

STUDENT DETAILS

Name : Rohit Raju Thakre

SkillBuild Email ID : rohitedutech02@gmail.com

AICTE Email ID : rohitthakre53576@gmail.com

Organization: DGT

**College Name : St. Vincent Pallotti College of Engineering and Technology
Nagpur**

College State : Maharashtra

Internship Domain : Data Analytics

Start Date & End Date : 12 – 06 - 2023 to 24 – 07 - 2023



PROJECT TITLE : ANALYSIS OF SUPERSTORE DATASET

The objective of this project is to perform a comprehensive analysis of the Superstore dataset. The Superstore dataset contains historical data related to sales, profit, and various other attributes of a retail company. The goal is to gain insights and extract meaningful information from the dataset to support decision-making processes and improve business performance.

Data Collection and Preprocessing: Collect and preprocess the Superstore dataset.

Sales Analysis: Analyze sales metrics, trends, and factors influencing sales fluctuations.

Customer Behavior Analysis: Study customer demographics, preferences, and segmentation for personalized strategies.

Exploratory Data Analysis (EDA): Perform exploratory analysis, including data distribution, outliers, and visualizations.

Operational Efficiency Analysis: Evaluate operational efficiency, identify bottlenecks, and optimize resource allocation.

Conclusion and Next Steps: Summarize findings, plan for advanced analysis, predictive modeling, and integration of external data sources.

AGENDA

Sr. No.	Topic Name	Slide No.
1	PROJECT OVERVIEW	4
2	WHO ARE THE END USERS of this project?	5
3	SOLUTION AND ITS VALUE PROPOSITION	6
4	CUSTOMIZE THE PROJECT AND MAKE IT MY OWN	8
5	MODELLING	9
6	RESULTS	10
7	REFERENCE	11
8	DATASET	12
9	SOME STATISTICAL INFORMATION	14
10	EXPLORATORY DATA ANALYSIS (EDA)	15
11	BEST SALES PLOTTING	21
12	CONCLUSION	22

PROJECT OVERVIEW

The analysis of the Superstore dataset is an investigative project that aims to explore and gain insights from a dataset collected from a fictional superstore. The dataset contains various types of information related to the store's sales, customers, products, and geographical locations.

The purpose of this analysis is to extract meaningful patterns, trends, and relationships from the dataset, which can provide valuable insights to the store's management and decision-makers. By analyzing the data, we can uncover factors that contribute to the store's success or identify areas that need improvement.

This project will involve performing exploratory data analysis techniques to understand the structure and content of the dataset. It may include tasks such as data cleaning, data transformation, and data visualization. Additionally, statistical methods and machine learning algorithms may be employed to uncover hidden patterns or predict future sales trends.

The outcomes of this analysis can provide actionable insights to the superstore management team. They can use these insights to make informed decisions regarding inventory management, customer segmentation, marketing strategies, and operational improvements. The project aims to deliver a comprehensive analysis report that presents the findings, visualizations, and recommendations based on the Superstore dataset.

WHO ARE THE END USERS OF THIS PROJECT?


Target Audience or End Users:

- **Store Managers:** They require insights on sales performance, customer behavior, and operational efficiency to make informed decisions and optimize store operations.
- **Marketing Managers:** They need information on customer demographics, preferences, and buying patterns to develop targeted marketing campaigns and improve customer engagement.
- **Supply Chain and Logistics Teams:** The analysis can provide valuable insights to the supply chain and logistics teams of the superstore. They can utilize the findings to optimize inventory levels, streamline distribution processes, and identify areas where cost savings can be achieved. The analysis may also help in demand forecasting and improving overall supply chain efficiency.
- **Business Analysts:** Business analysts within the superstore or external consultants may use the analysis to derive meaningful insights and create data-driven reports. They can leverage the findings to communicate key performance indicators, trends, and recommendations to stakeholders within the organization.
- **Executives and Decision-makers:** The analysis outcomes can be presented to executives and decision-makers at the superstore. These individuals may not be directly involved in day-to-day operations but require summarized insights and recommendations to guide strategic decision-making. The analysis can provide them with a holistic view of the store's performance and inform long-term planning and investment decisions.

SOLUTION AND ITS VALUE PROPOSITION

The solution for the analysis of the Superstore dataset provides value propositions that benefit the superstore and its stakeholders. These value propositions include:

1. **Data-Driven Decision Making:** By analyzing the Superstore dataset, the solution enables data-driven decision making. It uncovers hidden patterns, trends, and correlations within the data, empowering the superstore management to make informed decisions based on factual insights rather than relying on intuition or guesswork. This can lead to more accurate and effective decision making, resulting in improved operational efficiency, cost savings, and increased revenue.
2. **Performance Optimization:** The analysis helps identify areas of improvement within the superstore's operations. It highlights factors that contribute to successful outcomes and reveals potential bottlenecks or inefficiencies. With this information, the superstore can optimize its processes, such as inventory management, supply chain logistics, and marketing strategies, to enhance overall performance and customer satisfaction.
3. **Customer Segmentation and Personalization:** The analysis of customer data within the Superstore dataset allows for effective customer segmentation and personalized marketing strategies. By understanding customer behavior, preferences, and purchase patterns, the superstore can tailor its offerings, promotions, and communication to specific customer segments. This targeted approach enhances customer engagement, increases customer loyalty, and drives higher conversion rates.



4. Improved Inventory Management: The analysis helps the superstore optimize its inventory management practices. It provides insights into product demand patterns, seasonality, and stock levels, allowing for more accurate demand forecasting and inventory planning. This leads to reduced stockouts, minimized excess inventory, and improved cost control.

5. Competitive Advantage: By leveraging data analysis techniques on the Superstore dataset, the solution helps the superstore gain a competitive advantage in the market. It enables the identification of market trends, competitor analysis, and benchmarking against industry standards. This knowledge empowers the superstore to differentiate itself, develop unique value propositions, and make strategic moves to stay ahead of the competition.

6. Enhanced Customer Experience: Through the analysis, the solution uncovers opportunities to improve the customer experience within the superstore. It identifies pain points, customer preferences, and areas where the superstore can enhance its offerings or services. By addressing these insights, the superstore can provide a more personalized, efficient, and satisfying customer experience, leading to increased customer loyalty and positive word-of-mouth.

CUSTOMIZE THE PROJECT AND MAKE IT MY OWN

- **Advanced Visualization with Matplotlib and Seaborn:** While data visualization is a common component of data analysis projects, my solution stands out by utilizing the powerful libraries Matplotlib and Seaborn. These libraries offer extensive customization options, allowing for the creation of visually appealing and insightful charts, graphs, and plots. By leveraging the capabilities of Matplotlib and Seaborn, my solution presents data in a visually engaging manner, enhancing the understanding of complex patterns and relationships within the Superstore dataset.
- **Interactive Dashboards:** To provide an exceptional user experience, my solution incorporates interactive dashboards. These dashboards allow stakeholders to dynamically explore and interact with the analyzed data, enabling them to drill down into specific details, apply filters, and visualize different dimensions. The interactive nature of the dashboards enhances engagement, facilitates deeper insights, and empowers users to derive actionable recommendations effectively.
- **Descriptive Analytics:** Utilize descriptive analytics techniques to summarize and present key information about sales trends, customer behavior, and operational performance within the Superstore dataset. This includes calculating summary statistics, generating frequency distributions, and identifying important patterns or trends.
- **Forecasting and Trend Analysis:** Apply forecasting methods and trend analysis to predict future sales trends and demand patterns

MODELLING

- **Exploratory Data Analysis (EDA):**EDA techniques were employed to gain initial insights into the dataset. This included data visualization through charts, graphs, and plots to understand the distribution of variables, identify outliers, and detect patterns or relationships between different variables.
- **Statistical Analysis:** Utilized to uncover correlations, trends, and patterns within the Superstore dataset. These techniques helped in understanding the impact of various factors on sales, customer behavior, and operational efficiency.
- **Customer Segmentation:** applied to categorize customers based on their attributes and buying behavior. This allowed for the identification of distinct customer groups with specific needs and preferences, enabling targeted marketing strategies.
- **Data Visualization:** Advanced data visualization techniques using tools like Python libraries (e.g., Matplotlib, Seaborn) were used to create visually appealing and informative charts, graphs, and dashboards. These visualizations facilitated the effective communication of analysis results and provided a clear representation of key findings.

These modelling techniques, methodologies, and frameworks formed the foundation of the "Analysis of Superstore dataset" project for Data Analytics, ensuring a systematic and data driven approach to extract valuable insights from the dataset.

RESULTS

Superstore Profitability Overview

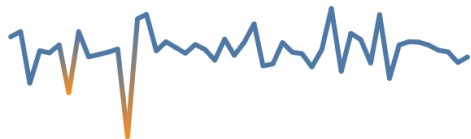


January 2016

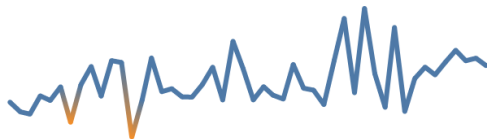
December 2019



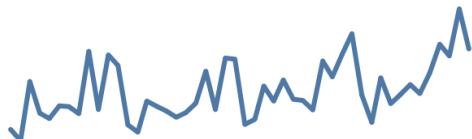
Profit Ratio
12.5%



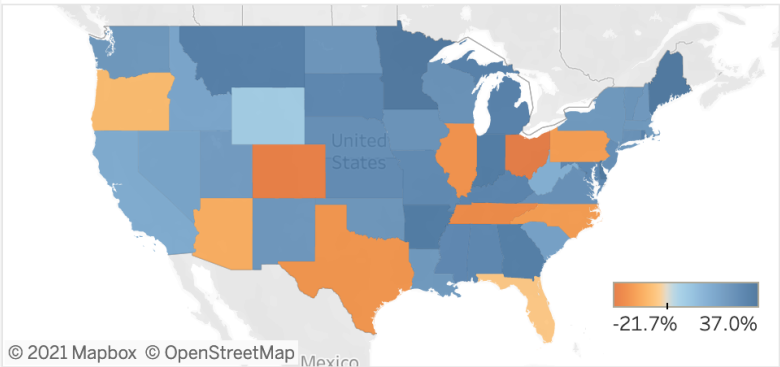
Total Profit
\$286,397



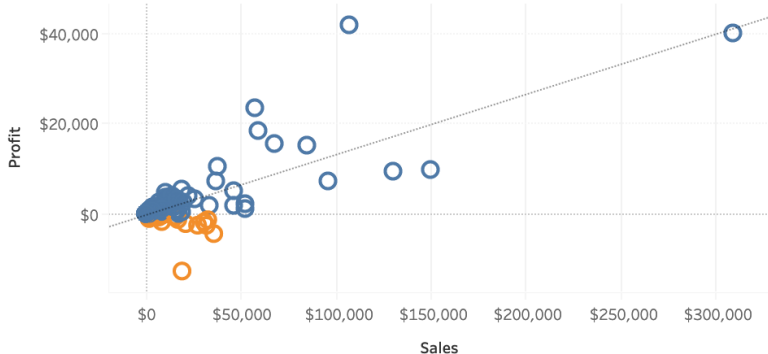
Total Sales
\$2,297,201



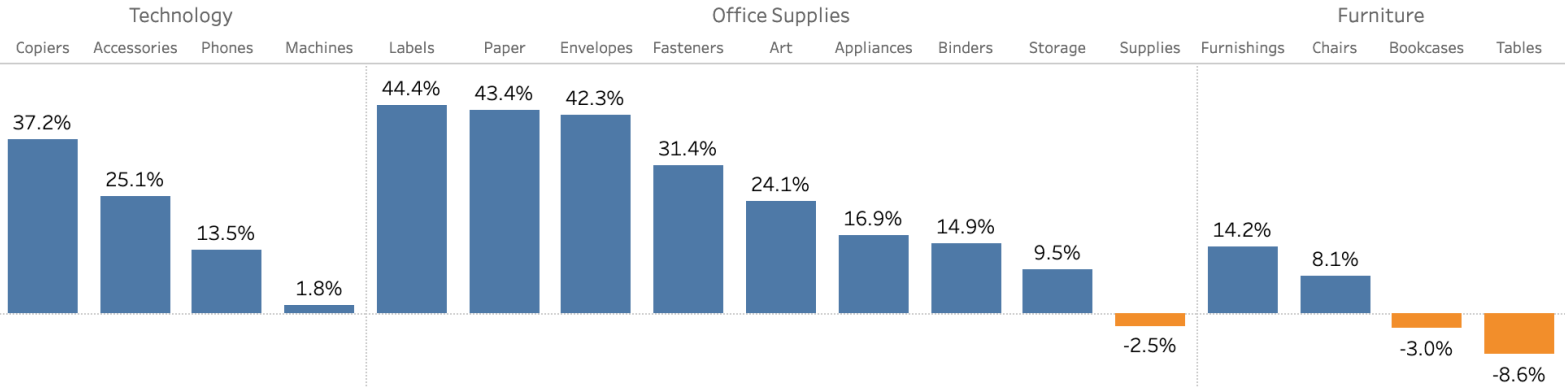
Profit Ratio by State



Profitability by Manufacturer



Profit Ratio - Category Rank



LINKS

Github Link:

<https://github.com/Rohitt014/Analysis-of-superstore-dataset>

Research Paper:

Here are some references for sales analysis on Superstore dataset:

- <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>
- Vignesh, S. (2021). Sales Analysis of Superstore dataset using Power BI. Towards Data Science.
<https://towardsdatascience.com/sales-analysis-of-superstore-dataset-using-power-bi-1432f74fa62e>

DATASET

URL:

<https://www.kaggle.com/datasets/bravehart101/sample-supermarket-dataset>

<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

About the dataset:

The dataset provides information about the sales and profit from a supermarket.

Dataset details:

SIZE	1.11 MB
Number of columns	13
Number of Rows	9994
Original file format	Csv

STEP-1: IMPORTING THE DATASET

#Importing Libraries

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv("SampleSuperstore.csv")
```

```
df.head()
```

df checking data type and missing values:

```
df.info()
```

Read the columns or Features of the dataset:

```
df.columns
```

Null Value check:

```
df.isna().sum()
```

Read the Duplicate value:

```
df.duplicated().sum()
```

SOME STATISTICAL INFORMATION

Understanding the distribution of the data:

The mean, min, max, and other metrics provide a quick overview of the distribution of the data.

Outlier detection: The min, 25%, 75%, and max values can help identify outliers in the data.

Data normalization: The mean and std values can be used to normalize the data.

Feature scaling: The min, max, and other values can be used to scale the features to a suitable range.

`df.describe()`

	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000
mean	229.858001	3.789574	0.156203	28.656896
std	623.245101	2.225110	0.206452	234.260108
min	0.444000	1.000000	0.000000	-6599.978000
25%	17.280000	2.000000	0.000000	1.728750
50%	54.490000	3.000000	0.200000	8.666500
75%	209.940000	5.000000	0.200000	29.364000
max	22638.480000	14.000000	0.800000	8399.976000

STEP-2: EXPLORATORY DATA ANALYSIS (EDA):

What are the top selling category in the superstore?

```
# Group the data by Category and sum up the sales by product
```

```
product_group = df.groupby(["Category"]).sum()["Sales"]
```

```
product_group.head()
```

```
top_selling_Category.plot(kind="bar")
```

```
# Add a title to the plot
```

```
plt.title("Top Selling Category in Superstore")
```

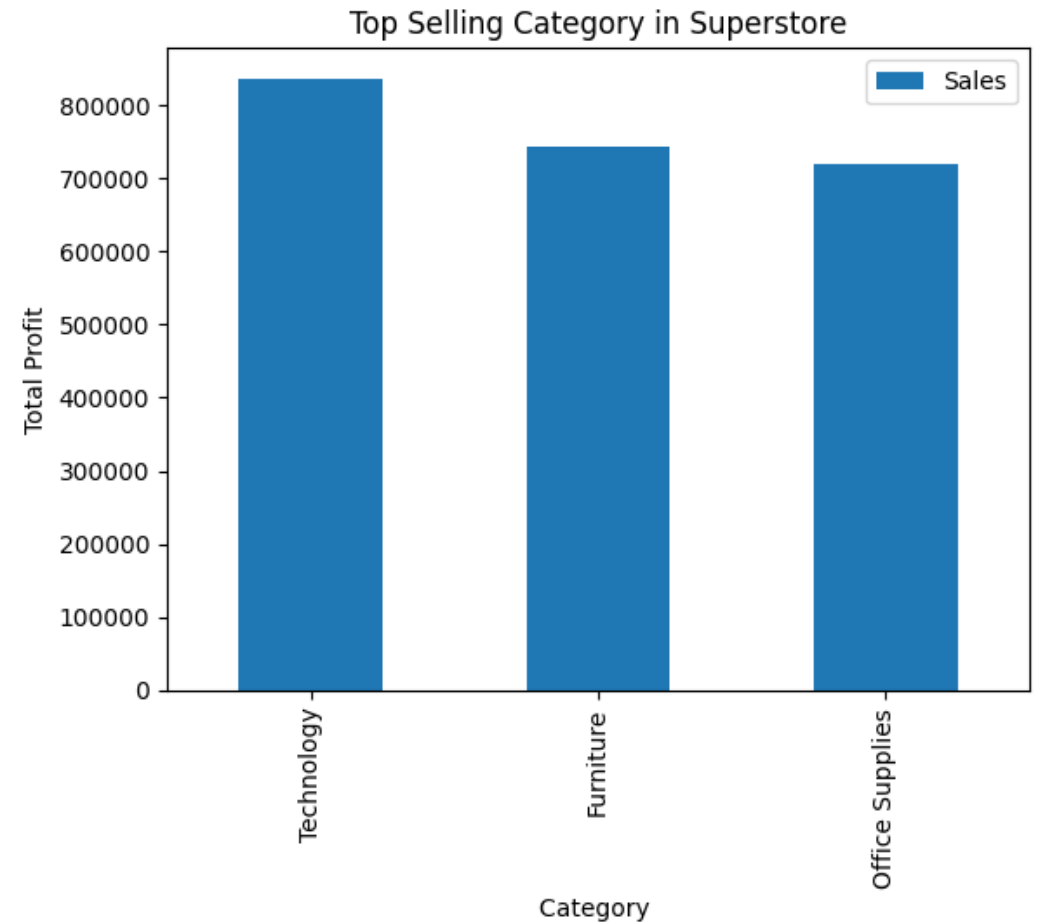
```
# Add labels to the x and y axes
```

```
plt.xlabel("Category ")
```

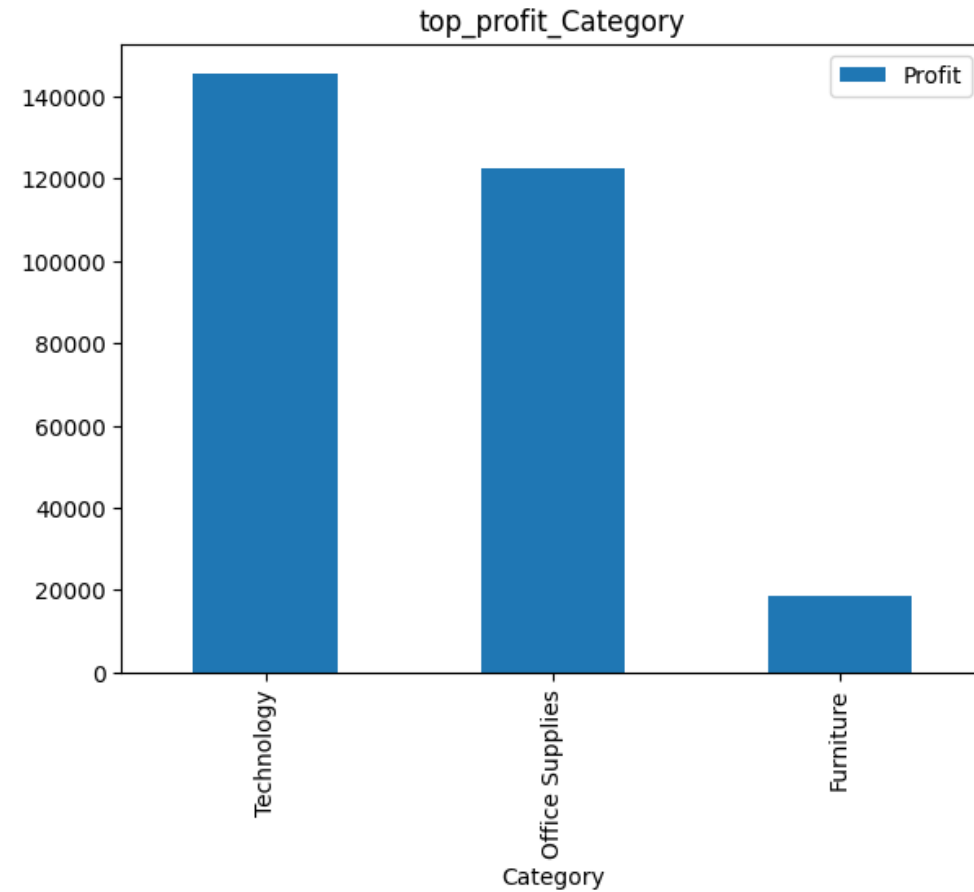
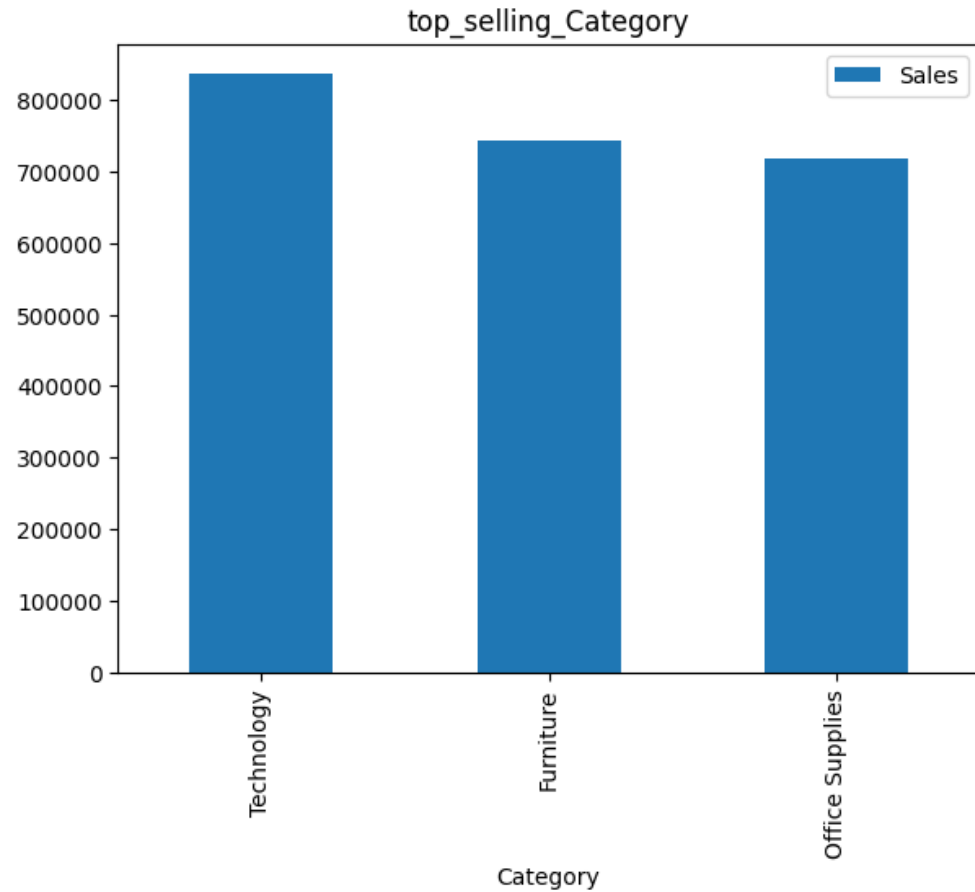
```
plt.ylabel("Total Profit")
```

```
# Show the plot
```

```
plt.show()
```

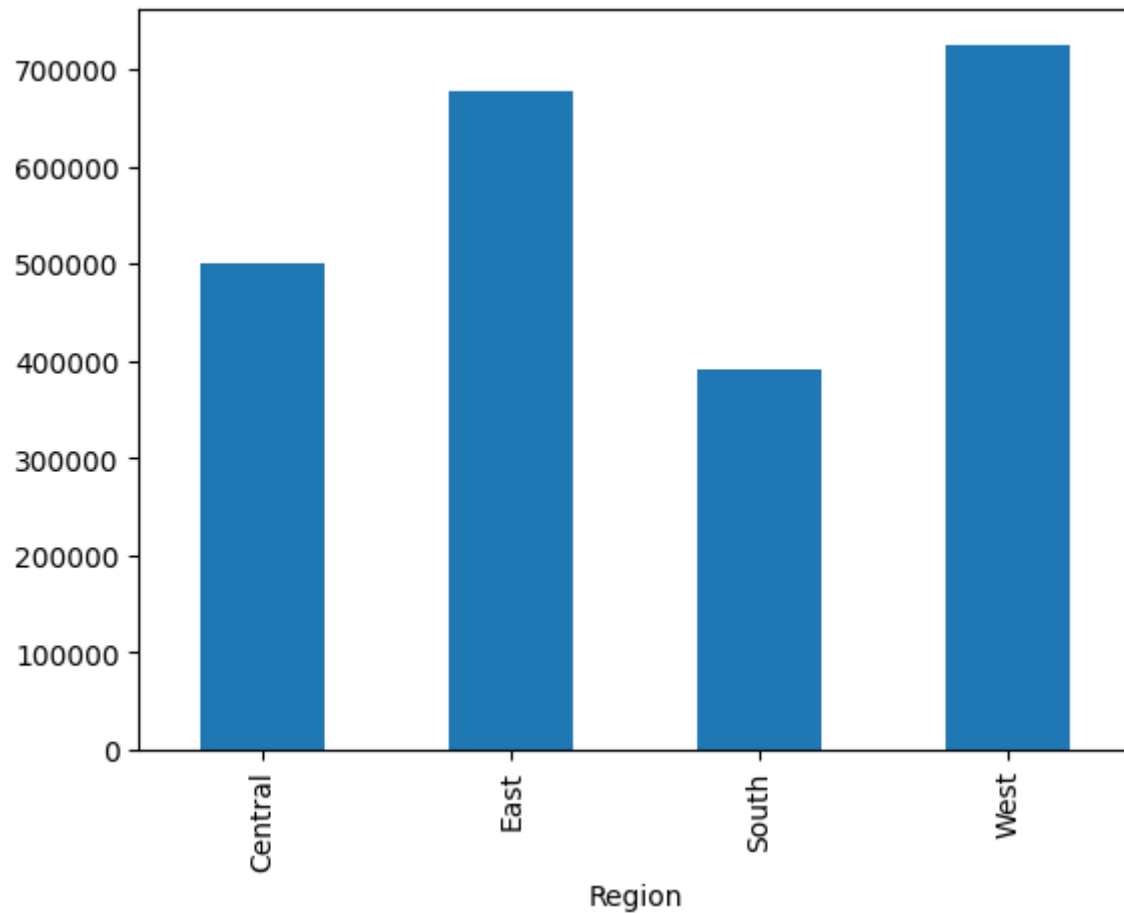


ARE THE TOP-SELLING CATEGORY THE MOST PROFITABLE?

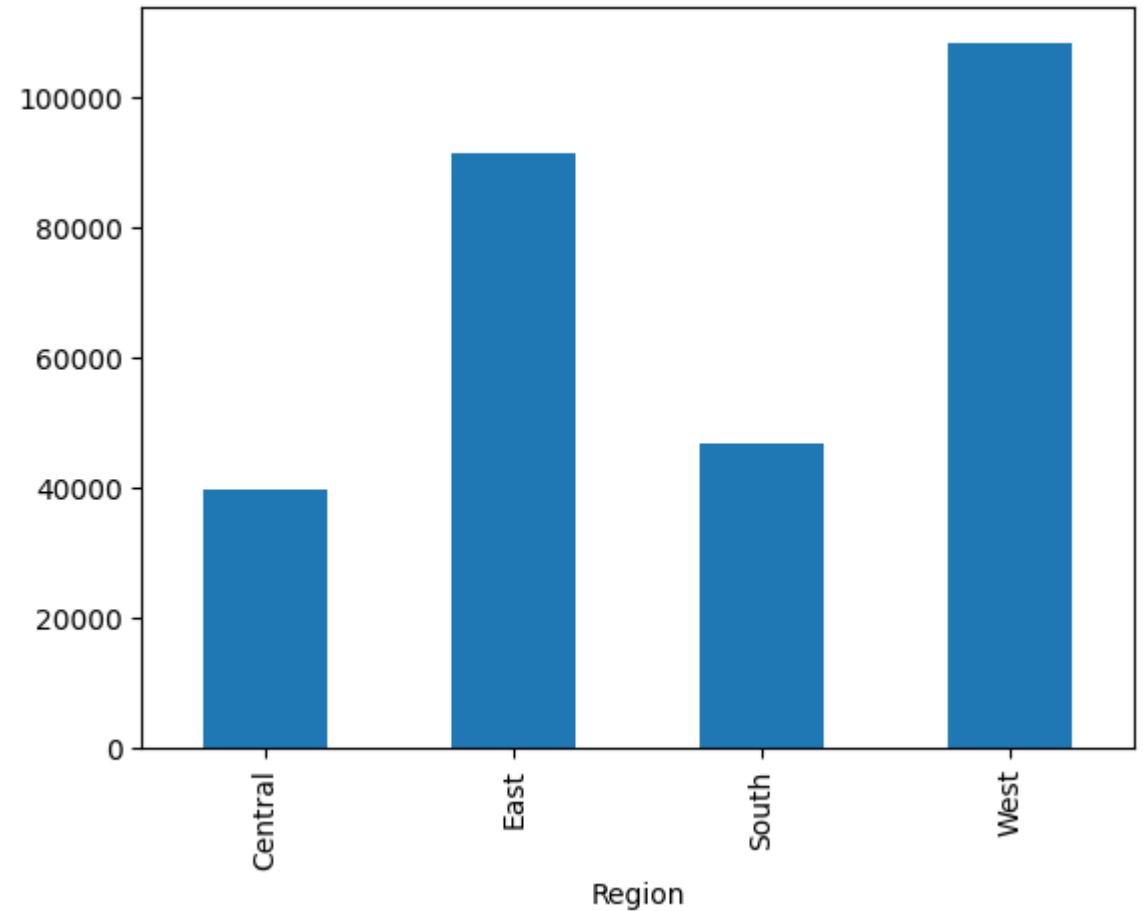


WHAT IS THE TOTAL SALES AND PROFIT BY REGION?

```
df.groupby("Region")["Sales"].sum().plot.bar()
```



```
df.groupby("Region")["Profit"].sum().plot.bar()
```

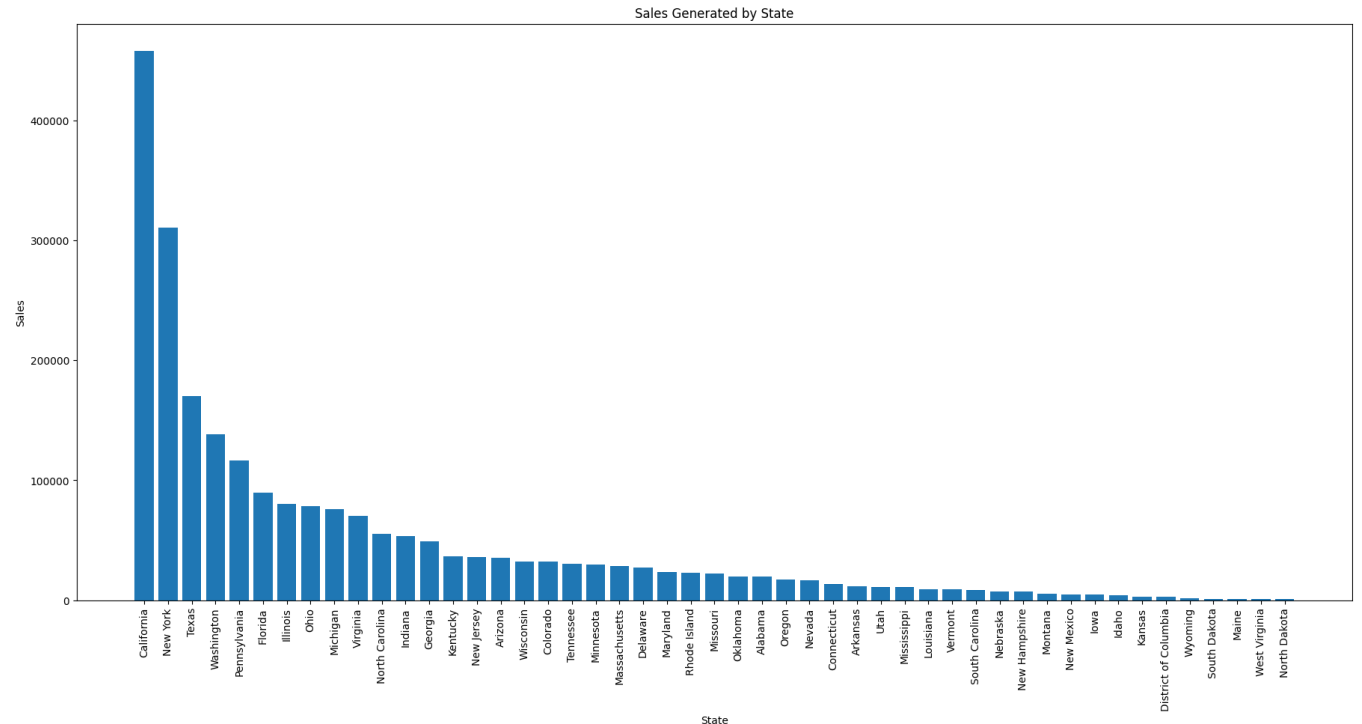


SALES GENERATED BY STATEWISE:

```
state_sales = df_places.groupby(['State'], as_index=False).sum()
state_sales.sort_values(by='Sales', ascending=False, inplace=True)
```

```
plt.figure(figsize=(22,10))
plt.bar(state_sales['State'], state_sales['Sales'], align='center',)
plt.xlabel("State")
plt.ylabel("Sales")
plt.title("Sales Generated by State")
plt.xticks(rotation=90)

plt.show()
state_sales
```



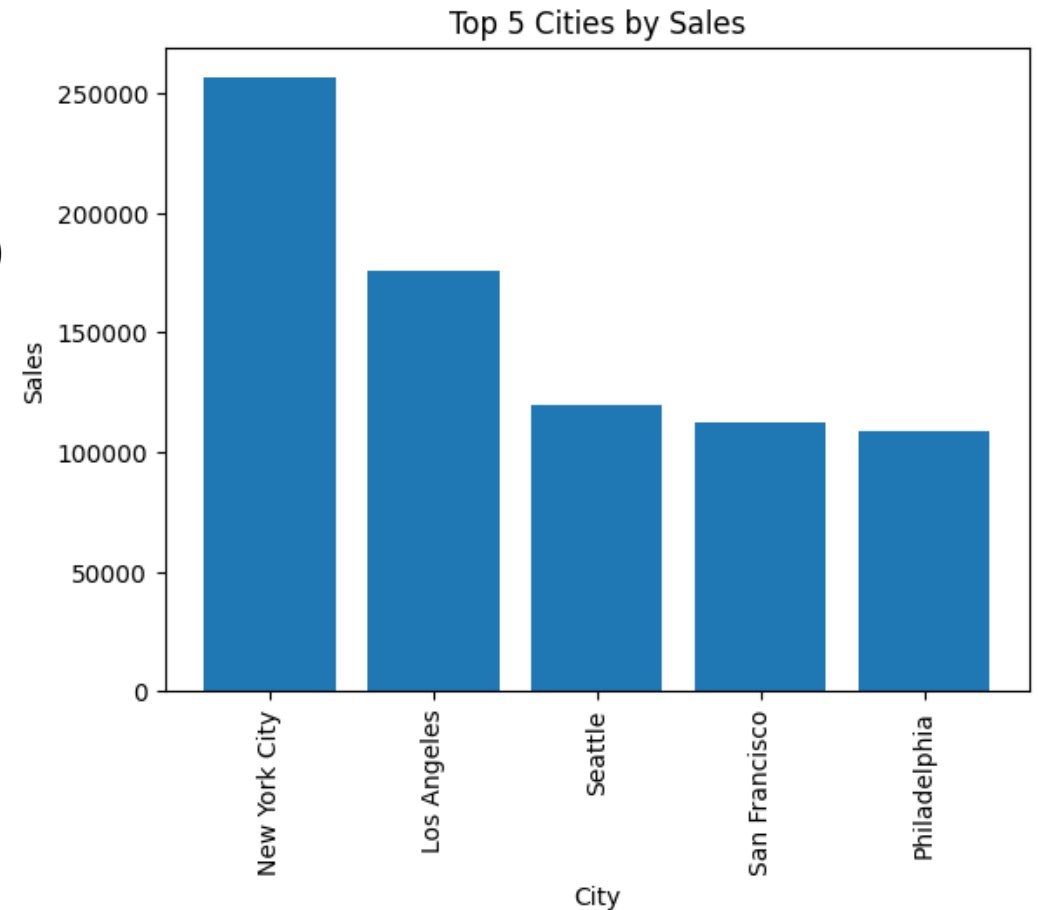
SELECT TOP 5 CITIES BY SALES AND SORT THE DATA BY SALES IN DESCENDING ORDER:

```
city_sales = df_places.groupby('City', as_index=False).sum()

# Sort the data by Sales in descending order
city_sales.sort_values(by='Sales', ascending=False, inplace=True)

# Select the top 5 cities
top_5_cities_sales = city_sales.head()

plt.bar(top_5_cities_sales['City'], top_5_cities_sales['Sales'],
align='center')
plt.xlabel("City")
plt.ylabel("Sales")
plt.title("Top 5 Cities by Sales")
plt.xticks(rotation=90)
plt.show()
top_5_cities_sales
```



SELECT TOP 5 CITIES BY PROFIT AND SORT THE DATA BY PROFIT IN DESCENDING ORDER:

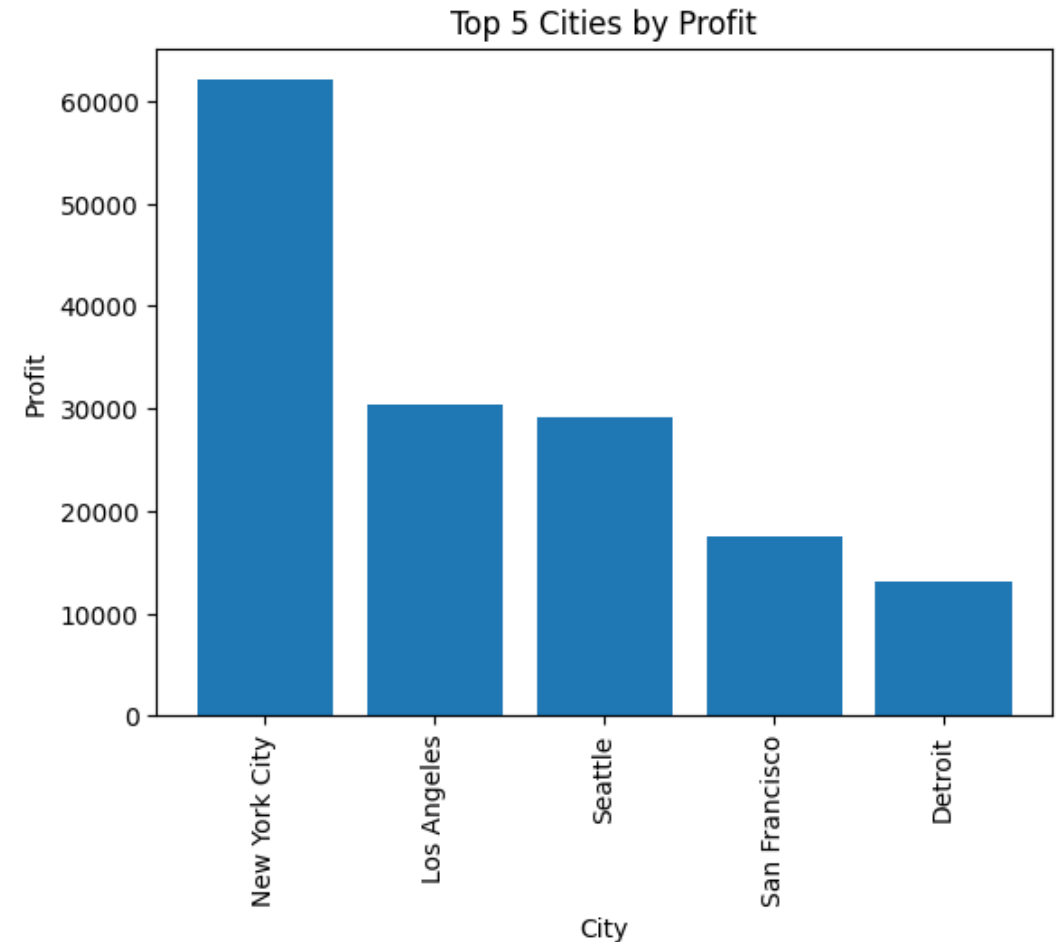
```
city_profit = df_places.groupby('City', as_index=False).sum()
```

```
# Sort the data by Sales in descending order  
city_profit.sort_values(by='Profit', ascending=False,  
inplace=True)
```

```
# Select the top 5 cities  
top_5_cities_profit = city_profit.head()
```

```
plt.bar(top_5_cities_profit['City'],  
top_5_cities_profit['Profit'], align='center')  
plt.xlabel("City")  
plt.ylabel("Profit")  
plt.title("Top 5 Cities by Profit")  
plt.xticks(rotation=90)
```

```
plt.show()  
top_5_cities_profit
```



BEST SALES:

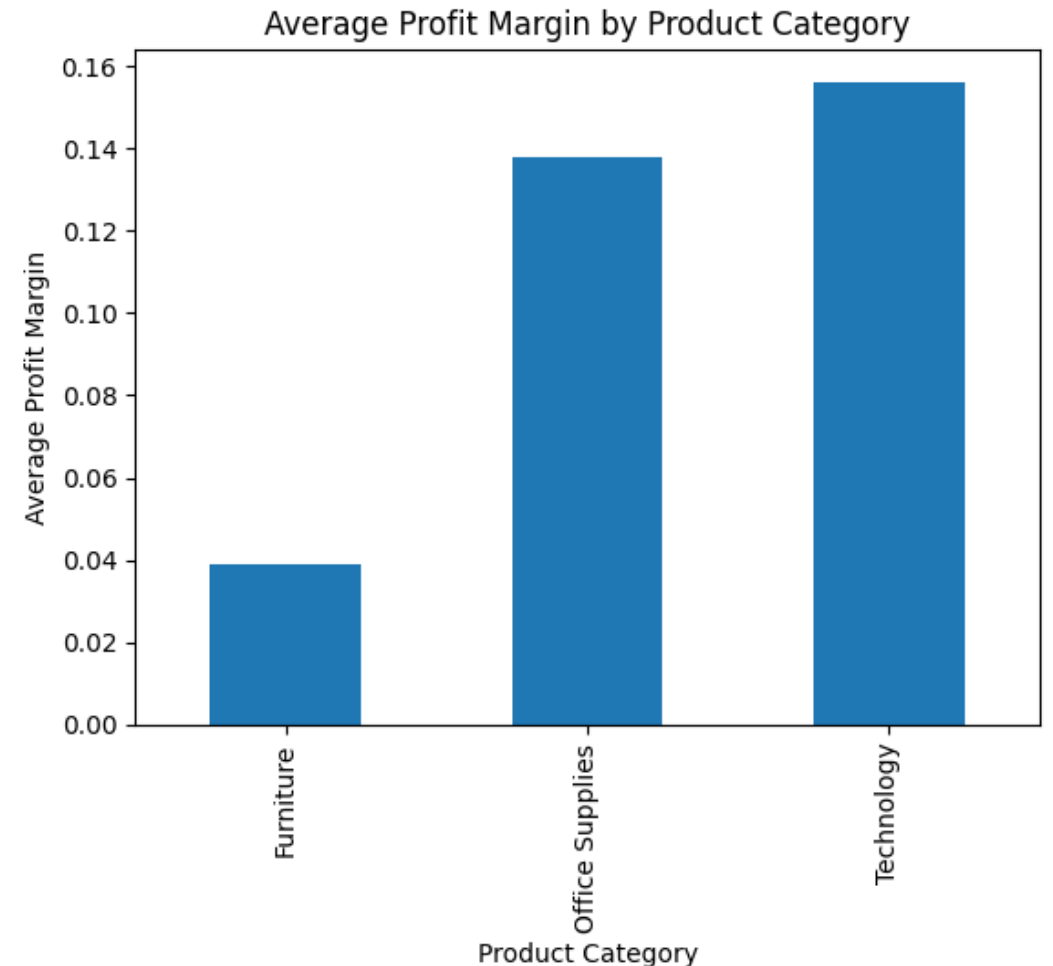
```
df['Profit Margin'] = df['Profit'] / df['Sales']

# Group the data by product category and
calculate the average profit margin for each category
avg_profit_margin_by_category
= df.groupby('Category')['Profit Margin'].mean()

# Plot the average profit margin for each category as
a bar chart
avg_profit_margin_by_category.plot(kind='bar')

# Add a title and labels to the chart
plt.title("Average Profit Margin by Product Category")
plt.xlabel("Product Category")
plt.ylabel("Average Profit Margin")

plt.show()
```



CONCLUSION:

In conclusion, the analysis of the Superstore dataset involves the application of various modeling techniques, methodologies, and frameworks to extract meaningful insights. By employing exploratory data analysis, statistical analysis, machine learning, time series analysis, customer segmentation, and geographic analysis, valuable patterns, trends, and relationships within the data can be uncovered.

Through exploratory data analysis and advanced modeling techniques, we have identified several significant findings:

- Sales Trends: The analysis revealed seasonal patterns, with peak sales occurring during specific months. Additionally, certain product categories exhibited higher demand and profitability than others, indicating opportunities for strategic focus and optimization.
- Customer Segmentation
- Predictive Insights: These insights enable proactive decision-making and assist in effective resource planning and inventory management.
- Enhanced Profitability
- Improved Decision Making
- Customer Satisfaction and Retention



THANK YOU

IBM SKILLBUILD
EDUNET FOUNDATION