Business Analysis

PRESENTATION

Predictive Analysis

-BY,
PRAVEEN R
36822018
MCA

Team Members

01 ANANYA MANOJ - Descriptive Analysis

02 HARINI K - Diagnostic Analysis

03 MAGHALAKSHMI R - Prescriptive Analysis

04 NANDHINIDEVI M - Location Analysis

05 NITHIYASRI S - Cost Analysis

06 **PRAVEEN R - Predictive Analysis**

Predictive Analysis

- The main objective of predictive analysis is to use data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data.
- Here, I have predicted the sales by considering the features such as ship mode, category, subcategory, quantity, profit, discount.

Overview

INTRODUCTION

Super Store is a small retail business located in the United States. They sell Furniture, Office Supplies and Technology products and their customers are the mass Consumer, Corporate and Home Offices.
Our task is to analyse the sales data and identify weak areas and opportunities for Super Store to boost business growth.

Dataset

DATASET

This dataset provides comprehensive view of the sales transactions, products, customers, discount, and geographical details.

```
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   object
 3   Ship Date      9994 non-null   object
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Subcategory    9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
```

Generating Word Cloud

```
# Generate Word Cloud for the 'Product Name' column
text_data = ' '.join(superstore_data['Product Name'].astype(str))
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text_data)
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

Inference

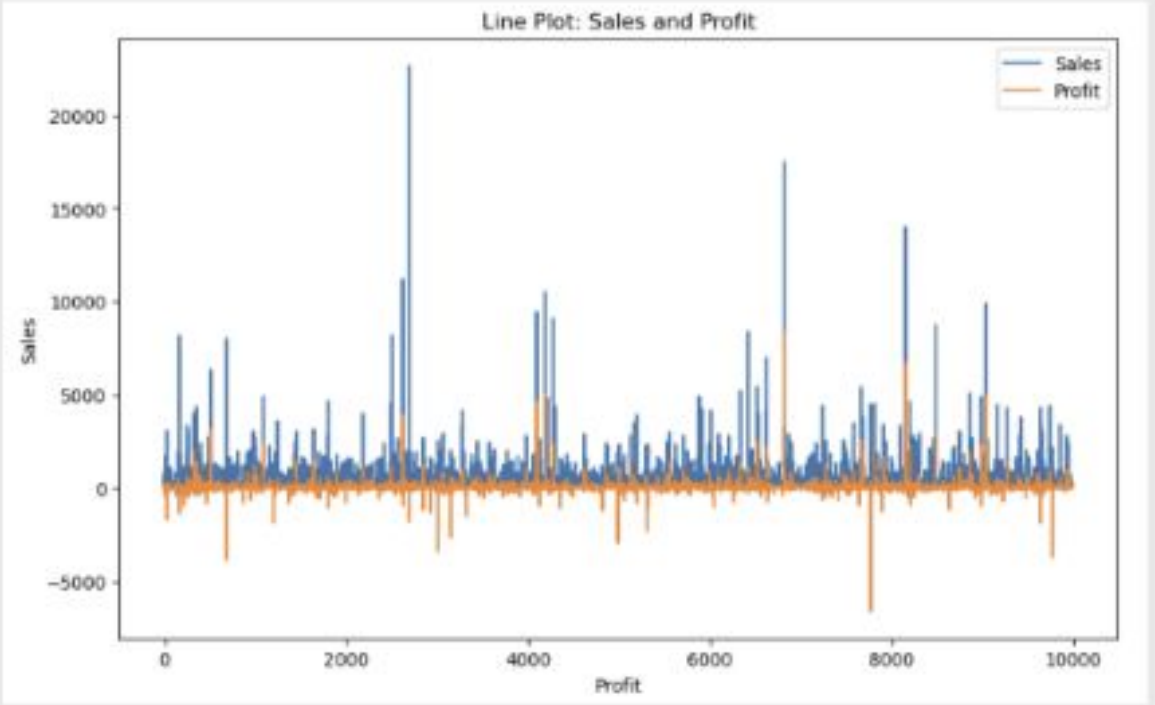
- By examining the word cloud, you can gain insights into the distribution of product names and identify which products are more prevalent or frequently mentioned in the dataset.
- This visualization is particularly useful for understanding the diversity of products in the dataset and identifying potential patterns or trends in product naming.

Interpretation

- The word cloud visually represents the frequency of words (in this case, product names) in the 'Product Name' column.
- Larger and bolder words in the cloud indicate higher frequency or prevalence of those words in the dataset.
- It provides a quick and intuitive way to identify the most common or prominent product names in the dataset.

Line plot for Sales & Profit

```
# Visualize Sales and Profit
plt.figure(figsize=(10, 6))
sns.lineplot(x=superstore_data.index, y='Sales', label='Sales', data=superstore_data)
sns.lineplot(x=superstore_data.index, y='Profit', label='Profit', data=superstore_data)
plt.title('Line Plot: Sales and Profit')
plt.xlabel('Profit')
plt.ylabel('Sales')
plt.legend()
plt.show()
```



Inference

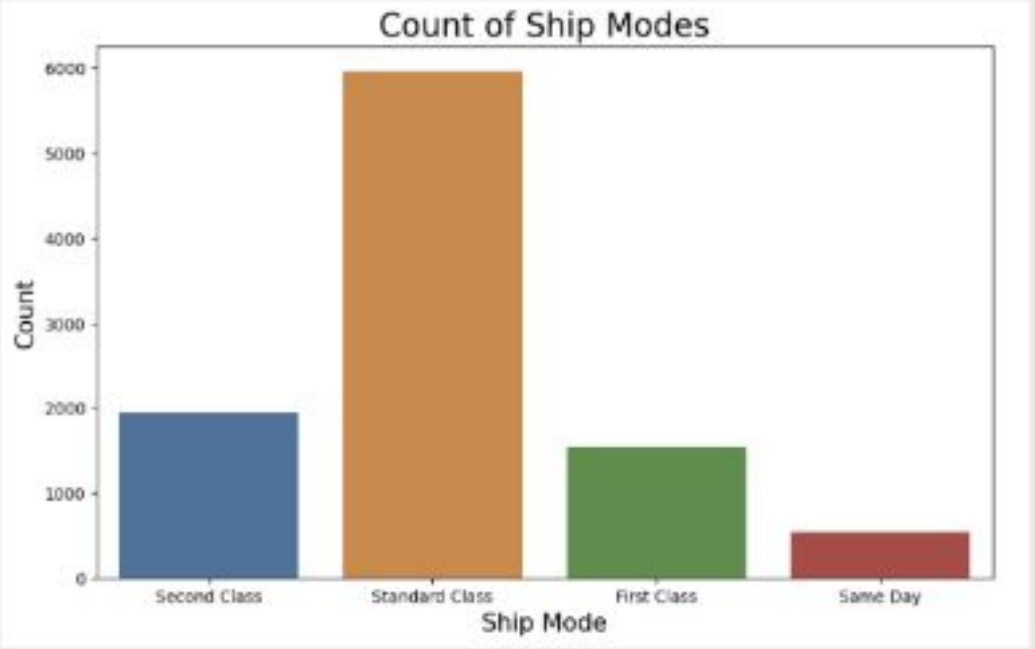
- Inference: The provided code generates a line plot depicting the trends of 'Sales' and 'Profit' over the index of the superstore_data DataFrame. The Seaborn library is used to create separate lines for 'Sales' and 'Profit,' and the resulting plot is displayed with labeled axes and a legend.

Interpretation

- **Sales and Profit Trends:** The line plot allows us to visually assess the trends in both 'Sales' and 'Profit' over the dataset's index. If the lines follow similar patterns, it suggests a potential positive correlation between sales and profit.
- **Fluctuations and Variability:** Fluctuations in the lines suggest variability in sales and profit over time. Understanding the causes behind these fluctuations could be crucial for strategic decision-making.

Count Of Each Ship Modes

```
# Visualize the count of each Ship Mode
plt.figure(figsize=(10, 6))
sns.countplot(x="Ship Mode", data=superstore_data)
plt.xlabel("Ship Mode", fontsize=15)
plt.ylabel("Count", fontsize=15)
plt.title("Count of Ship Modes", fontsize=20)
plt.show()
```



Inference

- By looking at the resulting plot, you can infer the distribution of different shipping modes used by the Superstore. Each bar represents a Ship Mode, and the height of the bar indicates the count or frequency of that particular mode.
- For example, if there are four types of ship modes (e.g., Standard Class, Second Class, First Class, and Same Day), the plot will have four bars, each showing how many times each mode appears in the dataset.

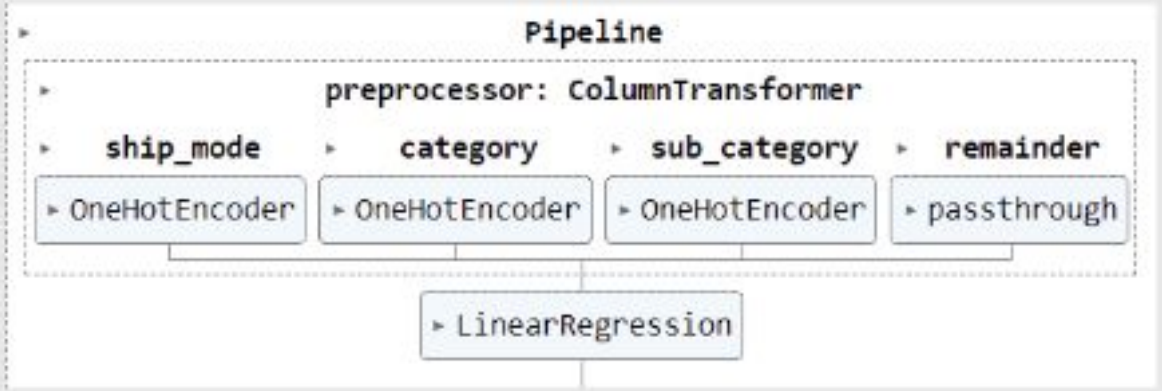
Interpretation

- The code uses the Seaborn library to create a count plot (sns.countplot) with "Ship Mode" on the x-axis and the count of each Ship Mode on the y-axis.

Implementation

```
# Apply one-hot encoding to the 'Ship Mode', 'Category', and 'Subcategory' columns
preprocessor = ColumnTransformer(
    transformers=[
        ('ship_mode', OneHotEncoder(drop='first'), ['Ship Mode']),
        ('category', OneHotEncoder(drop='first'), ['Category']),
        ('sub_category', OneHotEncoder(drop='first'), ['Subcategory'])
    ],
    remainder='passthrough'
)
# Create a pipeline with the preprocessor and the Linear Regression model
model = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('regressor', LinearRegression())
])
# Train the model
model.fit(X, y)
```

The Alphabetic values in the ship mode, category, subcategory are changed to numeric values using one hot encoder then combines in linear regression model..



Inference

- This code is setting up a machine learning pipeline for a linear regression model after preprocessing categorical features with one-hot encoding.
- The intention is to predict a target variable (y) based on features (X), considering the provided categorical features along with potential other features that are not explicitly mentioned.

Interpretation

- The use of one-hot encoding suggests that the 'Ship Mode', 'Category', and 'Subcategory' columns are categorical, and the model is being designed to handle these categorical variables appropriately.
- Linear regression is a simple, interpretable model that assumes a linear relationship between the features and the target variable. The interpretation of coefficients in linear regression can provide insights into the influence of each feature on the predicted outcome. However, the specifics would depend on the actual data used for training.

Make Prediction

```
# Make predictions for a new data point
new_data_point = pd.DataFrame([[ 'Second Class', 'Furniture', 'Bookcases', 5, 0.1, 50]],
                              columns=[ 'Ship Mode', 'Category', 'Subcategory', 'Quantity', 'Discount', 'Profit'])
predicted_sales = model.predict(new_data_point)

print(f'Predicted Sales: {predicted_sales[0]}')
```

```
Predicted Sales: 635.8765995173766
```

- Here, our model predicts sales considering features such as Ship mode, Category, Subcategory, Quantity, Discount, and Profit.
- We will provide the values for these features, and the model will predict the sales upon execution.

Inference

- **New Data Point Creation:** This data point represents a product with specific attributes, including ship mode, category, subcategory, quantity, discount, and profit.
- **Prediction Using the Model:** The model.predict() function takes the features of the new data point as input and returns the predicted sales value.
- **Print Predicted Sales:** The predicted sales value is printed to the console using the print statement.

Interpretation

- The new data point is an example of the input that the model would typically receive for making predictions.
- The prediction process involves passing the features of the new data point through the trained model to obtain an estimated sales value.
- The printed output provides the predicted sales value for the given input.

Reference

- https://chat.openai.com/
- https://www.kaggle.com/datasets/vivek468/superstore-dataset-final
- https://medium.com/analytics-vidhya/exploratory-data-analysis-super-store-cb91c37bcb06
- https://github.com/
- https://www.geeksforgeeks.org/step-by-step-predictive-analysis-machine-learning/

Thank You