

```
[3]: # TASK 4
# SUPERSTORE SALES PREDICTION
```

```
[78]: """
Time series analysis deals with time series based data to extract patterns for
predictions and other characteristics of the data.
• It uses a model for forecasting future values in a small time frame based on
previous observations.
• It is widely used for non-stationary data, such as economic data, weather data,
stock prices, and retail sales forecasting
"""
```

```
[78]: '\nTime series analysis deals with time series based data to extract patterns for\npredictions and other characteristics of the data.\n• It uses a model fo\nr forecasting future values in a small time frame based on\\previous observations.\n• It is widely used for non-stationary data, such as economic data, wea\nther data,\\nstock prices, and retail sales forecasting\\n'
```

```
[83]: import pandas as pd
# D:\\internship\\New folder\\SUPERSTORE SALES PREDICTION
train_data=pd.read_csv("D:\\\\internship\\\\New folder\\\\SUPERSTORE SALES PREDICTION\\\\train.csv")
train_data
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs
2	3	CA-2017-138688	12/6/2017	16/06/2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels
3	4	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	FUR-TA-10000577	Furniture	Tables
4	5	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	OFF-ST-10000760	Office Supplies	Storage
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
9795	9796	CA-2017-125920	21/05/2017	28/05/2017	Standard Class	SH-19975	Sally Hughsby	Corporate	United States	Chicago	Illinois	60610.0	Central	OFF-BI-10003429	Office Supplies	Binders
9796	9797	CA-2016-128608	12/1/2016	17/01/2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	OFF-AR-10001374	Office Supplies	Art
9797	9798	CA-2016-128608	12/1/2016	17/01/2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-PH-10004977	Technology	Phones
9798	9799	CA-2016-128608	12/1/2016	17/01/2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-PH-10000912	Technology	Phones
9799	9800	CA-2016-128608	12/1/2016	17/01/2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615.0	East	TEC-AC-10000487	Technology	Accessories

9800 rows × 18 columns

```
[ ]:
```

```
[85]: train_data.columns
```

```
[85]: Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State',
       'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category',
       'Product Name', 'Sales'],
      dtype='object')
```

```
[86]: train_data.head()
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs

```

2   3   CA- 12/6/2017 16/06/2017 Second DV-13045 Darrin Corporate United Los California 90036.0 West OFF-LA- Office Labels A
- 138688          Class           Van Huff          Corporate States Angeles          0000240 Supplies La Type

```

```

3   4   US- 2016- 11/10/2016 18/10/2016 Standard SO-20335 Sean Consumer United Fort Florida 33311.0 South FUR-TA- Furniture Tables E
- 108966          Class           O'Donnell          Consumer States Lauderdale          0000577 Rect Series

```

```

4   5   US- 2016- 11/10/2016 18/10/2016 Standard SO-20335 Sean Consumer United Fort Florida 33311.0 South OFF-ST- Office Eld- Storage 'N R
- 108966          Class           O'Donnell          Consumer States Lauderdale          0000760 Supplies Region

```

```
[87]: train_data.isna()
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
9795	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9796	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9797	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9798	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9799	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

9800 rows × 18 columns

```
[54]: train_data.isna().sum()
```

Row ID	0
Order ID	0
Ship Date	0
Ship Mode	0
Customer ID	0
Customer Name	0
Segment	0
Country	0
City	0
State	0
Postal Code	11
Region	0
Product ID	0
Category	0
Sub-Category	0
Product Name	0
Sales	0

```
[53]: train_data.describe()
```

	Row ID	Postal Code	Sales
count	9800.000000	9789.000000	9800.000000
mean	4900.500000	55273.322403	230.769059
std	2829.160653	32041.223413	626.651875
min	1.000000	1040.000000	0.444000
25%	2450.750000	23223.000000	17.248000
50%	4900.500000	58103.000000	54.490000
75%	7350.250000	90008.000000	210.605000
max	9800.000000	99301.000000	22638.480000

```
[59]: train_data.dtypes
```

Row ID	int64
Order ID	object
Ship Mode	object
Customer ID	object
Customer Name	object
Segment	object
Country	object
City	object
State	object
Postal Code	float64
Region	object
Product ID	object
Category	object
Sub-Category	object
Product Name	object
Sales	float64

```
[ ]:
```

```
[89]: # If there are missing values, you can fill them using interpolation
data=train_data.interpolate()
data.head()
```

C:\Users\Admin\AppData\Local\Temp\ipykernel\_9084\4154297262.py:2: FutureWarning: DataFrame.interpolate with object dtype is deprecated and will raise in a future version. Call obj.infer\_objects(copy=False) before interpolating instead.
data=train\_data.interpolate()

```
[89]: Row ID Order ID Order Date Ship Date Ship Mode Customer ID Customer Name Segment Country City State Postal Code Region Product ID Category Sub-Category P
```

0	1	CA-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases	Sc Co Bc
1	2	CA-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs	Hon Upch S C
2	3	CA-138688	12/6/2017	16/06/2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels	A & La Type
3	4	US-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	FUR-TA-10000577	Furniture	Tables	E S Rect
4	5	US-108966	11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311.0	South	OFF-ST-10000760	Office Supplies	Storage	Eld 'N R

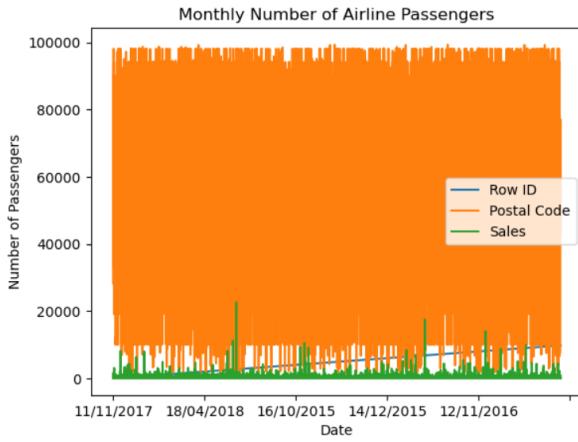
```
[90]: train_data.isna().sum()
```

```
[90]: Row ID      0
Order ID      0
Order Date    0
Ship Date     0
Ship Mode     0
Customer ID   0
Customer Name 0
Segment       0
Country       0
City          0
State          0
Postal Code   11
Region         0
Product ID    0
Category       0
Sub-Category   0
Product Name   0
Sales          0
dtype: int64
```

```
[ ]:
```

```
[69]: import matplotlib.pyplot as plt
```

```
train_data.plot()
plt.title('Monthly Number of Airline Passengers')
plt.xlabel('Date')
plt.ylabel('Number of Passengers')
plt.show()
```



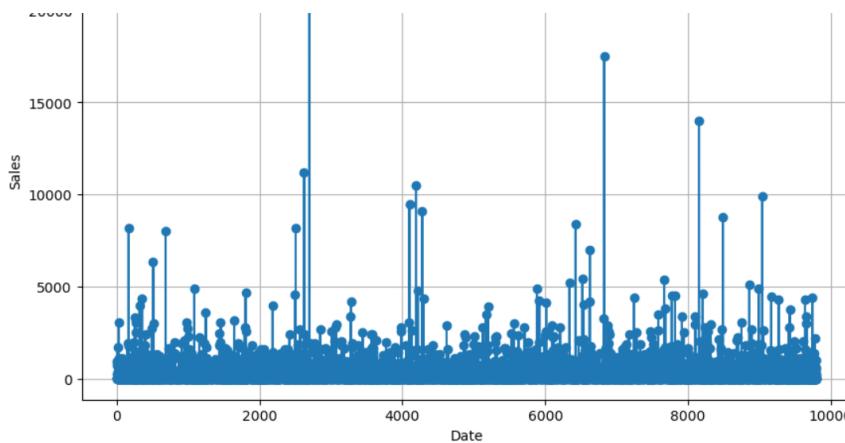
```
[71]: pip install statsmodels
```

```
Requirement already satisfied: statsmodels in c:\users\admin\anaconda3\lib\site-packages (0.14.0)
Requirement already satisfied: numpy>=1.18 in c:\users\admin\anaconda3\lib\site-packages (from statsmodels) (1.26.4)
Requirement already satisfied: scipy!=1.9.2,>=1.4 in c:\users\admin\anaconda3\lib\site-packages (from statsmodels) (1.11.4)
Requirement already satisfied: pandas>=1.0 in c:\users\admin\anaconda3\lib\site-packages (from statsmodels) (2.1.4)
Requirement already satisfied: patsy>=0.5.2 in c:\users\admin\anaconda3\lib\site-packages (from statsmodels) (0.5.3)
Requirement already satisfied: packaging>=21.3 in c:\users\admin\anaconda3\lib\site-packages (from statsmodels) (23.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\admin\anaconda3\lib\site-packages (from pandas>=1.0->statsmodels) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\admin\anaconda3\lib\site-packages (from pandas>=1.0->statsmodels) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\admin\anaconda3\lib\site-packages (from pandas>=1.0->statsmodels) (2023.3)
Requirement already satisfied: six in c:\users\admin\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

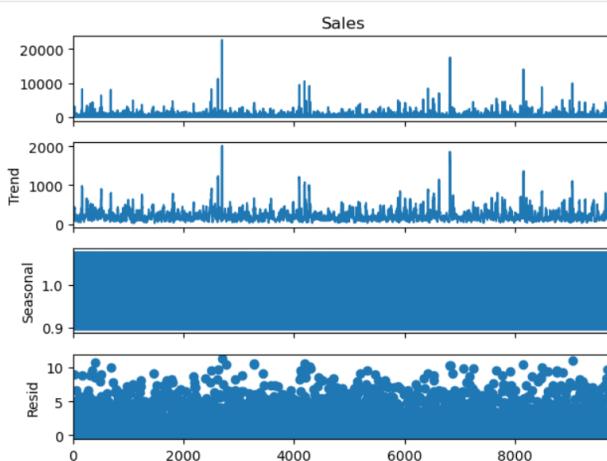
```
[92]: # EDA: Visualize the time series data
plt.figure(figsize=(10, 6))
plt.plot(train_data['Sales'], marker='o', linestyle='--')
plt.title('Superstore Sales Over Time')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.grid(True)
plt.show()
```

Superstore Sales Over Time





```
[104]: from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from sklearn.metrics import mean_squared_error
result = seasonal_decompose(data['Sales'], model='multiplicative', period=12)
result.plot()
plt.show()
```



```
[107]: train_size = int(len(train_data) * 0.8)
train_data, test_data = data.iloc[:train_size], data.iloc[train_size:]
predictions = model.forecast(len(test_data))
mse = mean_squared_error(test_data['Sales'], predictions)
print('Mean Squared Error:', mse)
```

```
Mean Squared Error: 390019.74701580574
```

```
[108]: future_predictions = model.forecast(steps=12)
```

```
[109]: plt.figure(figsize=(12, 6))
plt.plot(train_data.index, train_data['Sales'], label='Train')
plt.plot(test_data.index, test_data['Sales'], label='Test')
plt.plot(test_data.index, predictions, label='Predictions')
plt.plot(future_predictions.index, future_predictions, label='Future Predictions', marker='o')
plt.legend()
plt.title('Superstore Sales Prediction')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.show()
```

