EX:No.5 DATE:7/02/25	**Develop a linear regression model for forecasting time series data.**

AIM:
Develop a linear regression model for forecasting time series data.

OBJECTIVE:

To develop a linear regression model to forecast future air pollution levels based on historical data.

BACKGROUND:

• **Linear regression** models the relationship between dependent and independent variables.

• In time-series forecasting, **time** (e.g., year, month) can be an independent variable for predicting pollution levels.

• Linear regression can help predict future pollution trends based on historical data.

• The model is simple but effective for linear relationships and can be used for short-term forecasts.

SCOPE OF THE PROGRAM:
• Load and clean supermarket sales data (2012-2021).
• Use **time** (month/year) as a feature for regression.
• Develop a linear regression model to predict future pollution levels.
• Evaluate the model's performance using metrics such as **mean squared error (MSE)**.

CODE:

```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
import numpy as np

# Load the dataset
df = pd.read_csv(r'C:\Users\leela\OneDrive\Desktop\supermarket_sales - Sheet1.csv')

# Step 1: Convert 'Date' to datetime
df['Date'] = pd.to_datetime(df['Date'])

# Step 2: Aggregate total sales by date
daily_sales = df.groupby('Date')['Total'].sum().reset_index()

daily_sales['Date_ordinal'] = daily_sales['Date'].map(pd.Timestamp.toordinal)

X = daily_sales['Date_ordinal'].values.reshape(-1, 1)
y = daily_sales['Total'].values

model = LinearRegression()
model.fit(X, y)
```

```
daily_sales['Predicted'] = model.predict(X)

plt.figure(figsize=(14, 6))
plt.plot(daily_sales['Date'], daily_sales['Total'], label='Actual Sales', alpha=0.6)
plt.plot(daily_sales['Date'], daily_sales['Predicted'], label='Linear Regression Forecast', color='red')
plt.title('Linear Regression Forecasting of Sales')
plt.xlabel('Date')
plt.ylabel('Total Sales')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

OUTPUT:
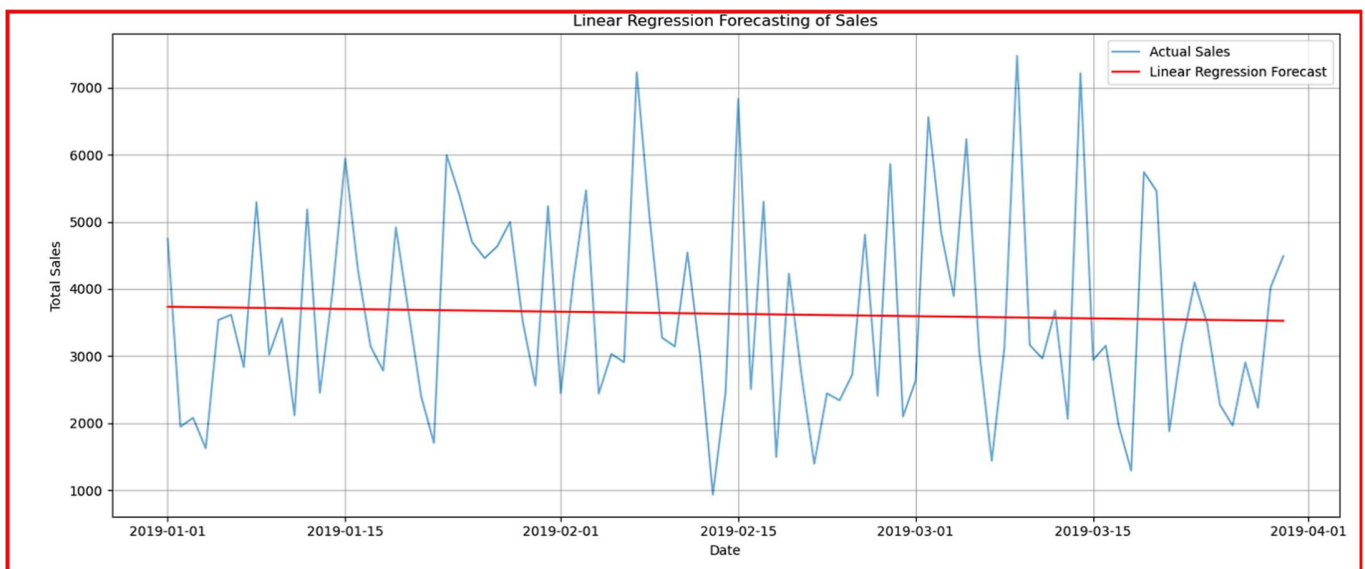
RESULT:

Thus, the program using the time series data implementation has been done successfully.