Name: Anurag Pani
anuragpani100@gmail.com

Organization: DGT
College name: Institute of Technical Education and Research
State: Odisha
Domain: Data Analytics
S/E date: 12.06.2023 - 24.07.2023

Project Title: Analysis of Superstore Dataset

Introduction: The goal of this project is to analyze the Superstore dataset to gain insights into sales trends, customer behavior, and operational efficiency. The dataset contains information about various aspects of the store's operations, including sales, customer demographics, product categories, and geographical regions. By conducting a comprehensive analysis, we aim to identify opportunities for improvement and make data-driven recommendations to optimize store 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:

S.no.	Topics Name	Page no
1	Project Overview	3-10
2	Data set 1. Dataset loading 2. About the Dataset	11-12
3	Some Statistical Information	13
4	Exploratory Data Analysis (EDA) • What are the top selling products in the superstore? • What are the top profit products in the superstore? • What is the total Sales and Profit by region? • Select top 5 cities by profit and Sort the data by profit in descending order	14-21
5	•The Best Sales Plotting • Conclusion	22-23

Project Overview

The analysis on Superstore dataset is a comprehensive study that aims to analyze the sales performance of a fictional retail company called "Superstore". The dataset used in this analysis contains information about sales transactions, customers, products, and geographical locations. The analysis involves using Power BI, a data visualization and reporting tool, to create interactive dashboards and reports that provide insights into the sales performance of Superstore.

Purpose: The purpose of the "Analysis of Superstore dataset" is to gain insights into sales trends, customer behavior, and operational efficiency in order to optimize store performance and make data-driven recommendations for improvement.

Scope: The scope of the analysis includes examining the Superstore dataset, which consists of sales transactions, customer demographics, product categories, and geographical regions. The analysis will involve data cleaning, exploratory data analysis, sales analysis, customer behavior analysis, and operational efficiency analysis.

Objectives:

•Identify sales trends, such as seasonal patterns and fluctuations, to optimize inventory management and sales forecasting.

•Understand customer behavior by analyzing demographics, preferences, and purchase patterns to develop targeted marketing strategies and enhance customer satisfaction.

•Improve operational efficiency by identifying bottlenecks, streamlining processes, and optimizing resource allocation for enhanced profitability.

•Provide data-driven recommendations to optimize store performance, improve customer experience, and increase overall profitability based on the analysis findings.

WHO ARE THE END USERS:

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.

Characteristics and Needs:

•They seek comprehensive data analysis, visualizations, and actionable recommendations to identify areas for improvement, enhance profitability, and streamline operations.

Benefits from the Solution:

•They will benefit from optimized inventory management, improved sales forecasting, and streamlined operations, leading to increased profitability and better customer satisfaction.

•They will benefit from targeted marketing campaigns, enhanced customer engagement, and improved customer retention, resulting in increased sales and brand loyalty.

Solution and its value proposition:

The solution for the "**Analysis of Superstore dataset**" project involves conducting a comprehensive analysis of the Superstore dataset to gain insights into sales trends, customer behavior, and operational efficiency. This analysis will be carried out using various statistical and data mining techniques, as well as advanced visualization tools.

Value Proposition: Our solution provides the following value propositions:

•Data-Driven Decision Making: By analyzing the Superstore dataset, we enable data-driven decision making for store managers and marketing managers. They can make informed decisions based on comprehensive analysis, leading to improved store performance, optimized operations, and targeted marketing strategies.

•Enhanced Profitability: Our analysis helps identify opportunities for increasing sales, improving inventory management, and reducing costs, ultimately leading to enhanced profitability for the Superstore. By optimizing pricing strategies, identifying high-demand products, and streamlining operations, the store can maximize its revenue and profitability.

•Customer Insights and Personalized Marketing: By analyzing customer behavior, demographics, and preferences, our solution provides valuable insights to marketing managers. This enables them to develop personalized marketing campaigns, tailor promotions, and enhance customer engagement, resulting in increased customer satisfaction, retention, and ultimately, higher sales.

•Competitive Advantage: Leveraging the power of data analysis, our solution provides the Superstore with a competitive advantage in the market.

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 techniques, methodologies, and frameworks were applied:

•**Exploratory Data Analysis (EDA):** EDA techniques were employed to gain initial insightsinto 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.

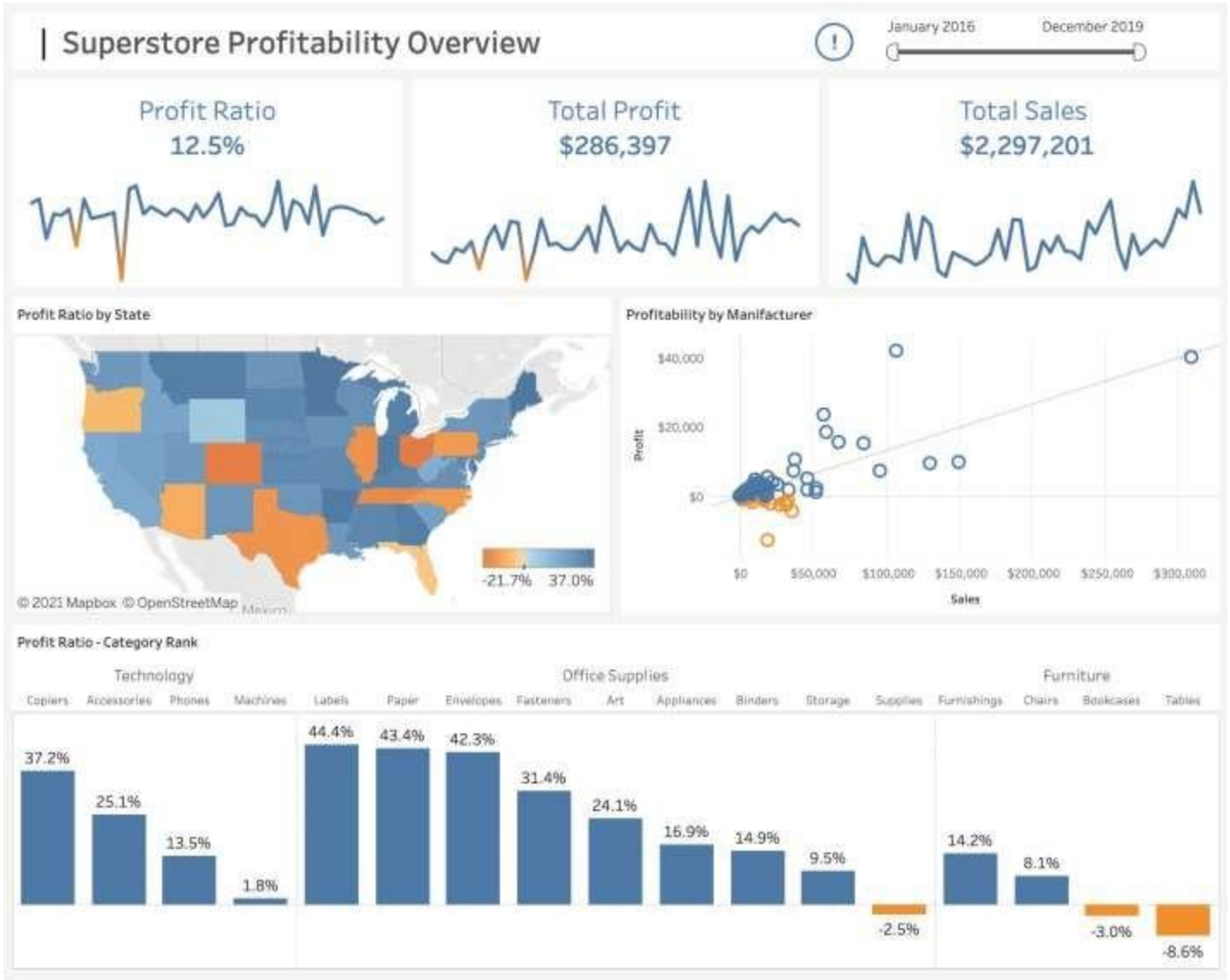
•Statistical Analysis: Utilized to uncover correlations, trends, and patterns within the Superstore dataset. These techniques helped in understanding the impact of variousfactors on sales, customer behavior, and operational efficiency.

•**Customer Segmentation:** applied o 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
-21.7%
37.0%
© 2021 Mapbox © OpenStreetMap
Profitability by Manifacturer
$40,000
$20,000
$0
Profit
$0
$50,000
$100,000
$150,000
$200,000
$250,000
$300,000
Sales
Profit Ratio - Category Rank
Technology
Office Supplies
Furniture
Copiers
Accessories
Phones
Machines
Labels
Paper
Envelopes
Fasteners
Art
Appliances
Binders
Storage
Supplies
Furnishings
Chairs
Bookcases
Tables
37.2%
25.1%
13.5%
1.8%
44.4%
43.4%
42.3%
31.4%
24.1%
16.9%
14.9%
9.5%
-2.5%
14.2%
8.1%
-3.0%
-8.6%

LINKS:

Github Link:

https://github.com/ImAnurag22/IBM_Internship_Project_for_DataAnalysis

Research Paper:

Here are some references for sales analysis on Superstore dataset:

•Chakraborty, M. (2020). Sales Analysis of Superstore using Power BI. Kaggle. https://www.kaggle.com/moumoyesh/sales-analysis-of-superstore-using-power-bi

•Microsoft. (n.d.). Analyse and visualize Superstore data in Power BI. https://powerbi.microsoft.com/en-us/tutorials/analyze-and-visualize-superstore-data/

•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

•Pranav, B. (2021). Sales Analysis of Superstore Data using Power BI. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/04/sales-analysis-of-superstore-data-using-power-bi/

•Microsoft. (n.d.). Analyse and visualize Superstore data in Power BI. https://powerbi.microsoft.com/en-us/tutorials/analyze-and-visualize-superstore-data/

•Wong, J. (2021). Sales Analysis of Superstore Dataset Using Power BI. LinkedIn. https://www.linkedin.com/pulse/sales-analysis-superstore-dataset-using-power-bi-jeremy-wong/

•Rajasekaran, D., & Mohan, K. V. (2018). A review of sales forecasting models for retail industry. International Journal of Business Forecasting and Marketing Intelligence, 4(1), 1-16.

•Suri, S., & Taneja, S. (2018). A comparative analysis of machine learning algorithms for sales forecasting in retail industry. International Journal of Engineering and Technology, 7(4.19), 66-70.

DATA SET:

1. Data set URL:
   https://www.kaggle.com/datasets/vivek468/superstore-dataset-final
1. About the dataset:
   The dataset provides information about the sales and profit from a supermarket.
1. Dataset details:

Size	563kb
Number of columns	21
Number of Rows	9994
Original file format	Csv

1. Colum e', 'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State', 'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category', 'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit']

Step-1: Importing the dataset

```
# Importing libraries
import pandas as pd
import numpy as np

df = pd.read_csv("/content/drive/MyDrive/IBM_Project/Superstoredataset.csv",
encoding='cp1252')
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()

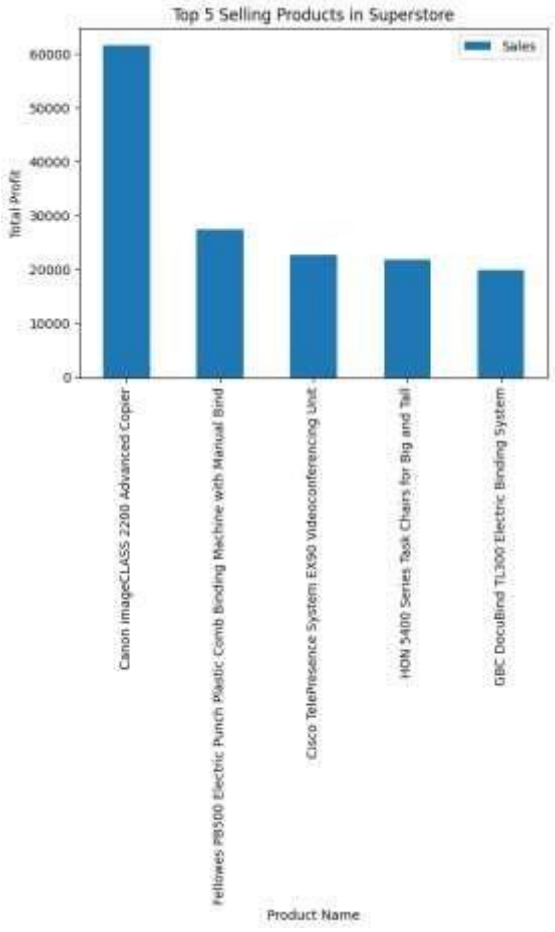
	Row ID	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	55190.379428	229.858001	3.789574	0.156203	28.656896
std	2885.163629	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1.000000	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	99301.000000	22638.480000	14.000000	0.800000	8399.976000

Step-2: Exploratory Data Analysis – EDA:

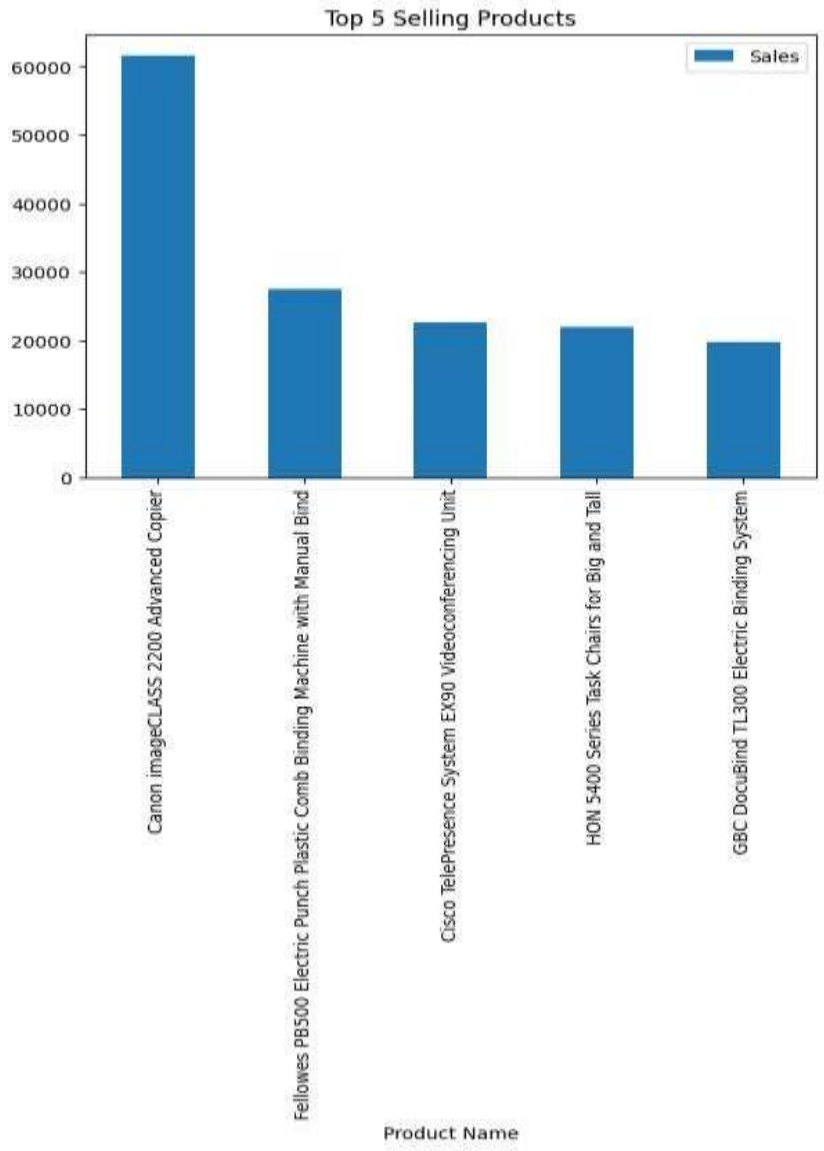
What are the top selling products in the superstore?

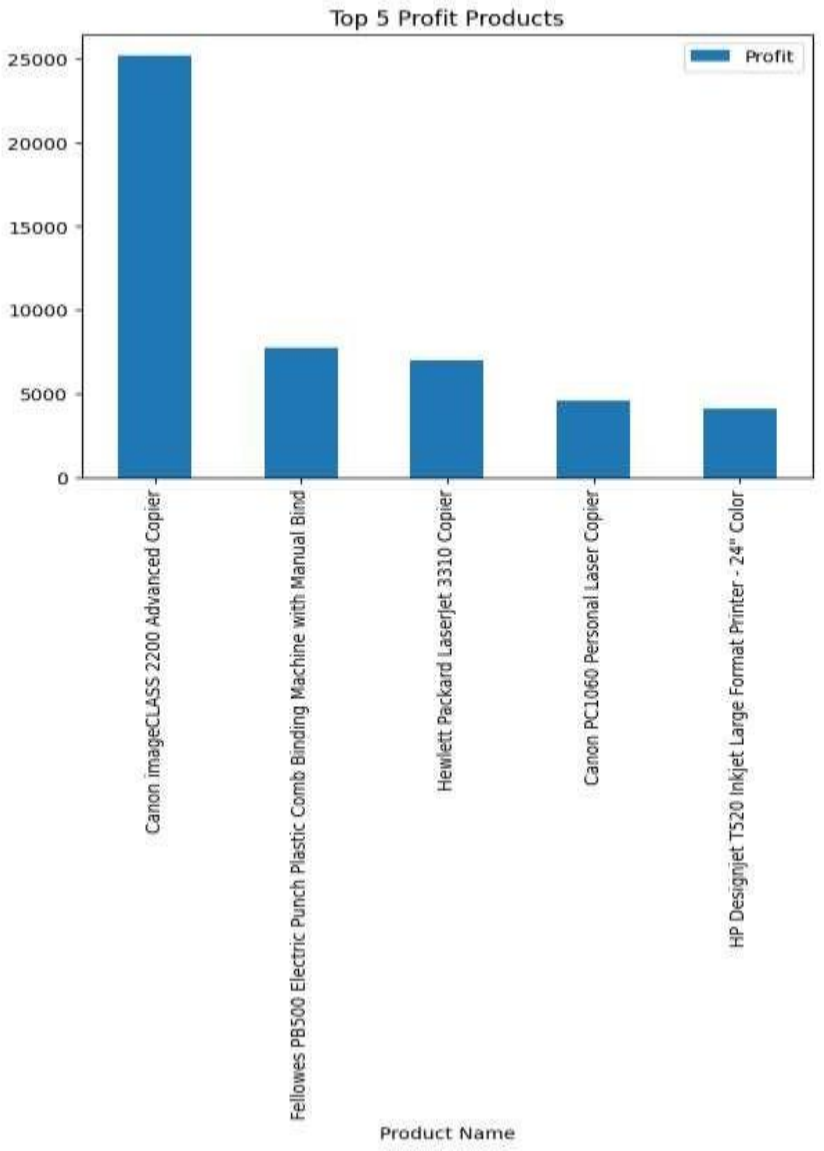
```
# Group the data by Product Name and sum up the sales by product
product_group = df.groupby(["Product Name"]).sum()["Sales"]
product_group.head()
```

```
top_5_selling_products.plot(kind="bar")

# Add a title to the plot
plt.title("Top 5 Selling Products in Superstore")

# Add labels to the x and y axes
plt.xlabel("Product Name")
plt.ylabel("Total Profit")

# Show the plot
plt.show()
```



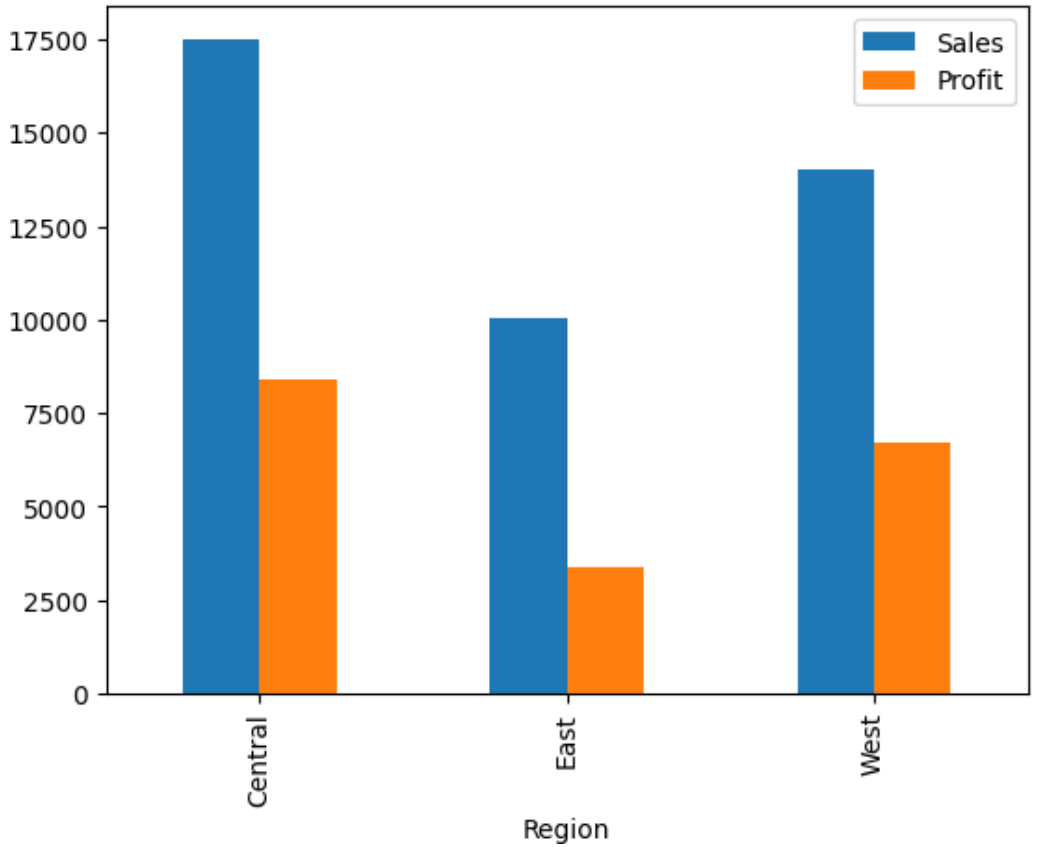
Are the top-selling products the most profitable?



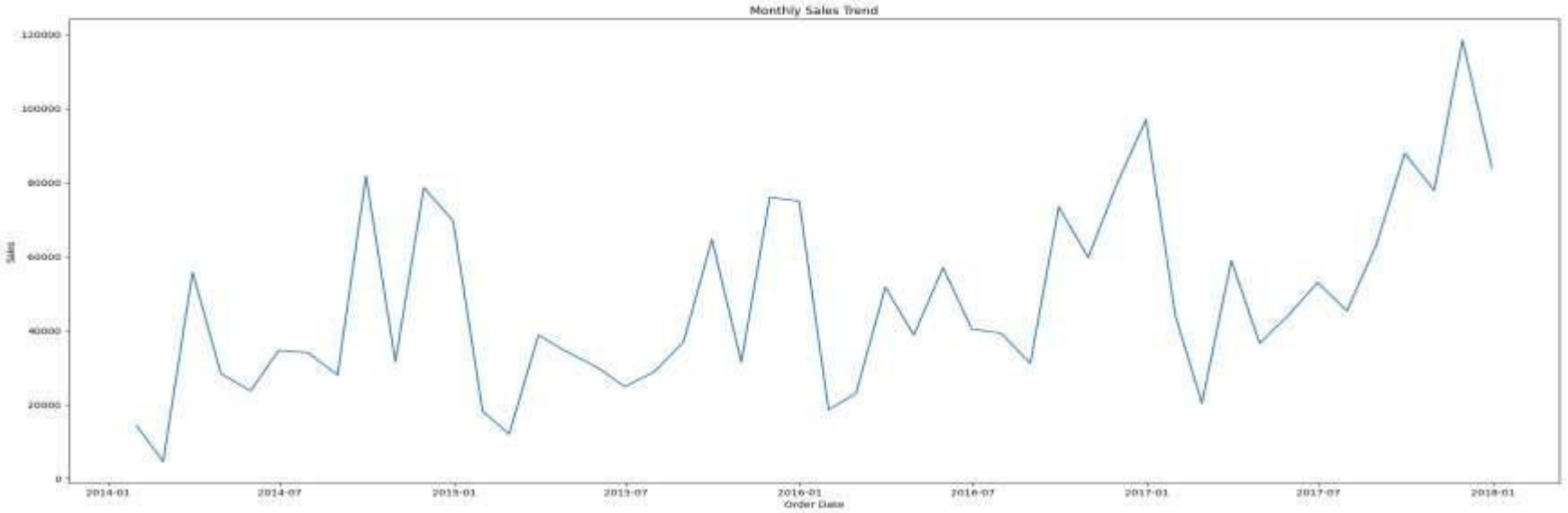


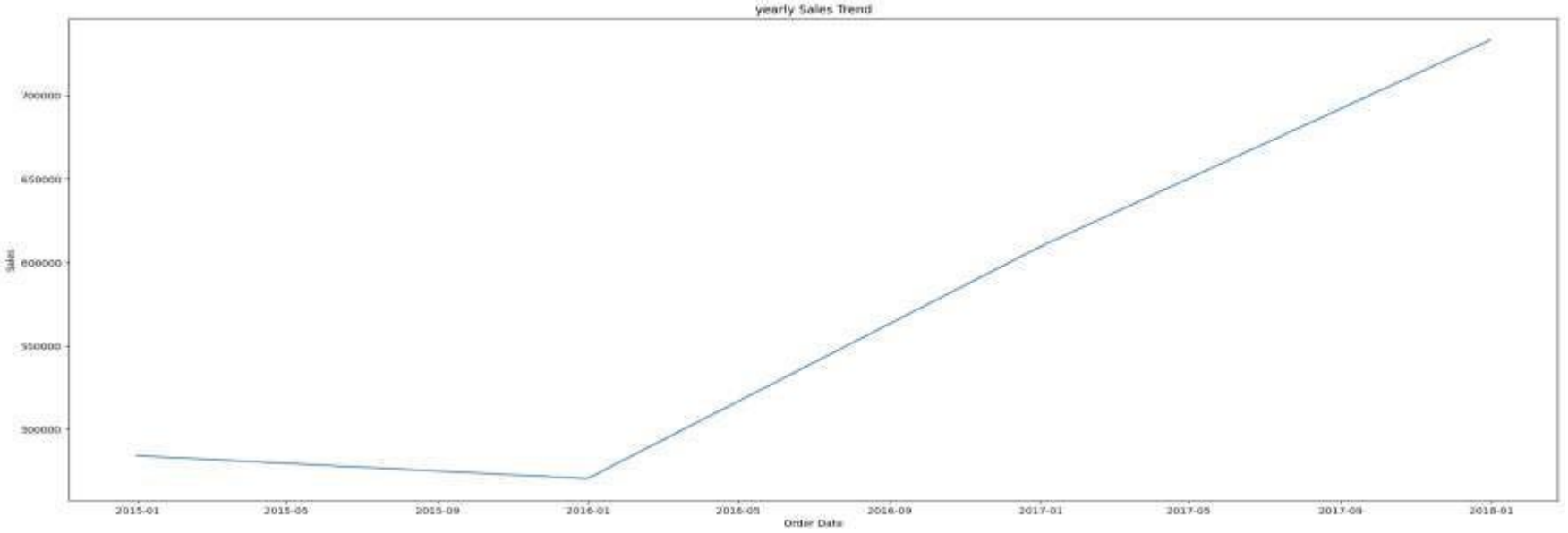
What is the total Sales and Profit by region?

```
# Filter the data to only include the Canon imageCLASS 2200 Advanced Copier
product = df[df["Product Name"] == "Canon imageCLASS 2200 Advanced Copier"]

# Group the data by Region
region_group = product.groupby(["Region"]).mean()[["Sales", "Profit"]]

# Ploting
region_group.plot(kind="bar")

plt.show()
```

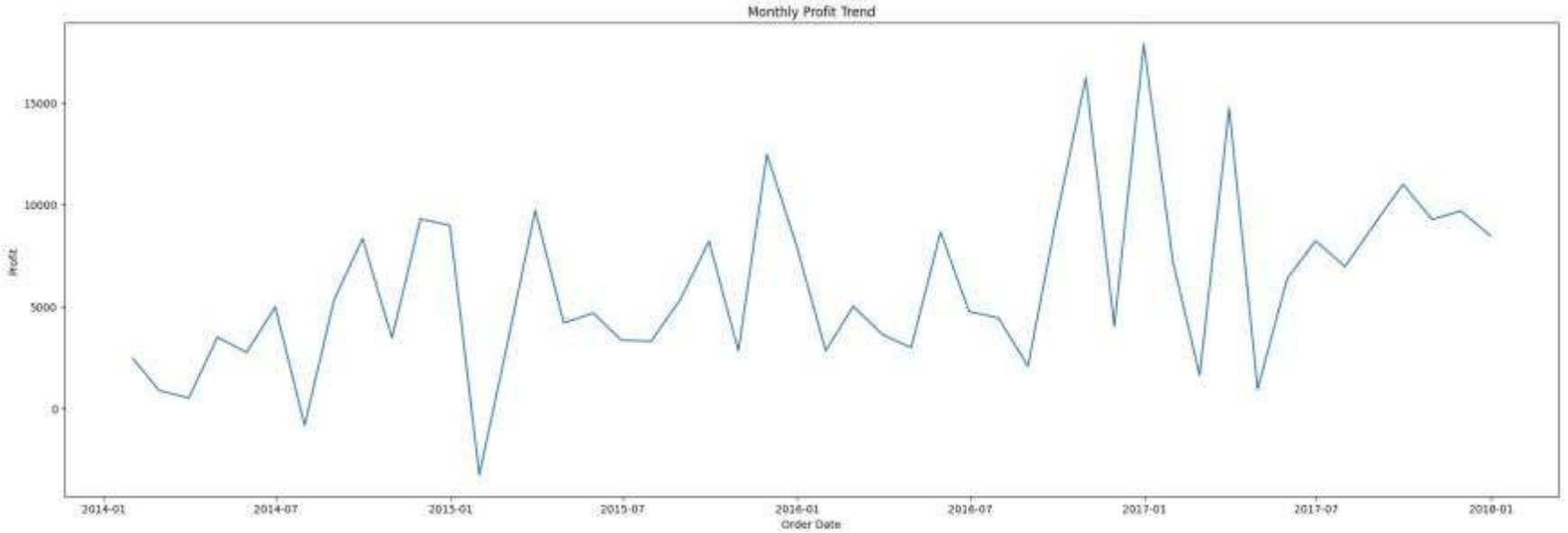


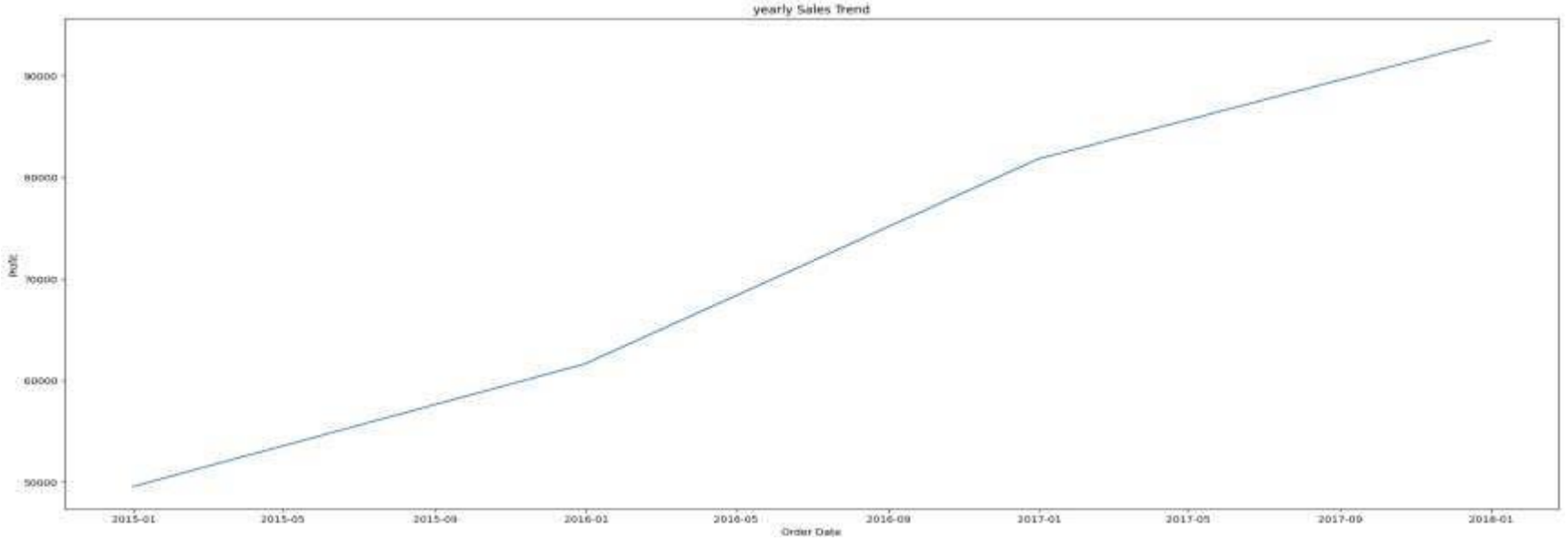
What is the sales trend over time (monthly, yearly)?





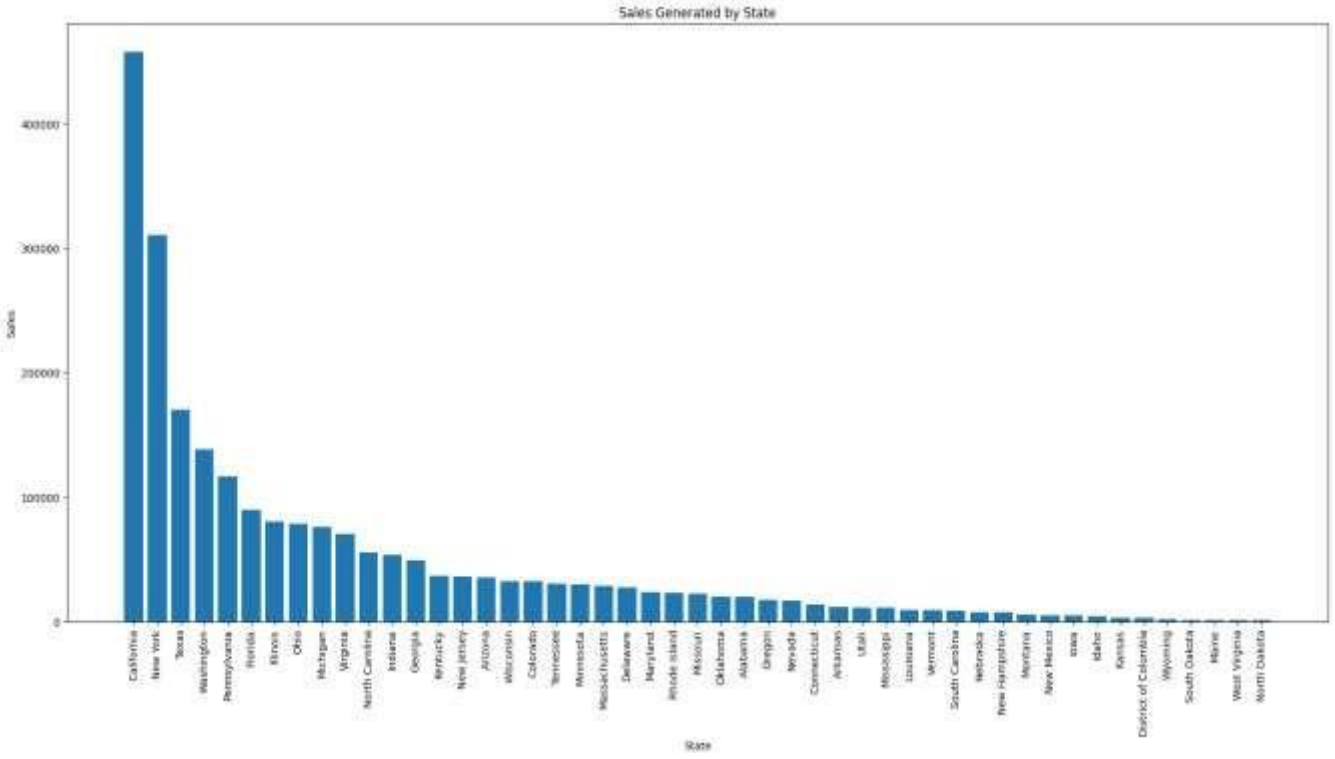
Profit over time:





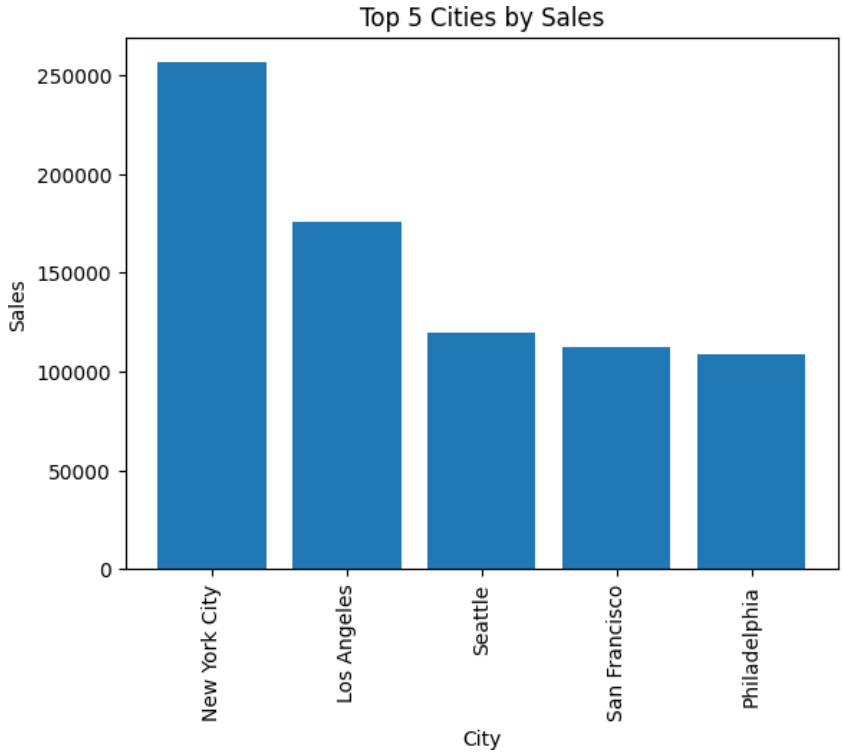
Sales Generated by Statewise:

```
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
```



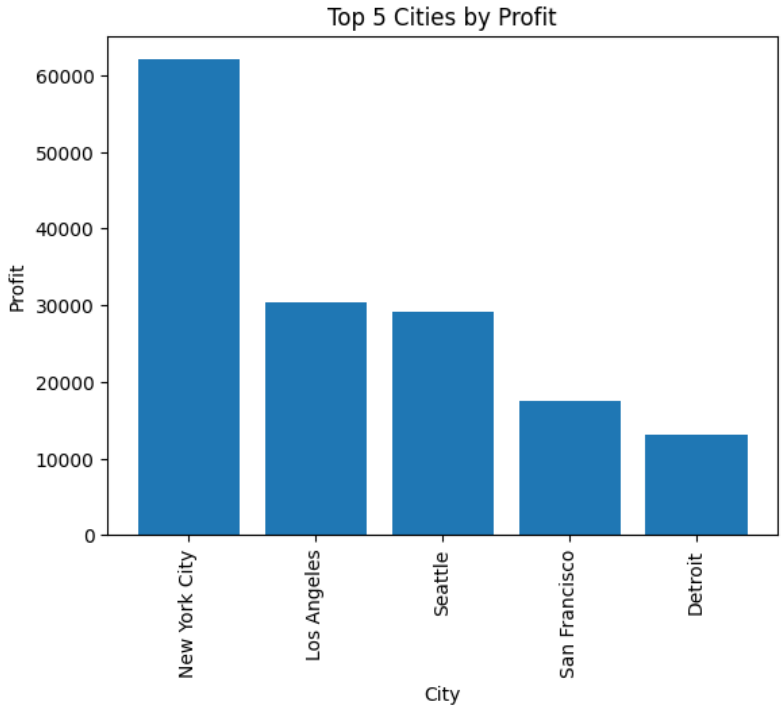
Select top 5 cities by sales and Sort the data by Sales in descending order:

```
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
```



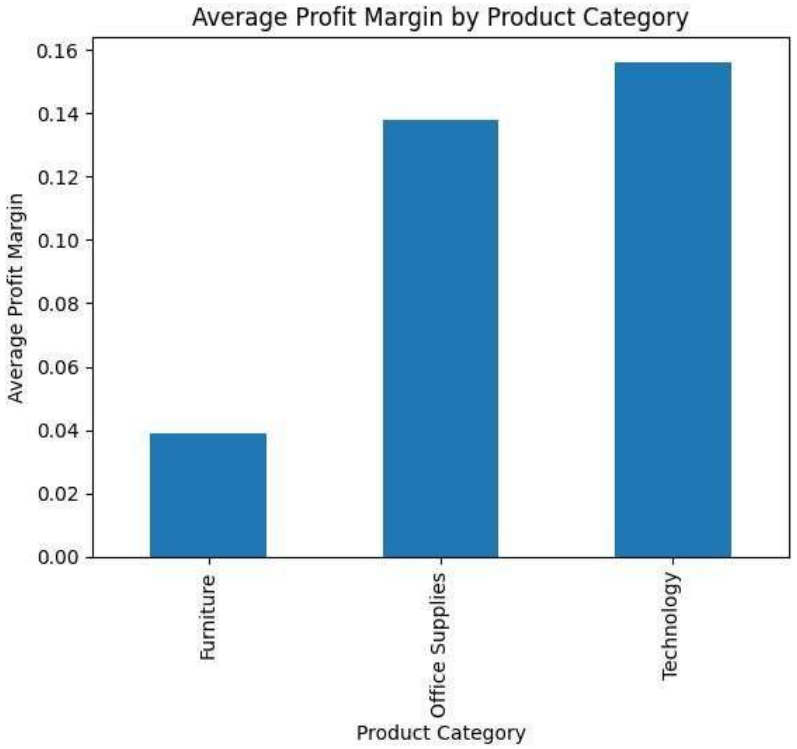
Select top 5 cities by profit and Sort the data by profit in descending order:

```
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
```



The best sales:

```
# Group the data by product category and calculate the average profit for each category
avg_profit_margin_by_category = df.groupby('Category')['Profit'].sum()
print(avg_profit_margin_by_category)
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
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:

The analysis of the Superstore dataset has provided valuable insights into sales trends, customer behavior, and operational efficiency. 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:

Moving forward, it is recommended that the Superstore continues to monitor sales performance, customer behavior, and operational metrics. This will allow for ongoing adjustments and improvements based on changing market dynamics and evolving customer preferences.

Overall, the "Analysis of Superstore dataset" project demonstrates the power of data analytics in uncovering insights that drive strategic decision-making, operational efficiency, and ultimately, the success of the Superstore in a competitive retail market.