

DEVELOP A LINEAR REGRESSION MODEL FOR FORECASTING

AIM:

To develop a linear regression model for forecasting time series data.

STEPS TO IMPLEMENT:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

file_path
= "C:/Users/Lenovo/Downloads/PRICE_AND_DEMAND_201801_NSW1.csv"

df = pd.read_csv(file_path)

df["timestamp"] = (df.index - df.index[0]).total_seconds()

X = df["timestamp"].values.reshape(-1, 1)
y = df["TOTALDEMAND"].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)

model = LinearRegression()
model.fit(X_train, y_train)

df["SETTLEMENTDATE"] = pd.to_datetime(df["SETTLEMENTDATE"],
format="%Y/%m/%d %H:%M:%S")

df = df.sort_values(by="SETTLEMENTDATE")
df.set_index("SETTLEMENTDATE", inplace=True)
```

```
LinearRegression ⓘ ?  
LinearRegression()
```

```
y_pred = model.predict(X_test)
```

```
index_values = df.index.to_numpy()
```

```
plt.figure(figsize=(10, 5))
```

```
plt.plot(index_values[len(X_train):], y_test, label="Actual Demand", color="blue")
```

```
plt.plot(index_values[len(X_train):], y_pred, label="Predicted Demand",  
color="red", linestyle="dashed")
```

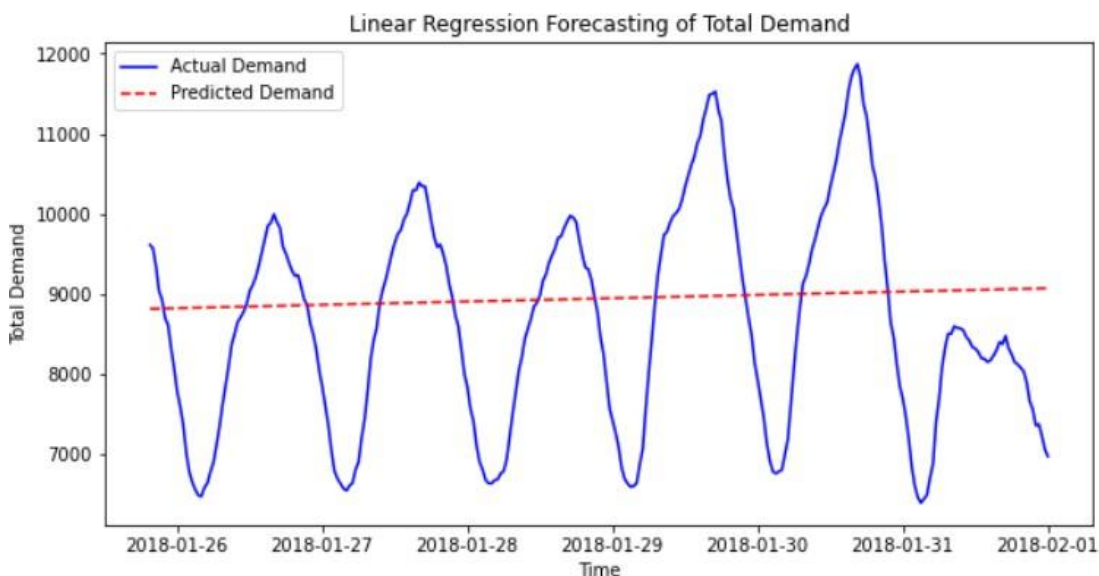
```
plt.xlabel("Time")
```

```
plt.ylabel("Total Demand")
```

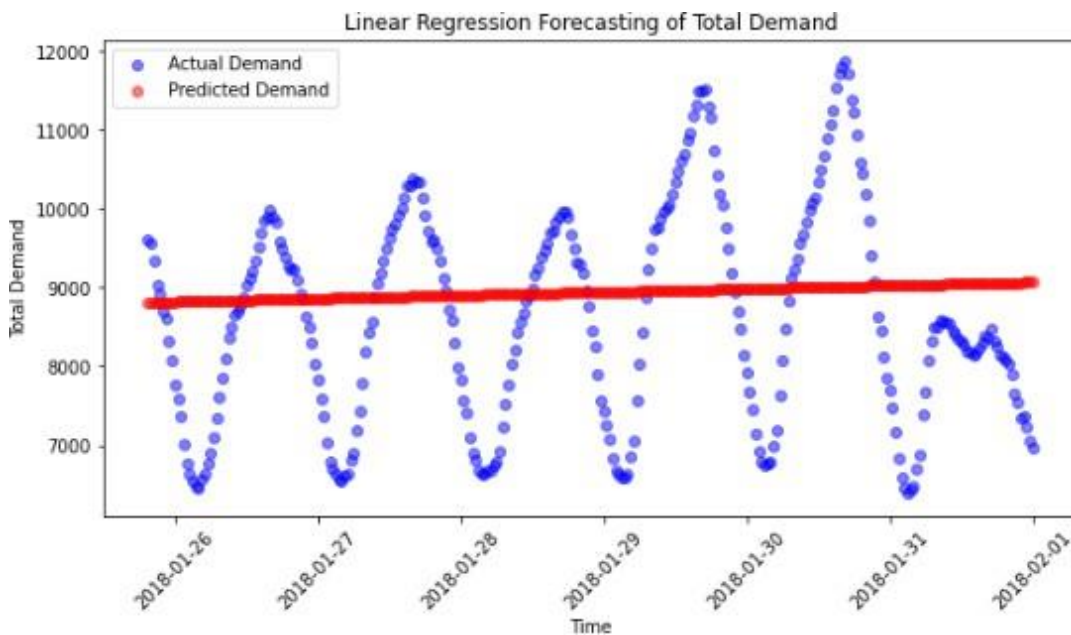
```
plt.title("Linear Regression Forecasting of Total Demand")
```

```
plt.legend()
```

```
plt.show()
```



```
index_values = np.array(df.index[len(X_train):])  
plt.figure(figsize=(10, 5))  
plt.scatter(index_values, y_test, label="Actual Demand", color="blue",  
alpha=0.5)  
plt.scatter(index_values, y_pred, label="Predicted Demand", color="red",  
alpha=0.5)  
plt.xlabel("Time")  
plt.ylabel("Total Demand")  
plt.title("Linear Regression Forecasting of Total Demand")  
plt.legend()  
plt.xticks(rotation=45)  
plt.show()
```



```
plt.figure(figsize=(12, 6))  
  
plt.plot(index_values, y_test, label="Actual Demand", color="blue",  
marker="o", linestyle="-", markersize=5)  
  
plt.plot(index_values, y_pred, label="Predicted Demand", color="red",  
marker="s", linestyle="--", markersize=5)  
  
plt.xlabel("Time")
```

Here are the steps:

1. Import the Required Libraries

Import the necessary libraries such as pandas, numpy, matplotlib, seaborn, and sklearn.

2. Load the Dataset

Load the CSV file into a pandas DataFrame using `pd.read_csv()`.

3. Convert the Date Column to DateTime

Convert the `SETTLEMENTDATE` column to a `datetime` format for easier manipulation.

4. Sort the Data by Date

Sort the DataFrame by the `SETTLEMENTDATE` column to ensure the data is in chronological order.

5. Set the Date Column as the Index

Set `SETTLEMENTDATE` as the index of the DataFrame for easier time-based analysis.

6. Create a New Column for Time (in Seconds)

Create a new column `timestamp` that represents the time elapsed (in seconds) from the first entry.

7. Prepare the Features and Target Variables

Define `x` (features) as the `timestamp` and `y` (target) as `TOTALDEMAND`.

8. Split the Data into Training and Testing Sets

Use `train_test_split` to split the data into training and testing sets. Ensure `shuffle=False` to maintain the temporal order.

9. Initialize and Train the Linear Regression Model

Create a Linear Regression model and fit it to the training data (`X_train, y_train`).

```
plt.ylabel("Total Demand")
```

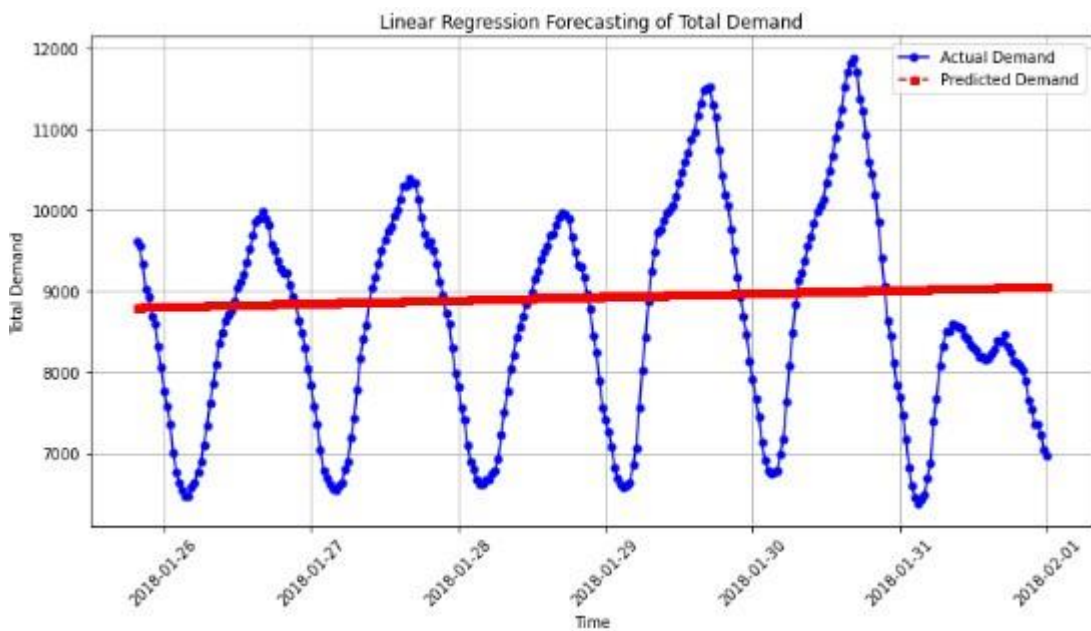
```
plt.title("Linear Regression Forecasting of Total Demand")
```

```
plt.legend()
```

```
plt.xticks(rotation=45)
```

```
plt.grid(True)
```

```
plt.show()
```



RESULT:

The program develop a linear regression model for forecasting time series data was executed successfully.