DATA ANALYTICS PROJECT

Name: Srijan Sasmal
Email: srijansasmal123@gmail.com
Organization: DGT
College name: St. Thomas' College of Engineering and Technology
State: West Bengal
Domain: Data Analytics
S/E date: 12.06.2023 - 24.07.2023

PROJECT TITLE

Analysis of Superstore Dataset

The goal of this project is to analyse the Superstore dataset to gain insights into sales trends, customer behaviour, 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 Pre-processing: Collect , pre-process the Superstore dataset.
- Sales Analysis: Analyse sales metrics, trends, and factors influencing sales fluctuations.
- Customer Behaviour Analysis: Study customer demographics, preferences, and segmentation
- Exploratory Data Analysis : Perform exploratory analysis, including data distribution, outliers, visualizations.
- Operational Efficiency Analysis: Evaluate operational efficiency, identify bottlenecks, and optimize resource allocation.
- Conclusion and Next Steps: Summarize findings, advanced analysis, predictive modelling, integration of external data sources.

Agenda

PROJECT OVERVIEW

3

DATA SET

11

STATISTICAL INFORMATION

13

EDA

14

CONCLUTION

22

PROJECT OVERVIEW

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: to gain insights into sales trends, customer behaviour, and operational efficiency in order to optimize store performance and make data-driven recommendations for improvement.

Scope: data cleaning, exploratory data analysis, sales analysis, customer behaviour analysis.

Objective:

- Identify sales trends, such as seasonal patterns and fluctuations, to optimize inventory management
- Understand customer behaviour by analysing demographics, preferences, and purchase patterns
- Improve operational efficiency by identifying bottlenecks, streamlining processes, optimizing resource allocation
- Provide data-driven recommendations to optimize store performance.

WHO ARE THE END USERS

Target Audience or End Users:

- Store Managers: They require insights on sales performance, customer behaviour, and operational efficiency
- Marketing Managers: They need information on customer demographics, preferences, and buying patterns

Characteristics and Needs:

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

Benefits from the Solution:

- optimized inventory management, improved sales forecasting, and streamlined operations, leading to increased profitability
- 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 behaviour, and operational efficiency. This analysis will be carried out using various statistical and data mining techniques, as well as advanced visualization tools.

Value Proposition:

- Data-Driven Decision Making: . 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.
- Customer Insights and Personalized Marketing: 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

MODELLING

- ❑ Exploratory Data Analysis (EDA):

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:

These techniques helped in understanding the impact of various factors on sales, customer behaviour, and operational efficiency.

- ❑ Customer Segmentation:

applied o categorize customers based on their attributes and buying behaviour. 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.

Customize the project and make it my own

- Advanced Visualization with Matplotlib and Seaborn:

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.

- Interactive Dashboards:

dashboards allow stakeholders to dynamically explore and interact with the analysed data, enabling them to drill down into specific details, apply filters, and visualize different dimensions, dashboards enhances engagement, facilitates deeper insights, and empowers users

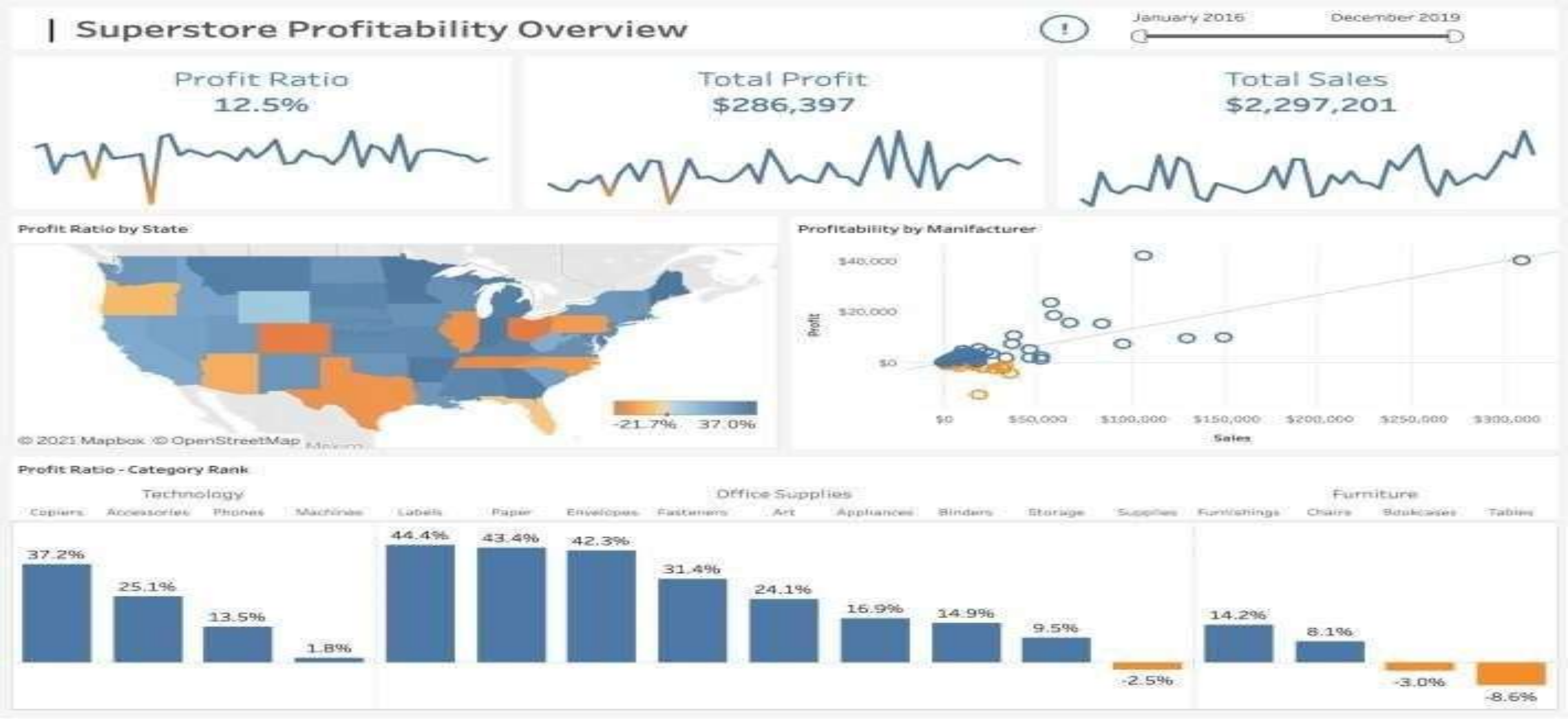
- Descriptive Analytics:

to summarize and present key information about sales trends, customer behaviour, operational performance within the Superstore dataset. This includes calculating summary statistics, generating frequency distributions, identifying important patterns or trends.

- Forecasting and Trend Analysis:

Apply forecasting methods and trend analysis to predict future sales trends and demand patterns.

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
© 2021 Mapbox © OpenStreetMap
-21.7%
37.0%
Profitability by Manufacturer
$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

https://colab.research.google.com/drive/13EgOOWNSO9XG8gjqxQ8lKSqBt5IuXwsv#scrollTo=ICGRezeDhkdk

Research Paper:

- 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/analyse-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/analyse-and-visualize-superstore-data/

DATA SET

DATA SET DETAILS

- Data set URL: 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:

1. Size: 563kb
2. Number of columns: 21
3. Number of Rows: 9994
4. Original file format: Csv

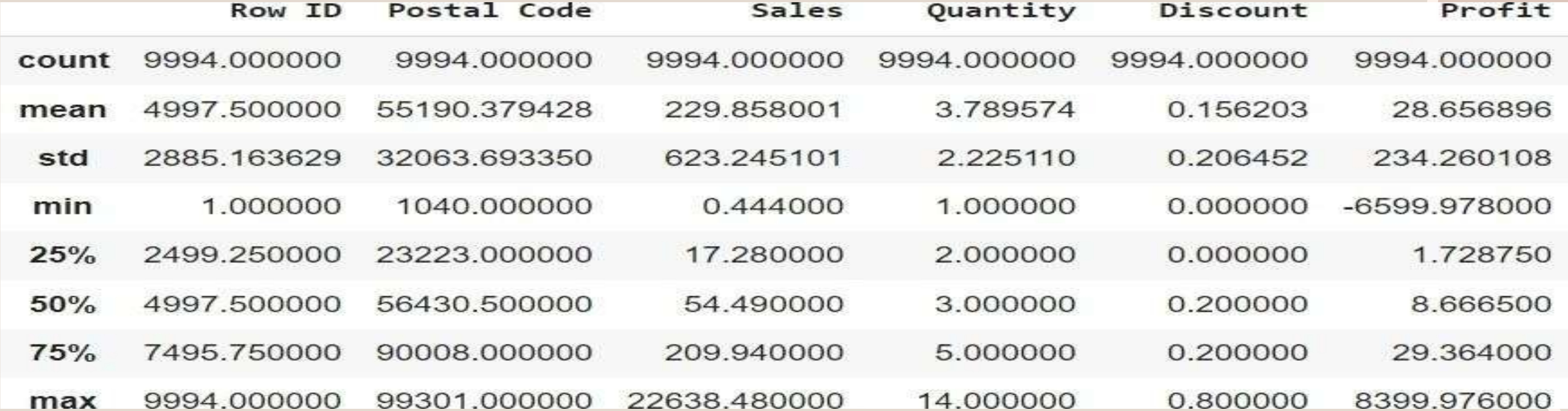
- Column:

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

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
```

EDA

Exploratory Data Analysis

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:

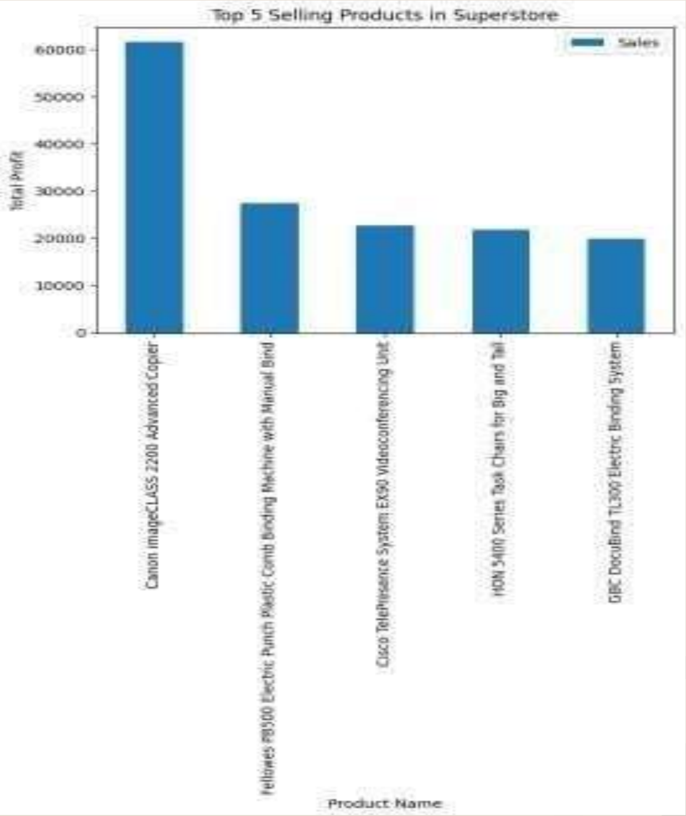
```
df.isna().sum()
```

Read the Duplicate value:

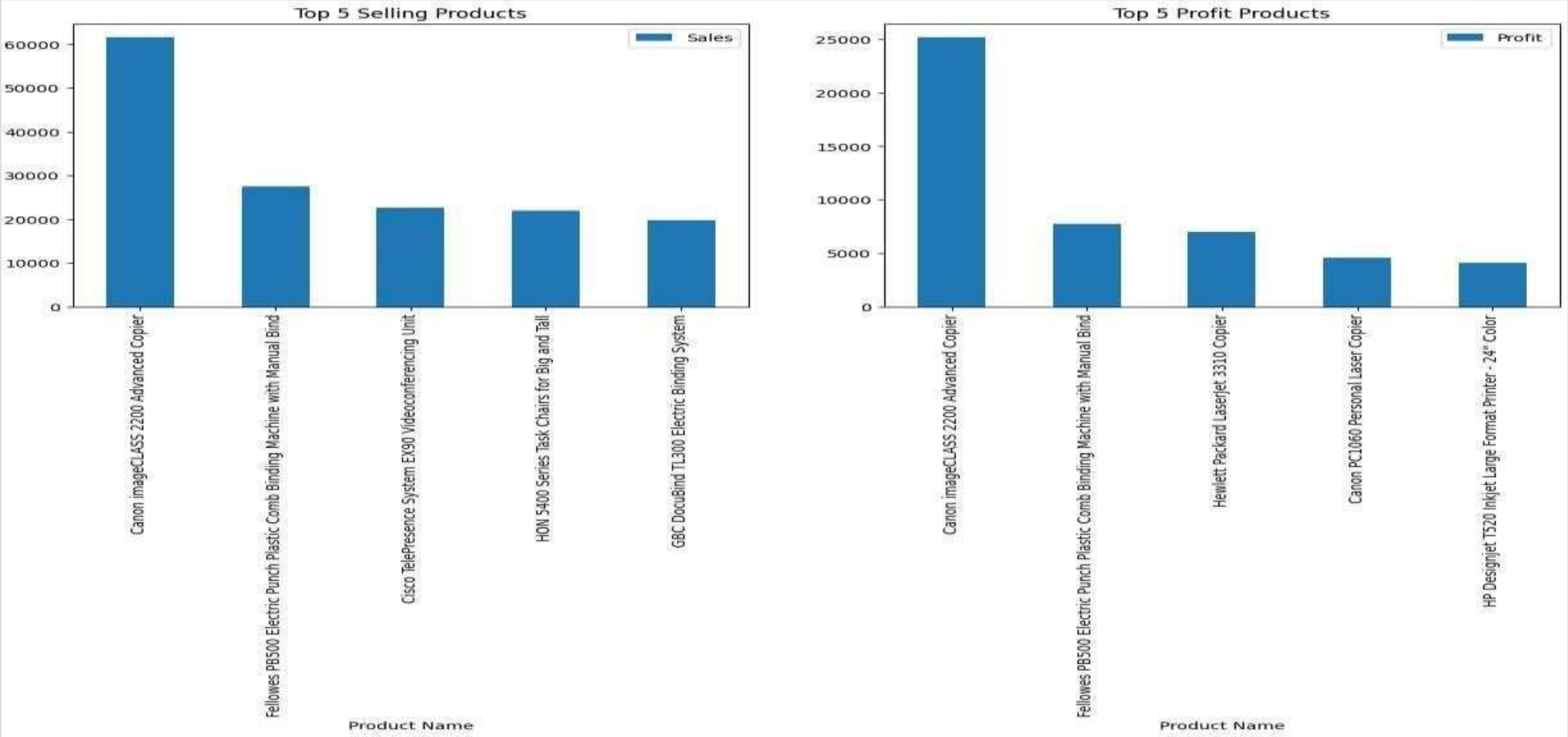
```
df.duplicated().sum()
```

Step-2: Exploratory Data Analysis - EDA

```
# 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()
```

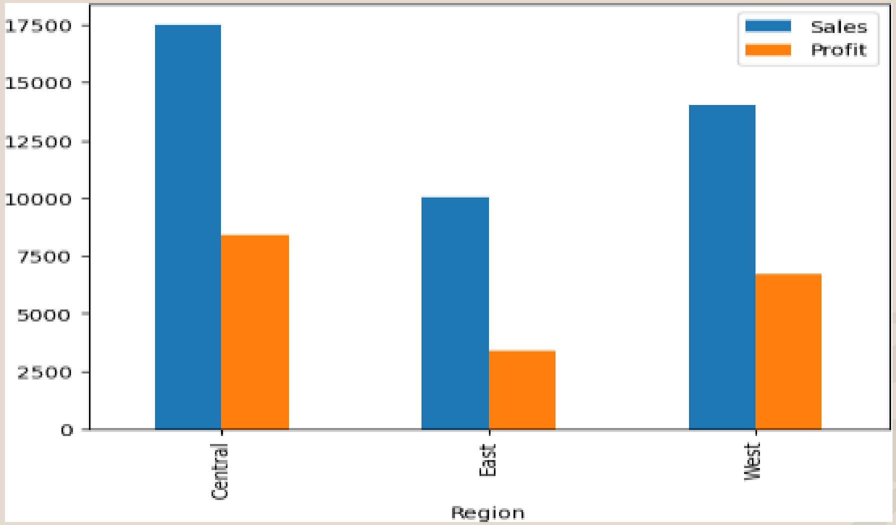


Are the top-selling products the most profitable?

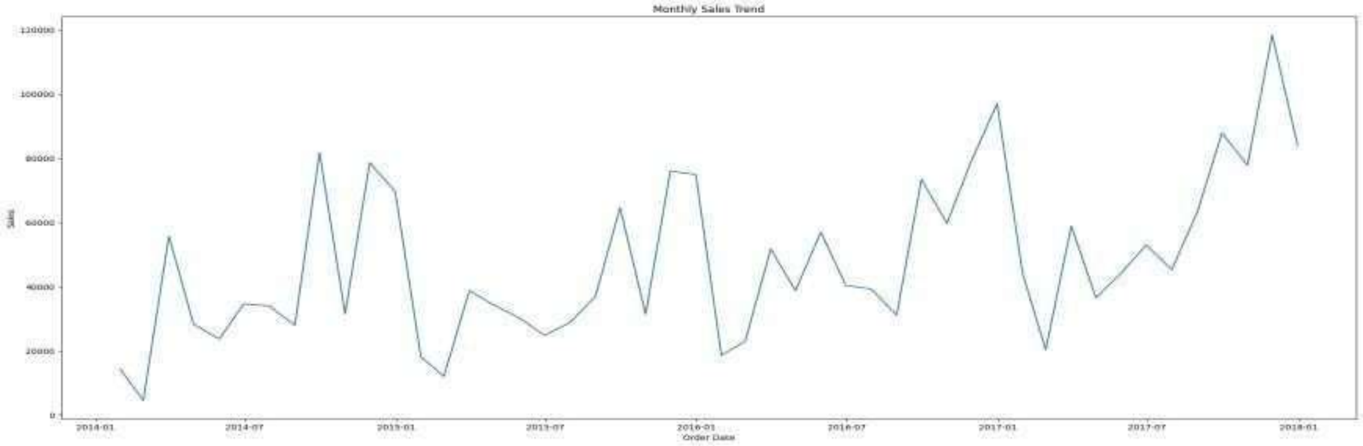


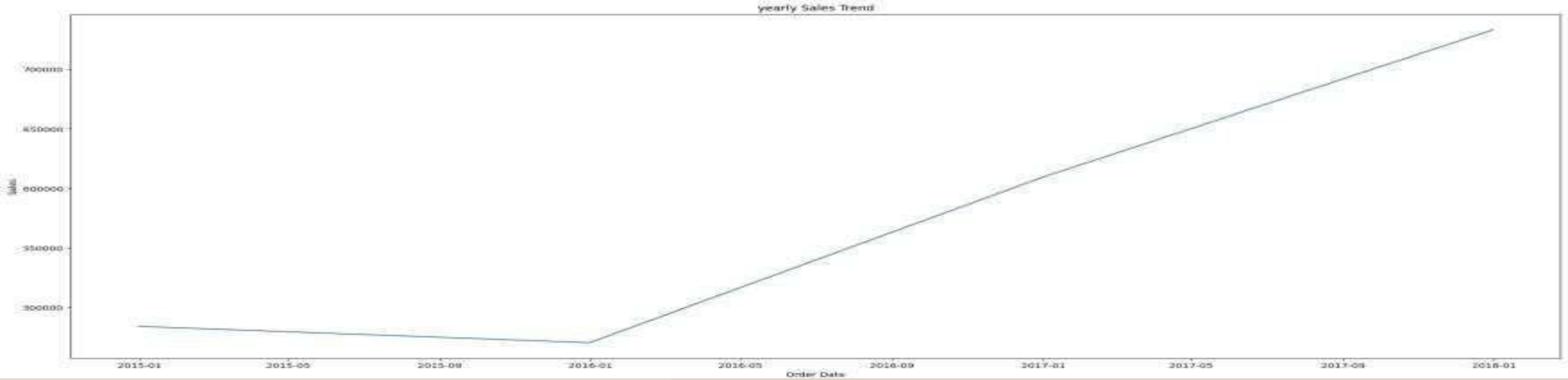
What is the total Sales and Profit by region?

```
# 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()
```

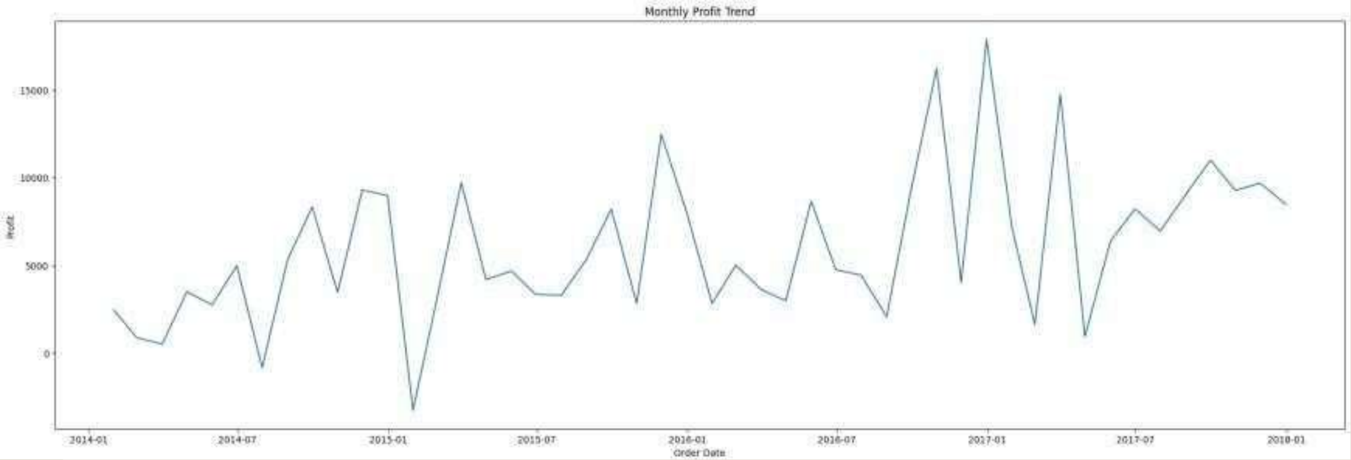


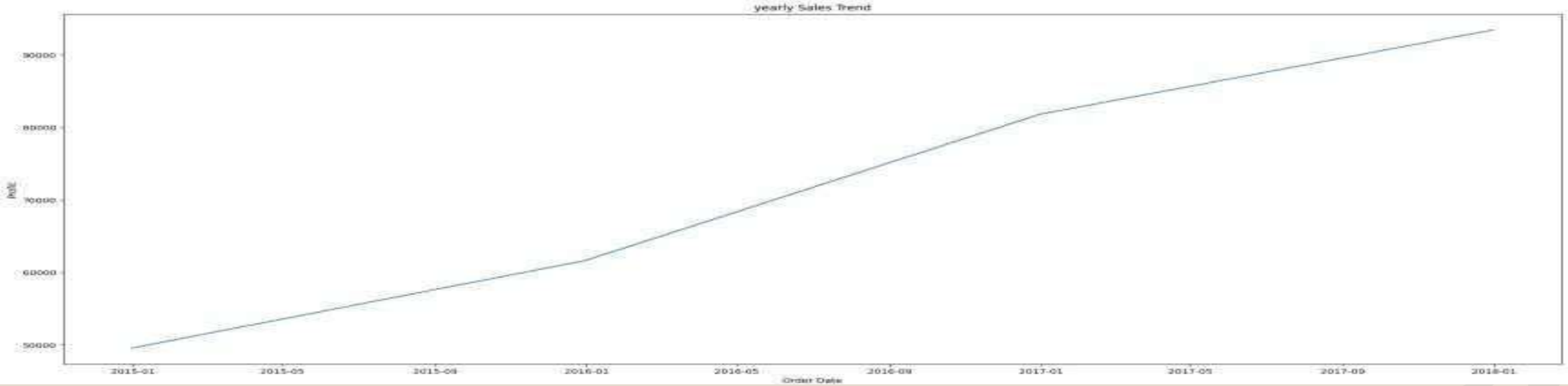
What is the sales trend over time (monthly, yearly)?





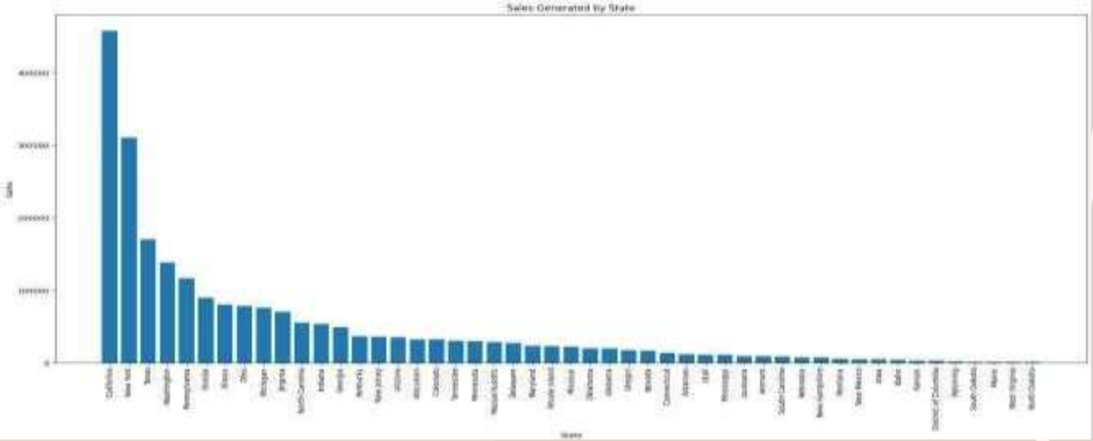
Profit over time





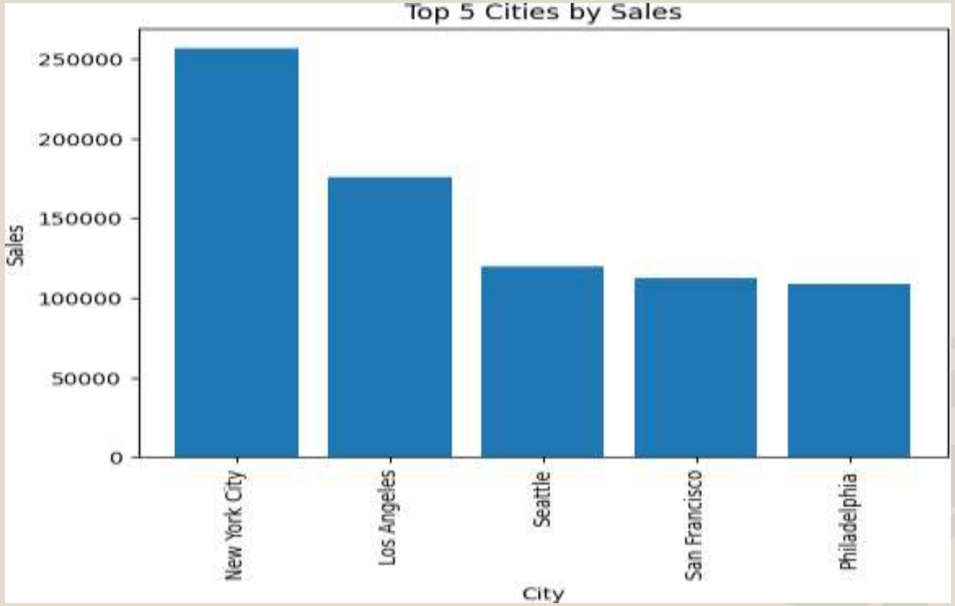
Sales Generated by Statewise

```
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
```



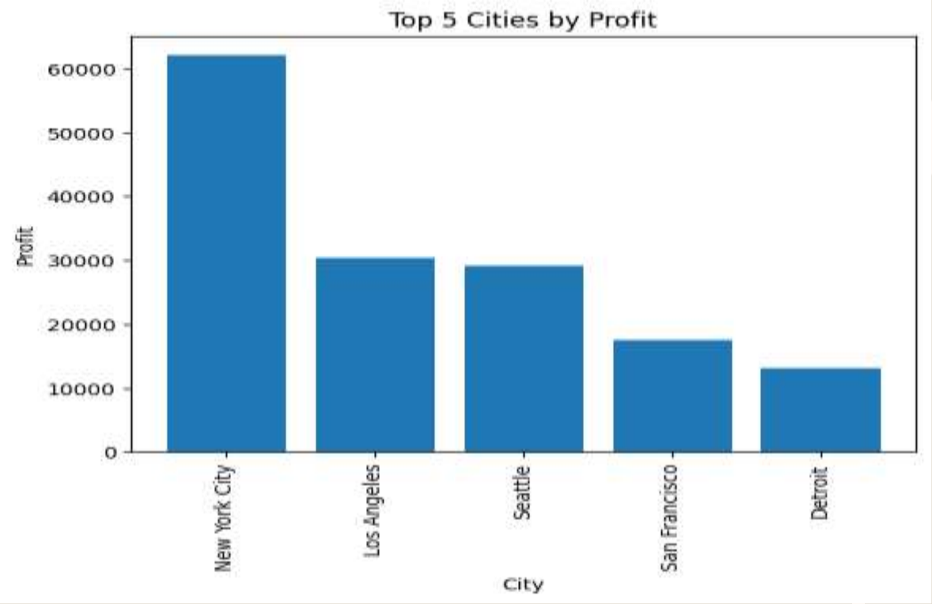
Select top 5 cities by sales and Sort the data by Sales in descending order

```
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
```



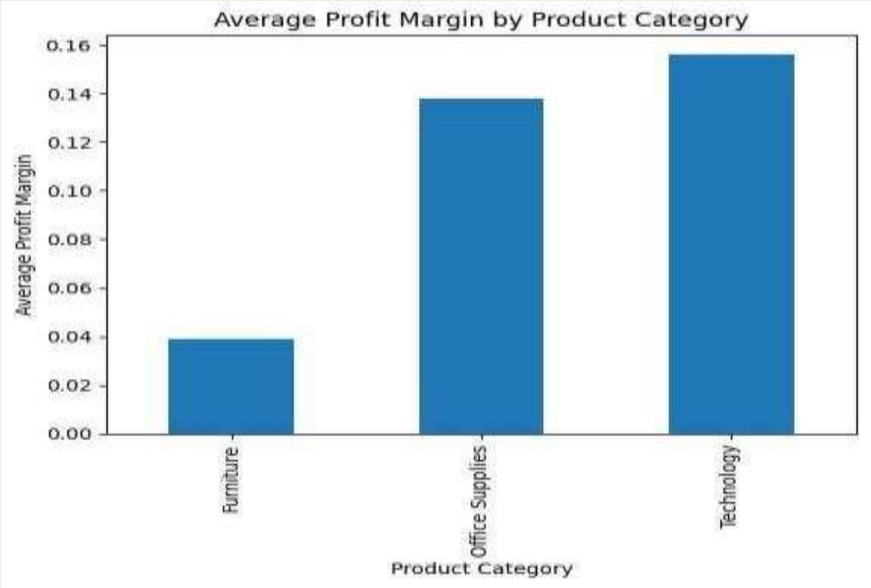
Select top 5 cities by profit and Sort the data by profit in descending order

```
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 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()
```



CONCLUTION

- SALES TRENDS
- CUSTOMER SEGMENTATION
- PREDICTIVE INSIGHTS
- ENHANCED PROFITABILITY
- IMPROVED DECISION MAKING
- CUSTOMER SATISFACTION AND RETENTION

Thank you

SRIJAN SASMAL

srijansasmal123@gmail.com