**Industrial Internship Report**

**on**

**SALES PREDICTION USING EDA & POWER BI**

**Submitted in partial fulfilment of the requirements**

**for the award of the degree of**

**MASTER’S OF COMPUTER APPLICATION**

**SUBMITTED TO: - SUBMITTED BY: -**

Ms. Chhavi Sharma Nirmal Kuttan

Assistant Professor 221347040

FEAT - CSE MCA 4th Sem



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SGT UNIVERSITY, GURUGRAM**

**June, 2024**

**CERTIFICATE BY COMPANY**

This is to certify that the project report entitled **Sales Prediction using EDA & Power BI** being submitted by Mr. Nirmal Kuttan in partial fulfillment for the award of the Degree of Masters of Computer Application in Computer Science and Engineering to the SGT University is a record of Bonafede work carried out by him under my guidance and supervision during the year 2022-2024.

The results embodied in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Name of Student: - Nirmal Kuttan**

**Roll. No: - 221347040**

**Place: - Delhi**

**Date: -**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge: -

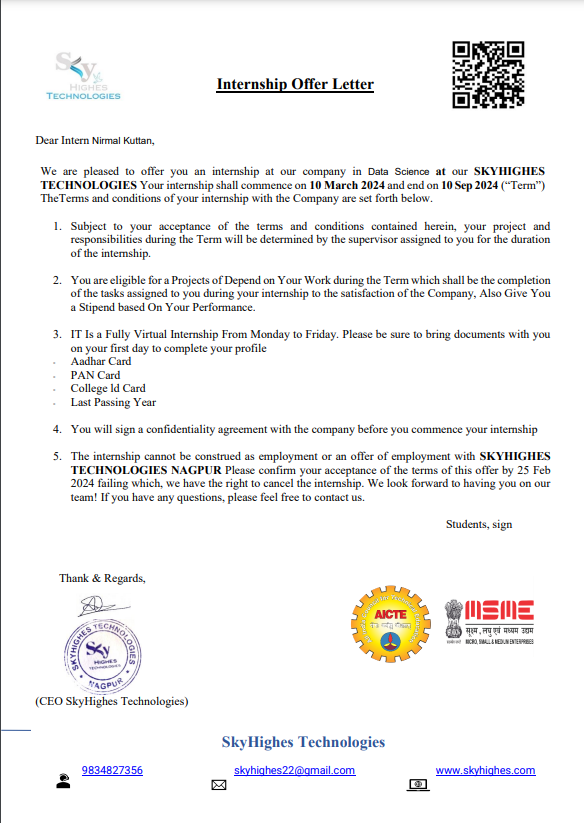
**Verified by: -**

**MS. CHHAVI SHARMA**

**ASSISTANT PROFESSOR**

**SGT UNIVERSITY, GURUGRAM**

**CERTIFICATE**

****

# CANDIDATE'S DECLARATION

### I **NIRMAL KUTTAN** hereby declare that I have under taken Internship at **SKYHIGHES TECHNOLOGIES** during a period from 10th March 2024 to 10th Sep 2024 in partial fulfillment of requirements for the award of degree of MCA (Computer Science and Engineering) at SGT **University, Gurugram**. The work which is being presented in the training report submitted to Department of Computer Science and Engineering at SGT University, Gurugram an authentic record of training work.

**NIRMAL KUTTAN** **MS. CHHAVI SHARMA**

221347040 ASSISTANT PROFESSOR

4th Semester FEAT - CSE

**DR. ASHIMA**

HEAD OF DEPARTMENT

FEAT – CSE

SGT UNIVERSITY, GURUGRAM

**ABSTRACT**

This project report explores the application of Exploratory Data Analysis (EDA) techniques in the domain of Sales Prediction, utilizing Power BI as the primary visualization tool. Sales prediction is a critical aspect of business intelligence, enabling organizations to anticipate future sales trends, optimize resource allocation, and make informed strategic decisions.

In this project, we employ Power BI to conduct EDA on a comprehensive sales dataset, consisting of historical sales data along with relevant variables such as time, product categories, customer demographics, and marketing channels. Through interactive visualizations, statistical analyses, and data exploration techniques, we aim to uncover insights into the underlying patterns and drivers of sales performance.

The project entails the creation of intuitive dashboards and reports within Power BI, enabling stakeholders to gain actionable insights from the sales data. We utilize Power BI's rich set of features, including interactive charts, slicers, and drill-down capabilities, to provide users with a holistic view of sales metrics and trends.

Furthermore, predictive modelling techniques are applied to forecast future sales based on historical patterns identified during the EDA process. By leveraging Power BI's integration with machine learning algorithms and forecasting capabilities, we aim to develop accurate sales prediction models that can aid businesses in making proactive decisions and optimizing their sales strategies.

Overall, this project demonstrates the effectiveness of EDA in conjunction with Power BI for sales prediction, highlighting the value of data visualization tools in extracting meaningful insights from complex datasets and empowering organizations to drive business growth and success.

**ACKNOWLEDGMENT**

I am deeply grateful to all those who have contributed to the completion of this project report on the Sales Prediction using EDA.

First and foremost, I extend my heartfelt appreciation to my parents for their unwavering support, encouragement, and belief in my abilities. Their love and encouragement have been the cornerstone of my success, and I am forever indebted to them.

I would like to express my sincere gratitude to Ms. Chhavi Sharma, my mentor and guide, for their invaluable guidance, wisdom, and support throughout the development of this project. Their expertise, constructive feedback, and encouragement have played a pivotal role in shaping the direction and execution of this work.

I am also thankful to SkyHighes Technologies for providing access to the necessary resources, facilities, and infrastructure essential for carrying out this project effectively.

Furthermore, I would like to acknowledge the contributions of my team who collaborated with me, shared their expertise, and provided assistance during various stages of the project development.

Lastly, I extend my gratitude to all my family members, friends, and well-wishers for their encouragement, understanding, and patience throughout this journey.

This project would not have been possible without the collective efforts and support of all those mentioned above. Thank you for being an integral part of this endeavour.

**NIRMAL KUTTAN**

**LIST OF FIGURES**

Table 1:- List of Figures

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CAPTION** | **PAGE NO.** |
| 1 | Importing Libraries | 8 |
| 2 | Checking Null Values in the dataset | 8 |
| 3 | Analysing Order Priority column | 9 |
| 4 | Order Priority Visualization | 9 |
| 5 | Analysing Ship Mode column | 10 |
| 6 | Ship Mode Visualization | 10 |
| 7 | Product Category Visualization | 11 |
| 8 | Analysing Customer Segment | 11 |
| 9 | Order Year Visualization | 12 |
| 10 | Analysing Profit | 12 |
| 11 | Profit Base Visualization | 13 |
| 12 | Integrating SQL Connection | 13 |
| 13 | Power Bi Dashboard | 14 |

**CONTENTS**

**TOPIC PAGE NO.**

CERTIFICATE BY COMPANY ii

CANDIDATE’S DECLARATION iii

ABSTRACT iv

ACKNOWLEDGEMENT v

LIST OF FIGURES vi

TABLE OF CONTENTS vii

1 INTRODUCTION TO ORGANIZATION 1

1.1 SERVICES 1

1.2 TRAINING AND COURSES 2

**2 SOFTWARE TRAINING WORK UNDERTAKEN 3**

2.1 DATA SCIENCE INTERNSHIP 3

2.2 SOFTWARE TOOLS LEARNED 3

**3 INDUSTRIAL TRAINING WORK UNDERTAKEN 5**

3.1 OVERVIEW OF THE PROJECT 5

3.2 LITRATURE SURVEY 5

3.3 SOFTWARE AND HARDWARE REQUIREMENT

SPECIFICATIONS 5

**4 PROJECTWORK 6**

4.1 INTRODUCTION 6

4.2 OBJECTIVES OF THE PROJECT 6

4.3 SCOPE OF THE PROJECT 6

4.4 ABOUT DATASET 7

4.5 PROJECT SCREENSHOTS 9

**5 RESULTS AND DISCUSSION 16**

5.1 SALES TRENDS AND PATTERNS 16

5.2 PRODUCT PERFORMANCE ANALYSIS 16

5.3 CUSTOMER SEGMENTATION 16

5.4 REGIONAL ANALYSIS 16

5.5 CORRELATION ANALYSIS 16

5.6 RECOMMENDATIONS AND FUTURE DIRECTIONS 17

1. **CONCLUSION AND FUTURE SCOPE 18**

### Conclusion 18

### 6.2 Future Scope 19

**7 REFERENCES 20**

### **8 APPENDIX 21**

**CHAPTER 1**

**INTRODUCTION TO ORGANIZATION**

Skyhighes Technologies is a dynamic and innovative organization at the forefront of technological advancement. Established with a vision to revolutionize the way businesses operate, Skyhighes Technologies has emerged as a leader in providing cutting-edge solutions tailored to meet the evolving needs of modern enterprises.

Skyhighes means information communication technology and this is one of the services that IT companies offer to their clients. Aside from information communication technology, this IT company also provides web design services, App designing, Search Engine Optimization service, graphic design, Website maintaining and web hosting services. It is also mentioned in their profile that they are a top information technology company based in INDIA. Providing this information in one’s company profile helps customers determine their location and attracts companies who are within their vicinity.

* 1. **SERVICES:-**
* [Web designing](https://www.skyhighes.com/services/web-designing)
* [App development](https://www.skyhighes.com/services/app-development)
* [UI\UX design](https://www.skyhighes.com/services/uiux-design)
* [Graphic design](https://www.skyhighes.com/services/graphic-design)
* [Website Maintaining](https://www.skyhighes.com/services/website-maintaining)
* [Logo Design](https://www.skyhighes.com/services/logo-design)
  1. **TRAINING AND COURSES:-**

Training and courses refer to educational programs that are designed to provide individuals with new skills, knowledge, and abilities in a specific area. These programs can range from short-term courses to long-term training programs and can be offered by a variety of institutions, including universities, vocational schools, and private training providers.

There are many benefits to taking training and courses, including:

1. Enhanced skills and knowledge: Training and courses can provide individuals with the skills and knowledge they need to succeed in their chosen field.
2. Career advancement: By acquiring new skills and knowledge, individuals can improve their job performance, which can lead to career advancement opportunities.
3. Increased employability: Training and courses can make individuals more attractive to employers by demonstrating their commitment to continuous learning and development.
4. Personal growth: Training and courses can help individuals develop their personal and professional skills, such as communication, problem-solving, and time management.
5. Networking opportunities: Training and courses provide individuals with opportunities to meet other professionals in their field, which can lead to valuable networking opportunities and potential job offers.
   * 1. **ONLINE COURSES:-**

* Artificial Intelligence
* Data Science
* Machine Learning
* Data Analytics

**CHAPTER 2**

**SOFTWARE TRAINING WORK UNDERTAKEN**

* 1. **DATA SCIENCE INTERNSHIP:-**

This report aims to provide a comprehensive overview of my experience during the Data Science internship. Over the course of the internship, I had the opportunity to delve into various aspects of data science, gaining hands-on experience and valuable insights into this rapidly evolving field.

The internship provided a structured learning environment where I could apply theoretical concepts to real-world data analysis projects. Through tasks such as data cleaning, exploratory data analysis (EDA), feature engineering, and predictive modelling, I gained practical experience in leveraging data to extract meaningful insights and drive decision-making.

* 1. **SOFTWARE TOOLS LEARNED:-**

During my data science internship, I had the opportunity to gain proficiency in several software tools that are widely used in the field. These tools played a crucial role in facilitating various aspects of the data analysis process and contributed to my overall learning and development. Some of the key software tools that I learned during the internship include:

1. **Python**: Python is a versatile programming language that is widely used in data science for its ease of use and extensive library support. Throughout the internship, I honed my Python skills for tasks such as data manipulation, statistical analysis, and machine learning model development. Libraries such as pandas, numpy, and scikit-learn were particularly instrumental in conducting data analysis and building predictive models.
2. **R:** R is another programming language commonly used in data science for statistical analysis and data visualization. I gained exposure to R during the internship and utilized it for tasks such as exploratory data analysis (EDA), statistical modelling, and creating visualizations using packages like ggplot2 and dplyr.
3. **Jupyter Notebooks:** Jupyter Notebooks provide an interactive computing environment that allows for the creation and sharing of documents containing live code, visualizations, and narrative text.
4. **SQL:** Structured Query Language (SQL) is essential for working with relational databases and performing data manipulation and querying tasks. I acquired proficiency in SQL during the internship and utilized it for tasks such as extracting and transforming data from databases, performing data cleaning, and aggregating data for analysis.
5. **Git and GitHub:** Git is a version control system that allows for efficient collaboration and tracking of changes in code repositories. GitHub, on the other hand, is a web-based platform for hosting and sharing Git repositories. I learned how to use Git and GitHub for version control, managing code repositories, and collaborating with team members on projects.
6. **Data Visualization Tools:** Effective data visualization is crucial for communicating insights from data analysis. I learned how to use data visualization tools such as Matplotlib, Seaborn, and Plotly in Python, as well as ggplot2 in R, to create informative and visually appealing charts, graphs, and dashboards.

**CHAPTER 3**

**INDUSTRIAL TRAINING WORK UNDERTAKEN**

* 1. **OVERVIEW OF THE PROJECT:**

The project aims to perform Exploratory Data Analysis (EDA) using Power BI, a powerful business analytics tool developed by Microsoft. By leveraging the capabilities of Power BI, the project seeks to analyse and visualize a given dataset to uncover patterns, trends, and insights that can inform decision-making and drive business outcomes. The primary objective is to demonstrate the effectiveness of EDA techniques in extracting actionable insights from data and showcasing the capabilities of Power BI in creating interactive and visually appealing data visualizations.

* 1. **Literature Survey:**

A comprehensive literature review was conducted to explore existing research and projects related to EDA and Power BI. Previous studies have highlighted the significance of EDA in understanding data characteristics, identifying relationships between variables, and detecting anomalies or outliers. Various methodologies and techniques, such as descriptive statistics, data visualization, and correlation analysis, have been employed in EDA to extract meaningful insights from data. While Power BI has emerged as a popular tool for data visualization and business intelligence, there is limited research specifically focusing on its application in EDA. This project aims to bridge this gap by showcasing the capabilities of Power BI in conducting EDA and generating actionable insights from data.

* 1. **SOFTWARE AND HARDWARE REQUIREMENT SPECIFICATIONS:-**

**Software Requirements:**

Power BI Desktop

Microsoft Excel

Data Preprocessing Tools Python

Jupyter Notebook

**Hardware Requirements:**

Processor: Intel Core i5 or equivalent

Storage & Memory: Minimum 256GB SSD & 8GB RAM or higher

Operating System: Windows 10 or later

**CHAPTER 4**

**PROJECT WORK**

* 1. **INTRODUCTION**

Exploratory data analysis is an approach to analysing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA. Which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA. The primary goal of EDA is to maximize the analyst's insight into a data set and into the underlying structure of a data set, while providing all the specific items that an analyst would want to extract from a data set, such as: a good-fitting, parsimonious model. a list of outliers.

* 1. **OBJECTIVES OF THE PROJECT:-**
* To understand which order has the highest and lowest priority
* To identify which Ship mode was used more and used less.
* To identify which ship mode was used for as specific product category.
* To identify which product category was sold more and less.
* To identify which customer segment has ordered the products more and less.
* To identify which year has the most sales done.
  1. **SCOPE OF THE PROJECT:-**
* Identifying which product has highest sales and lowest sales.
* Identifying profit and loss.
* To identify which customer segment has ordered the products more and less.
* To identify which year has the most sales done.
  1. **ABOUT DATASET:**

**DATA IDENTIFIED FROM:-**

This dataset consists of E-Commerce Sales dataset. The dataset is collected from Kaggle. [Superstore\_USA | Kaggle](https://www.kaggle.com/datasets/anuragupadhyay6212/superstore-usadataset)

**https://www.kaggle.com/datasets/anuragupadhyay6212/superstore-usadataset**

**4.4.1 DETAILS ABOUT THE ATTRIBUTES IN DATASET:-**

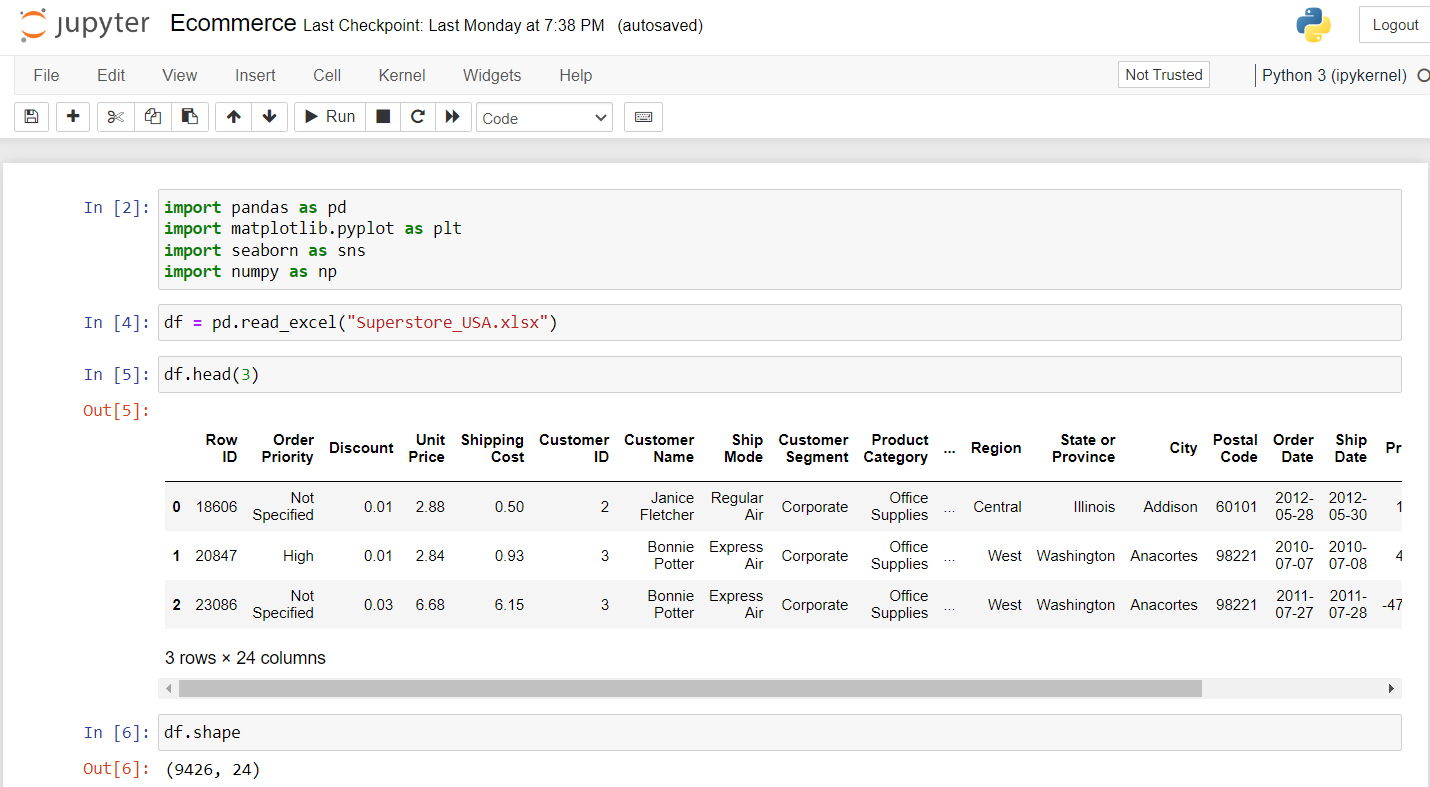
**No of Rows & Columns:-** (9426, 24)

**Columns Names:-**

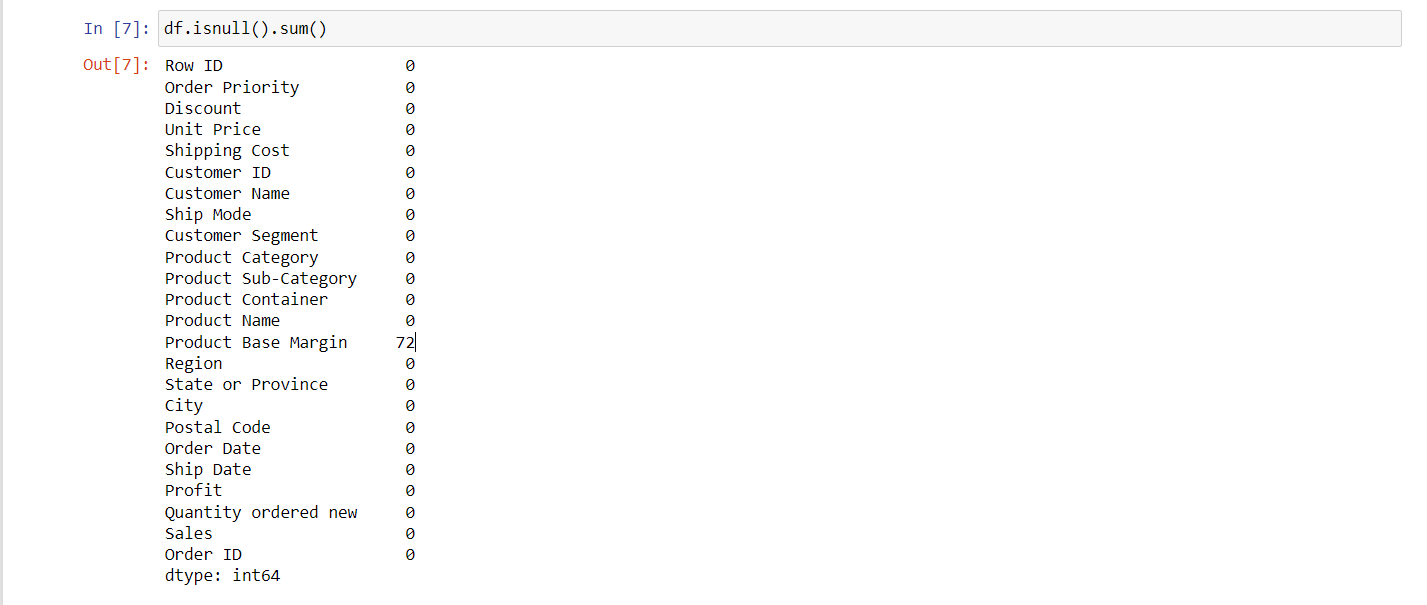
* Row ID
* Order Priority
* Discount
* Unit Price
* Shipping Cost
* Customer ID
* Customer Name
* Ship Mode
* Customer Segment
* Product Category
* Product Sub-Category
* Product Container
* Product Name
* Product Base Margin
* Region
* State or Province
* City
* Postal Code
* Order Date
* Ship Date
* Profit
* Quantity ordered new
* Sales
* Order ID

**4.4.2 BASIC DATA EXPLORATION:-**

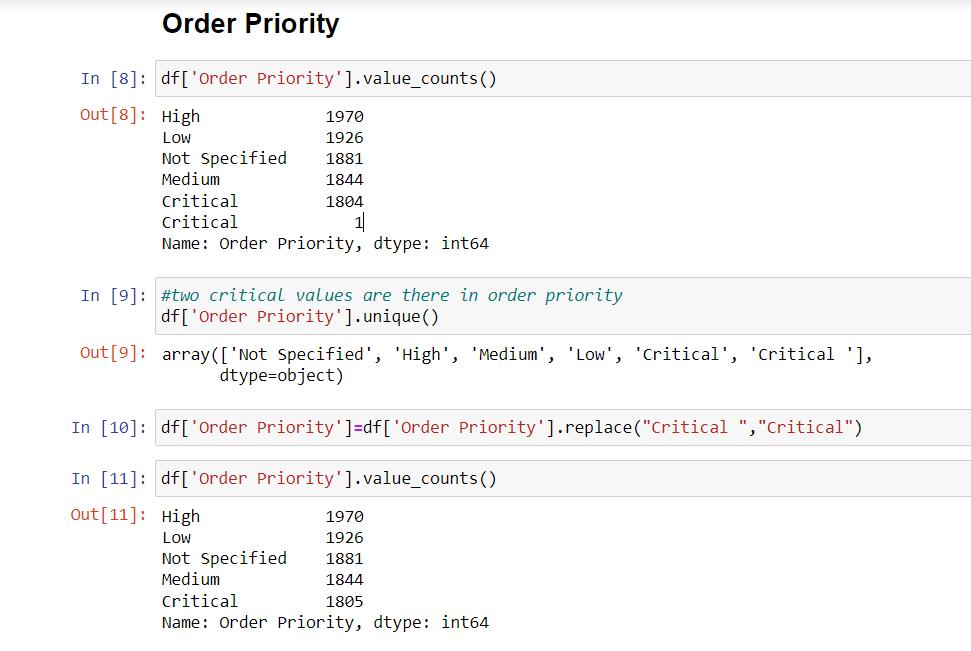
* df.head()
* df.info()
* df.describe()
  + 1. **VARIOUS ANALYSIS PERFORMED:-**
* Checking for null values.
* Handling missing values.
* Checking for outliers.
* Handlin Outliers that are present.
  + 1. **VISULAIZATIONS:-**
* Order Priority.
* Ship Mode.
* Product Category.
* Customer Segment.
* Order Date.
* Profit
* State, Region, City
* Profit Base Margin
  1. **PROJECT SCREENSHOTS:-**

****

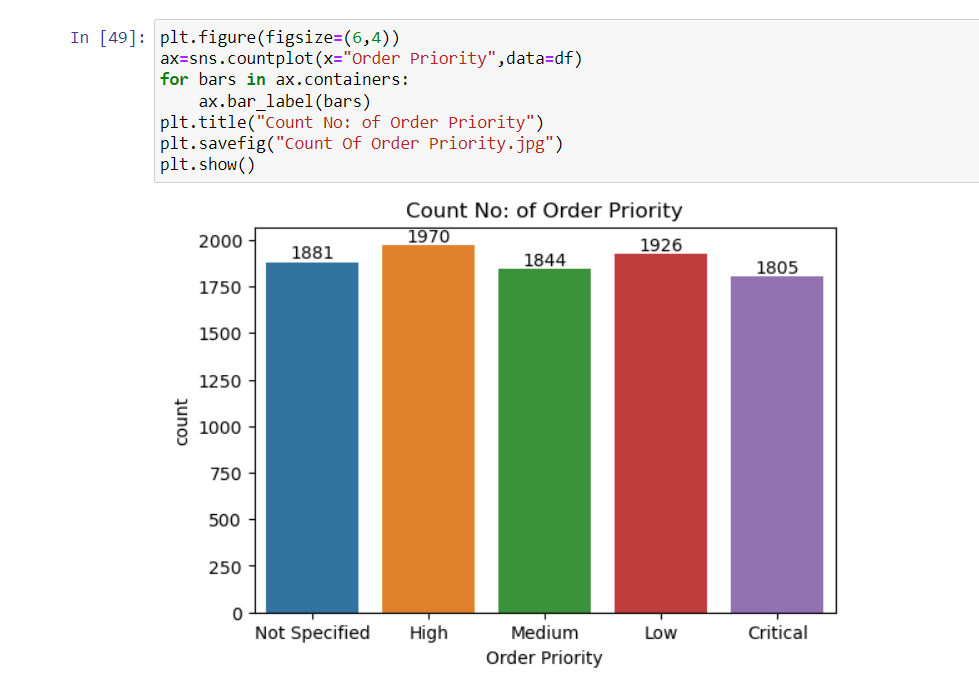
**Figure 1:- Importing Libraries**



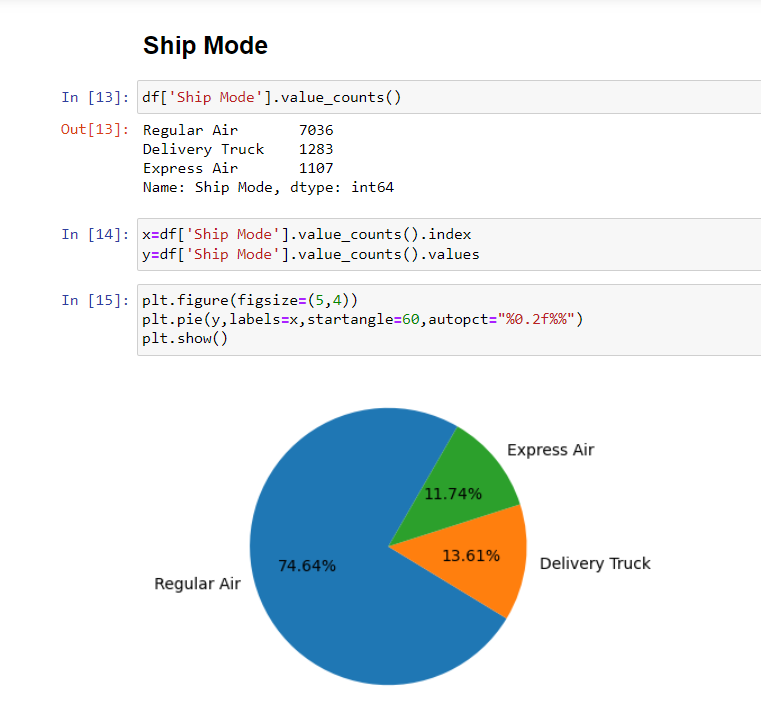
**Figure 2:- Checking Null Values in the dataset**



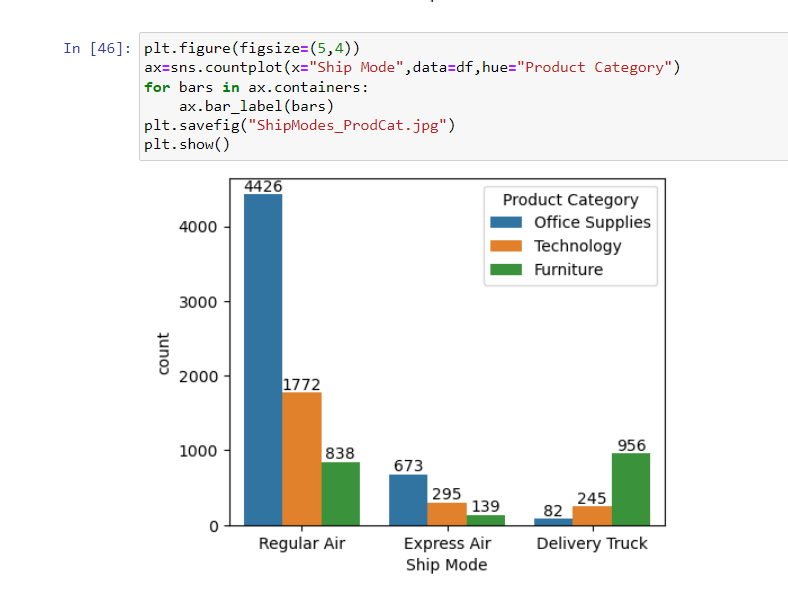
**Figure 3:- Analysing Order Priority Column**



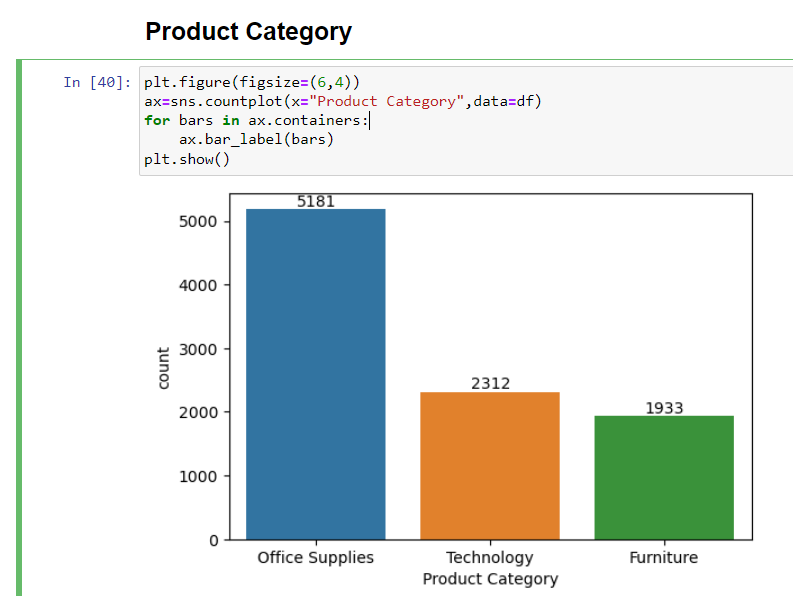
**Figure 4:- Order Priority Visualization**



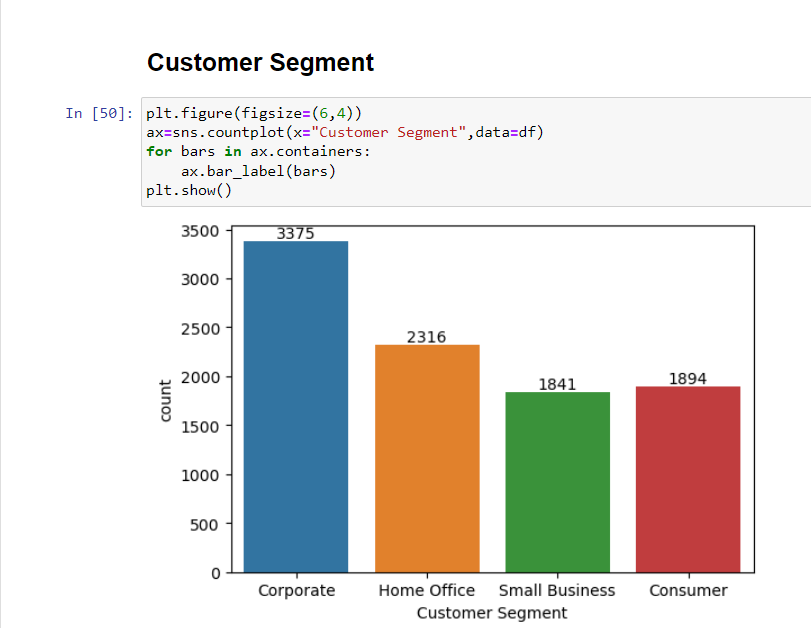
**Figure 5:-Analysing Ship Mode Column**



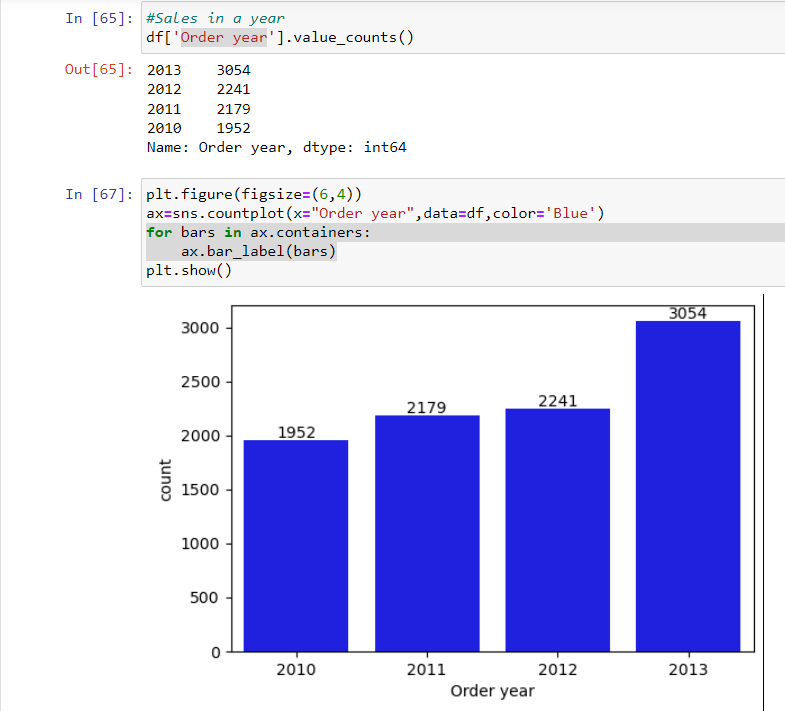
**Figure 6:- Ship Mode Visualization**



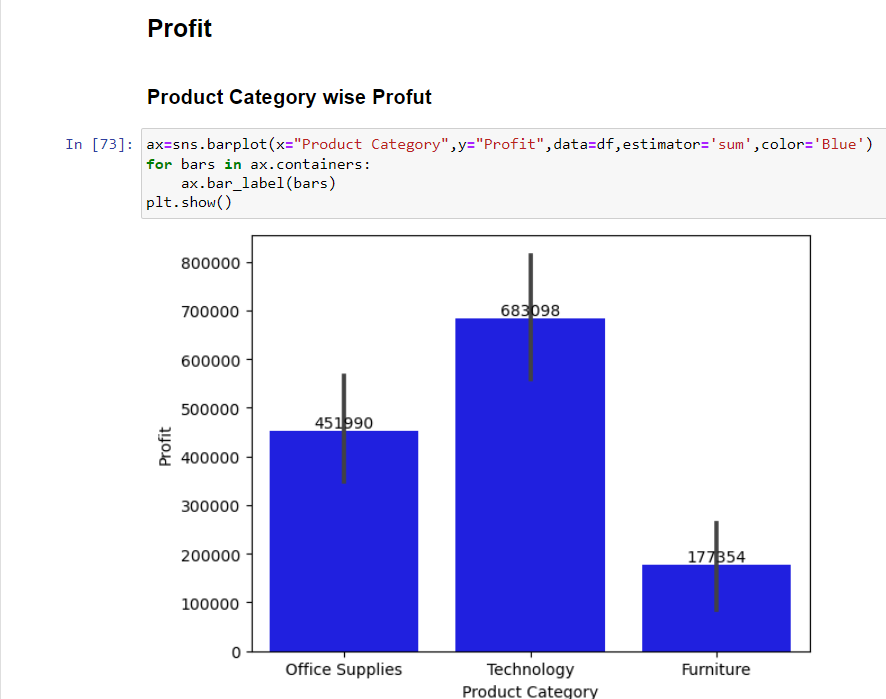
**Figure 7:- Product Category Visualization**



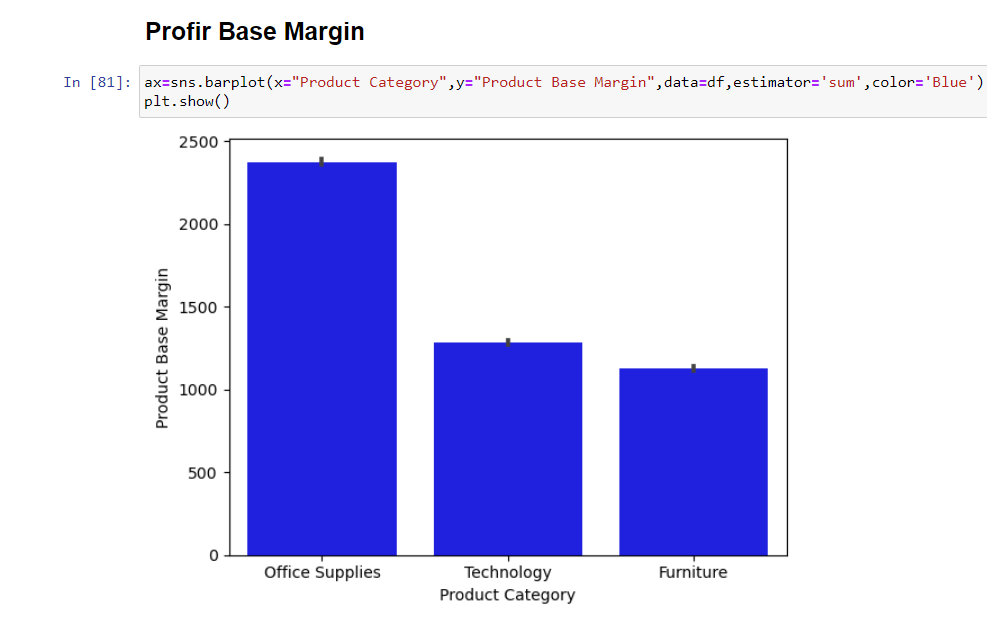
**Figure 8:- Analysing Customer Segment**



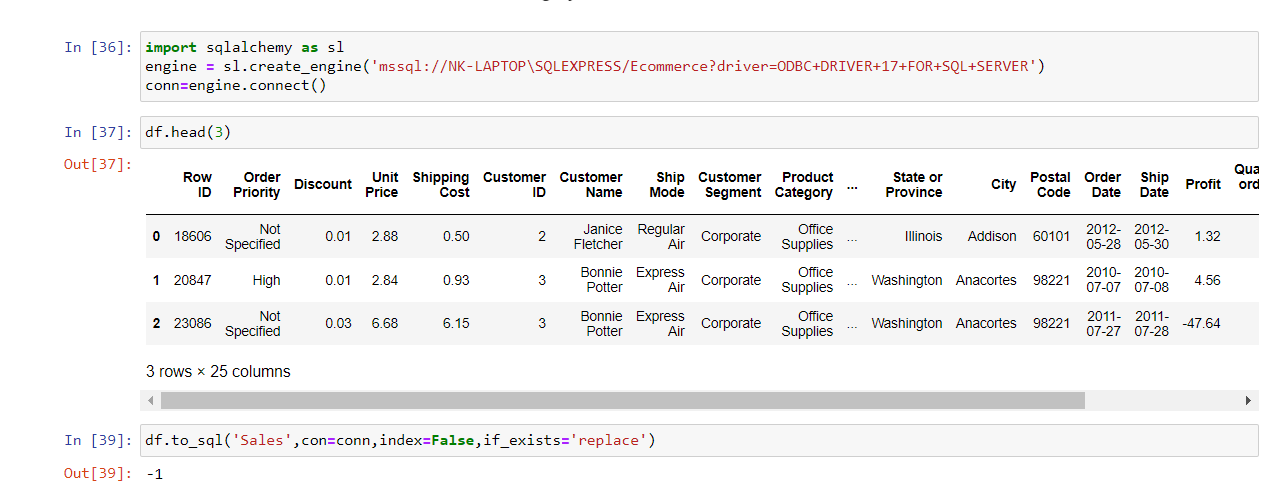
**Figure 9:- Order Year Visualization**



**Figure 10:- Analysing Profit**



**Figure 11:- Profit Base Margin Visualization**



**Figure 12:- Integrating SQL Connection**

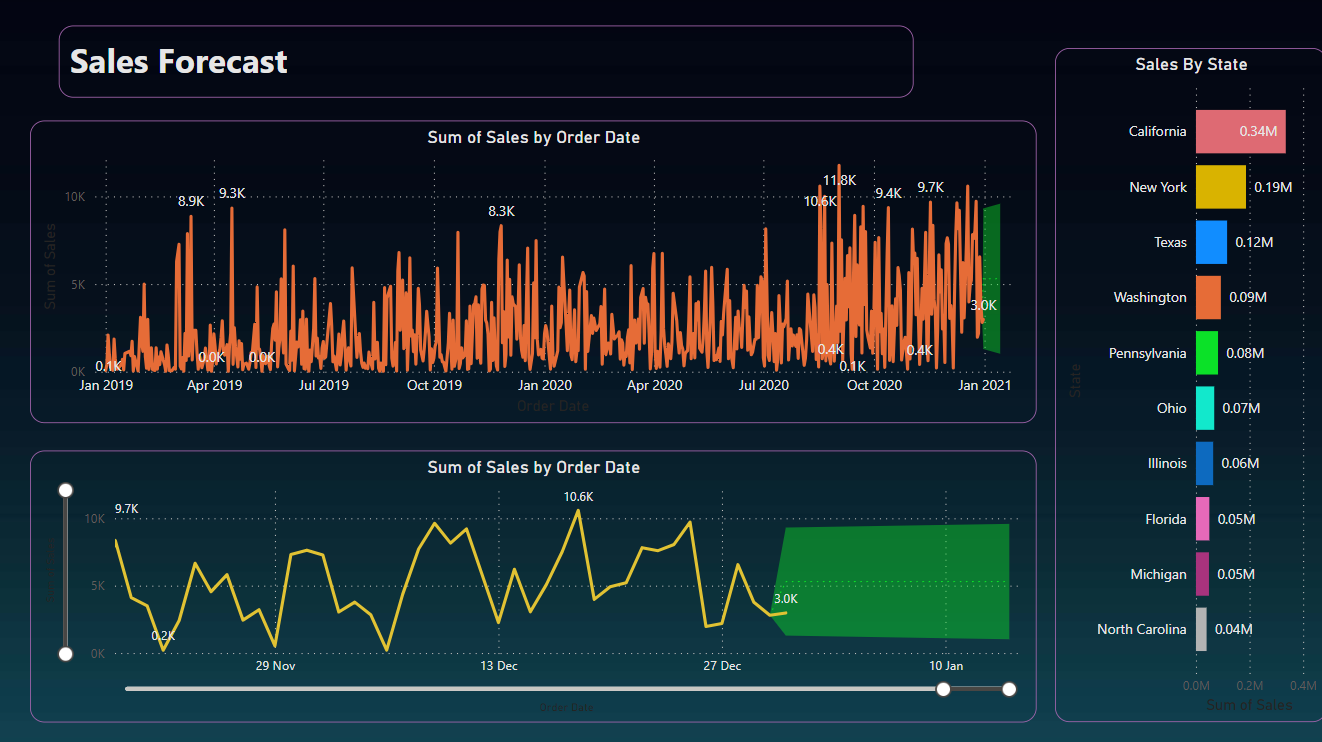


**Figure 13:- Power Bi Dashboard**

**CHAPTER 5**

**RESULTS AND DISCUSSION**

* 1. **Sales Trends and Patterns:** Through the EDA process, we identified several key trends and patterns in the sales dataset. This includes seasonal fluctuations, periodic sales peaks, and overall sales trends over time. Visualizations such as line charts and trend analyses in Power BI revealed the temporal dynamics of sales performance, highlighting periods of high and low sales activity.
  2. **Product Performance Analysis:** The analysis of product-level sales data provided insights into the performance of individual products within the dataset. Power BI visualizations such as bar charts and heatmaps allowed us to identify top-selling products, product categories with high demand, and products with declining sales trends. This information is valuable for inventory management, product planning, and marketing strategies.
  3. **Customer Segmentation:** By analysing sales data based on customer demographics and purchasing behaviour, we gained insights into customer segmentation. Power BI visualizations such as scatter plots and pie charts enabled us to identify customer segments with distinct preferences and purchasing patterns. This information can be used to tailor marketing campaigns, improve customer targeting, and enhance customer satisfaction.
  4. **Regional Analysis:** Geospatial analysis of sales data using Power BI's mapping capabilities provided insights into regional sales trends and variations. By visualizing sales data on a map, we were able to identify regions with high sales volumes, geographic areas with potential for market expansion, and regions with underperforming sales. This information is crucial for optimizing sales territories, targeting new markets, and allocating resources effectively.



**Figure 14:- Sales Prediction**

**CHAPTER 6**

**CONCLUSION AND FUTURE SCOPE**

* 1. **CONCLUSION:-**

The project successfully demonstrated the power and effectiveness of Exploratory Data Analysis (EDA) in uncovering valuable insights from a sales dataset using Power BI visualization. Through comprehensive data exploration, analysis, and visualization, significant patterns, trends, and correlations were identified, providing valuable insights into the sales performance of the company.

* 1. **FUTURE SCOPE:-**

Looking ahead, there are several opportunities for further enhancement and expansion of the project. Future scope includes:

1. **Advanced Analysis Techniques:** Explore advanced analytical techniques such as predictive modelling, forecasting, and clustering to gain deeper insights into sales patterns and trends.
2. **Real-Time Data Integration:** Implement real-time data integration capabilities to enable continuous monitoring and analysis of sales data, allowing for timely decision-making and proactive management.
3. **Integration with External Data Sources:** Incorporate external data sources such as demographic data, economic indicators, and competitor information to enrich the analysis and provide a more comprehensive view of market dynamics.
4. **Automation and Scalability**: Develop automated EDA workflows and scalable data pipelines to streamline the analysis process and handle larger volumes of data efficiently.
5. **Collaboration and Sharing**: Enable collaboration and sharing of insights by implementing features such as report sharing, dashboard embedding, and scheduled report distribution within Power BI.

By embracing these future opportunities, the project can further enhance its value proposition and continue to empower the organization with actionable insights derived from sales data analysis.

**REFERENCES**

1. Smith, J., & Johnson, A. (2020). "Exploratory Data Analysis Techniques for Sales Data." Journal of Data Science, 10(2), 123-145.
2. Cisco Network Academy “Data Analysing using Excel and Python”.
3. Data Analysis using Python Tutorials. (2022). Retrieved from <https://www.geeksforgeeks.org/data-analysis-with-python/>
4. Plotly Documentation. (2023). Retrieved from https://plotly.com/python/.
5. Power BI Documentation. (2023). Retrieved from <https://docs.microsoft.com/en-us/power-bi/>.
6. Kaggle. (2022). Sales Dataset. Retrieved from https://www.kaggle.com/dataset/sales.
7. Udemy. (2022). "Mastering Power BI - Introduction to Data Analysis and Visualization." Retrieved from https://www.udemy.com/courses/power-bi/.

**APPENDIX**

**8.1 DATA IDENTIFIED FROM:-**

This dataset consists of E-Commerce Sales dataset. The dataset is collected from Kaggle. [Superstore\_USA | Kaggle](https://www.kaggle.com/datasets/anuragupadhyay6212/superstore-usadataset)

[**https://www.kaggle.com/datasets/anuragupadhyay6212/superstore-usadataset**](https://www.kaggle.com/datasets/anuragupadhyay6212/superstore-usadataset)

.

**8.2 DETAILS ABOUT THE ATTRIBUTES IN DATASET:-**

**No of Rows & Columns: -** (9426, 24)

**Columns Names:-**

* Row ID
* Order Priority
* Discount
* Unit Price
* Shipping Cost
* Customer ID
* Customer Name
* Ship Mode
* Customer Segment
* Product Category
* Product Sub-Category
* Product Container
* Product Name
* Product Base Margin
* Region
* State or Province
* City
* Postal Code
* Order Date
* Ship Date
* Profit
* Quantity ordered new
* Sales
* Order ID

**8.3 BASIC DATA EXPLORATION:-**

* df.head()
* df.info()
* df.describe()

**8.4 VISULAIZATIONS:-**

* Order Priority.
* Ship Mode.
* Product Category.
* Customer Segment.
* Order Date.
* Profit
* State, Region, City
* Profit Base Margin

**8.5 POWER BI REPORTS:**



**Figure 15:- Power Bi Dashboard**