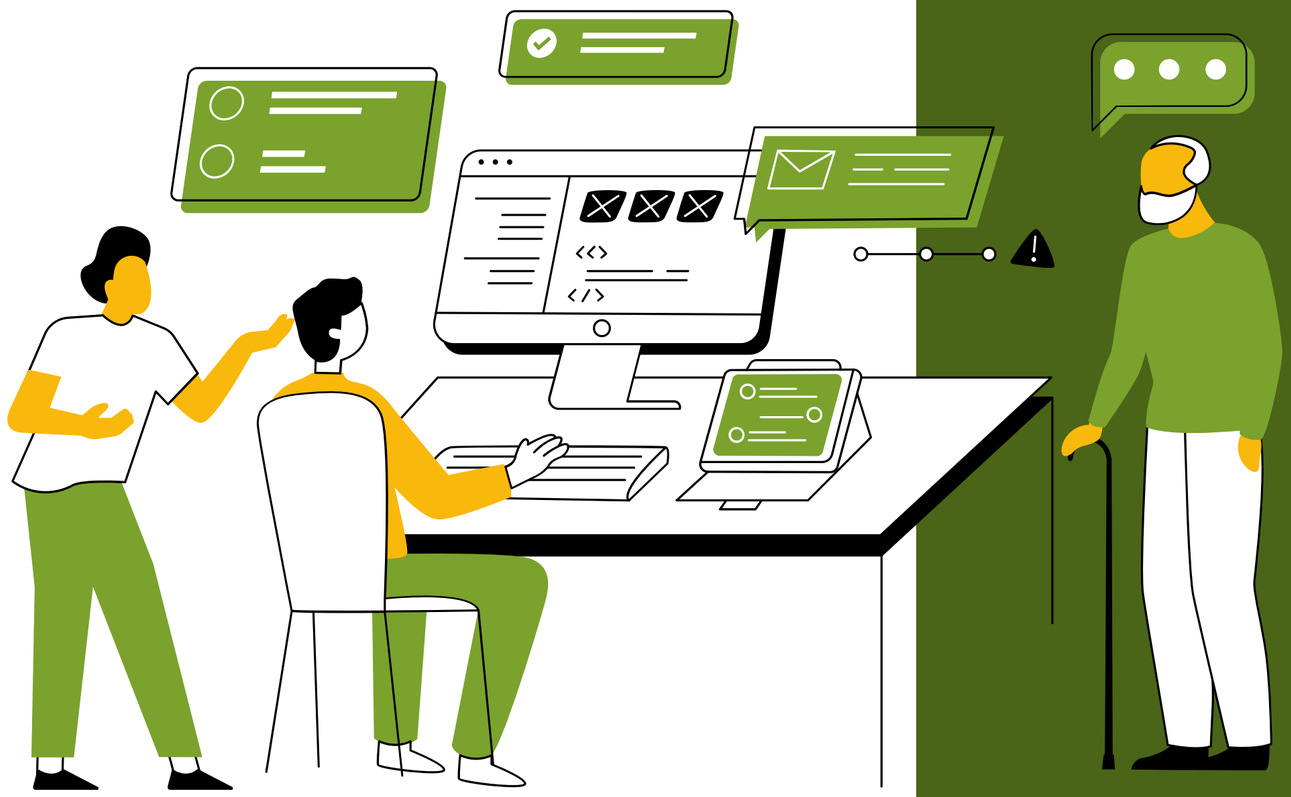
The analysis reveals that the Home Office segment contributes significantly to the overall profit.

Profit is high in South & less in central part

Highest profit is earned in copiers while selling price for chairs & phones is extremely high compared to other products

Technology category is more number in count

In region wise more count is in west region and many more conclusions can be made.



THANK YOU