I

STUDENT DETAILS

NAME: DIVYA GUPTA

EMAIL ID: DIVYAG.EC.20@NITJ.AC.IN

COLLEGE: D R B R AMBEDKAR NIT JALANDHAR

COLLEGE STATE: PUNJAB

DOMAIN: DATA ANALYTICS

S/E DATE: 12.06.2023-24.07.2023

SUPERSTORE DATABASE ANALYSIS

- **<u>Introduction:</u>** This project is to analyze the superstore dataset to gain insights into sales trends, customer behavior and operational efficiency. The various aspects taken care of are: *sales, customer demographics, product categories and geographical regions*. The aim of this project is to identify opportunities for improvement and make data-driven recommendations to optimize store performance.

3 AGENDA

S.no	Topic Name	Slide No
1.	Project Overview	4
2.	End Users	5
3.	My Solution and its Value Proportion	6
4.	Customizing the Project	7
5.	Modelling Techniques	8
6.	Results	9
7.	Links	12

4 PROJECT OVERVIEW

- **Purpose:** The purpose of this project is to collect and clean the raw data of the superstore and optimize store performance by effectively analyzing the information regarding sales transactions, customers, products and geographical locations.

- **Scope:** The project deals with *data collection and cleaning*, *exploratory* analysis, *customer behavior* analysis and *operational efficiency* analysis.

- **Objectives:** 1. To understand sales trends like seasonal patterns and fluctuations.

  2. To analyze customer preferences and purchase patterns.

  3. To optimize resource allocation for enhanced profitability.

THE 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 increase customer engagement.

- **Characteristics and Needs:** Comprehensive data analysis, visualizations and actionable recommendations required to identify areas for improvement, profitability and streamline operations.

- **Benefits from the Solution:** Improved sales forecasting, profitability, customer satisfaction and brand loyalty.

| SOLUTION	VALUE PROPOSITION
• Various statistical and data mining techniques were carried out along with advanced visualization tools to help improve the performance of superstore. • Comprehensive analysis with python libraries like numpy, pandas, matplotlib, seaborn was done to help achieve the target.	• **Data-driven decision making:** Store and marketing managers can now make informed decisions due to the analysis carried out on the dataset. • **Enhanced Profitability:** By optimizing pricing strategies, identifying high-demand products and streamlining operations, the store can maximize its revenue and profits. • **Increased Customer Engagement:** By taking care of customer behavior, demographics and preference patterns as depicted by the analysis, the store can tailor personalized marketing campaigns which in turn will lead to customer satisfaction and retention. • **Competitive Advantage:** Leveraging the power of data analytics, the solution proposed definitely provides a competitive advantage in the market to the superstore.

CUSTOMIZATION OF THE PROJECT

- **<u>Advanced Visualization with Matplotlib and Seaborn:</u>** By leveraging the capabilities of these powerful python libraries, the solution proposed presents data in a visually engaging manner in the form of insightful charts, graphs and plots which ultimately helps in enhancing the understanding of complex patterns and relationships within the superstore dataset.

- **<u>Interactive Dashboards:</u>** Interactive dashboards are incorporated to provide an exceptional user experience. 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.

- **<u>Descriptive Analytics:</u>** This includes calculating summary statistics, generating frequency distributions, and identifying important patterns or trends.

- **<u>Forecasting and Trend Analysis:</u>** This includes applying forecasting methods and trend analysis to predict future sales trends and demand patterns.

DATA MODELLING

- Advanced data visualization tools including python libraries used to create visually appealing and informative charts, graphs and plots. These visualizations facilitated the effective communication of analysis results and provided a clear representation of key findings.

- Pandas library effectively used to collect and clean data and perform exploratory data analysis effectively.

- 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.

- Statistical Analysis was carried out to uncover correlations, trends, and patterns within the Superstore dataset.
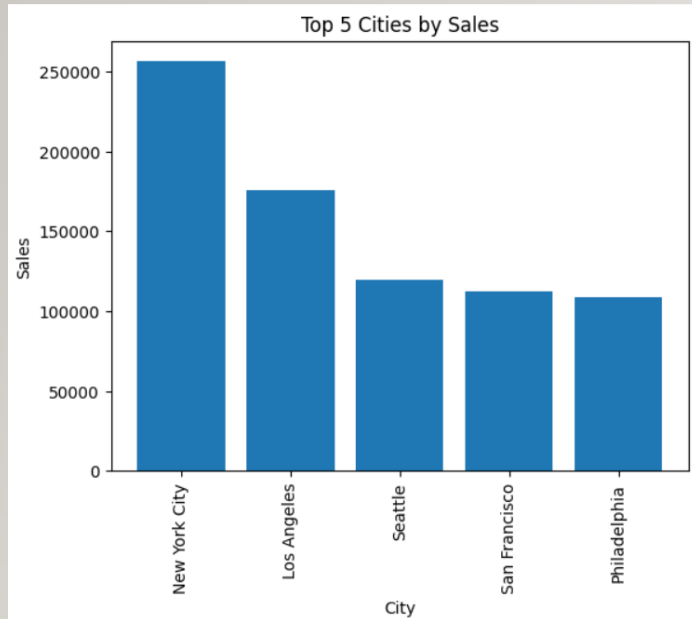
RESULTS

MONTHLY SALES TREND:
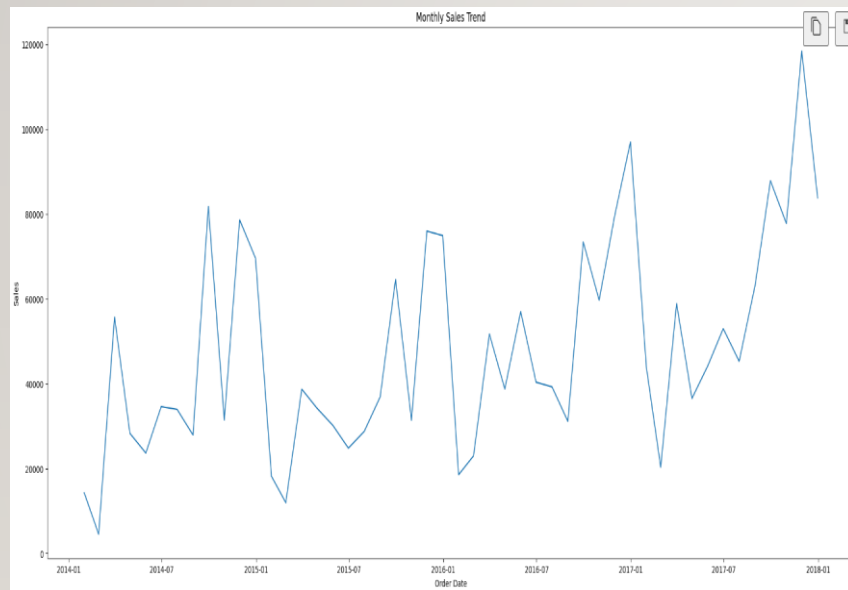
MONTHLY PROFIT TREND:

10
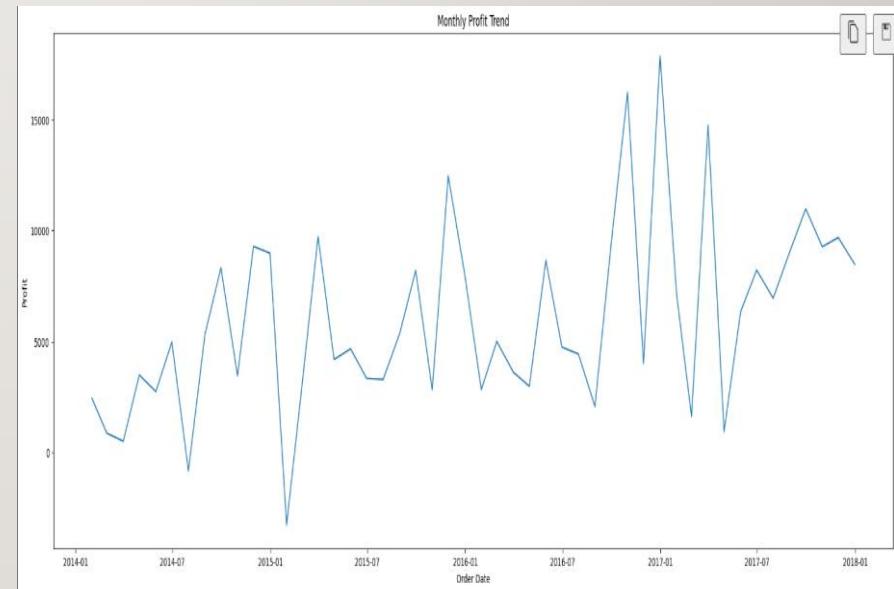
Top 5 Cities by Sales

```python
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
```

```python
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
```
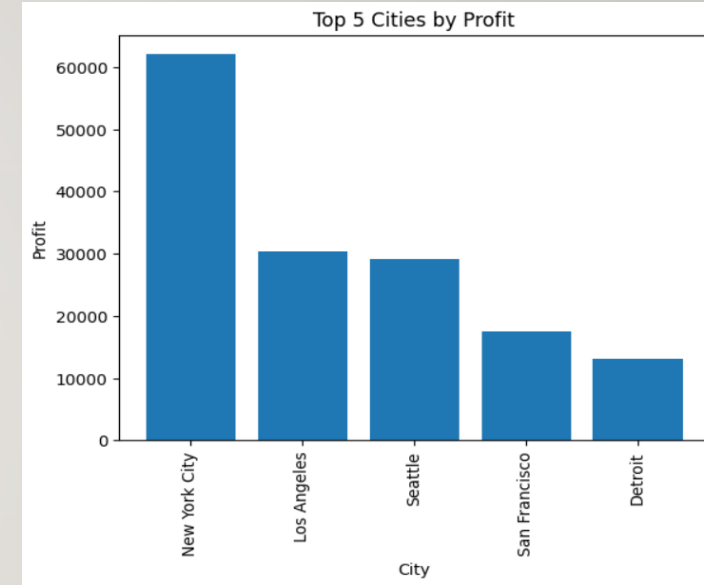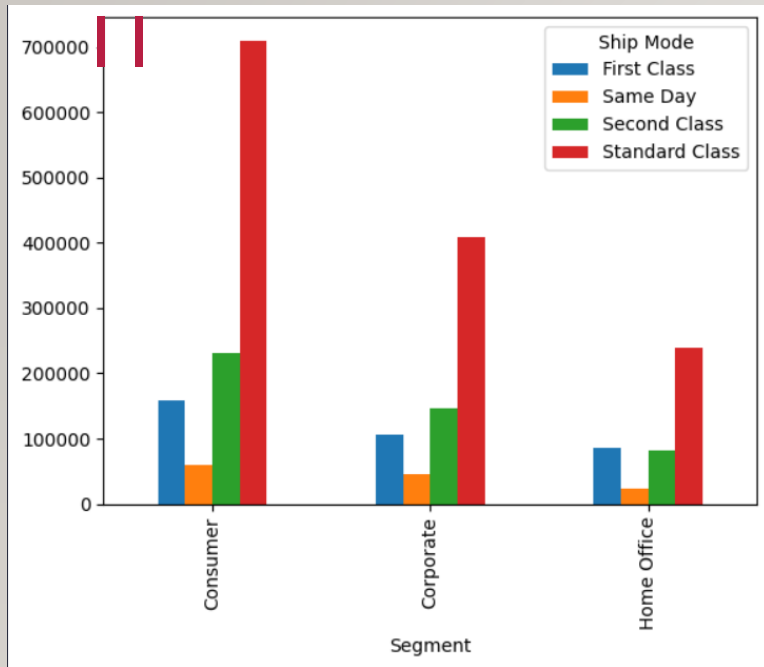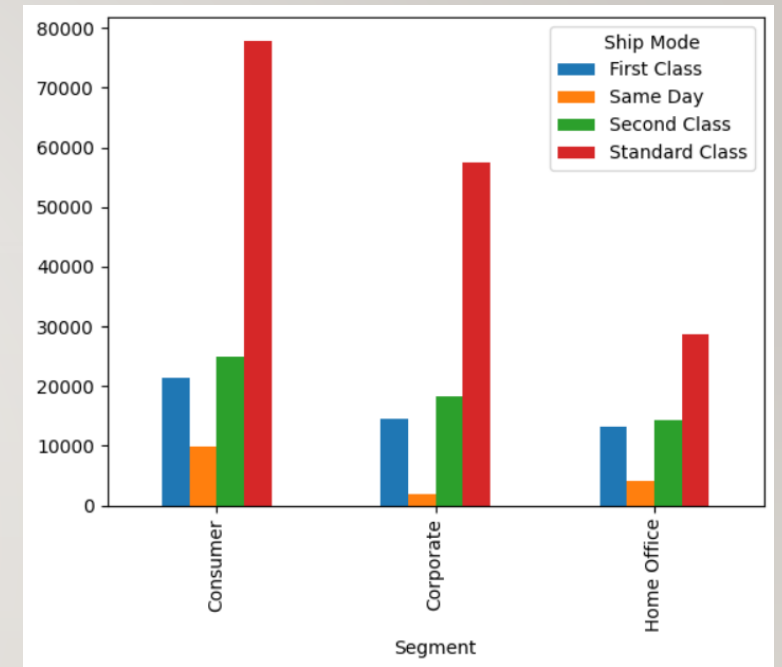

Top 5 Cities by Profit

SALES

PROFIT


```
pivot_table = df.pivot_table(index='Segment', columns='Ship Mode', values='Sales', aggfunc='sum')
```

```
#profit
pivot_table = df.pivot_table(index='Segment', columns='Ship Mode', values='Profit', aggfunc='sum')

pivot_table.plot(kind='bar', stacked=False)

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

LINKS

- [Github Repository](#)

- RESEARCH PAPERS:

  - [Chakraborty, M. (2020). Sales Analysis of Superstore using Power BI. Kaggle](#)

  - [Vignesh, S. (2021). Sales Analysis of Superstore dataset using Power BI. Towards Data Science](#)

  - [Wong, J. (2021). Sales Analysis of Superstore Dataset Using Power BI. LinkedIn](#)