**Develop a linear regression model for forecasting time series data.**

**EX:No.5**

**DATE:**

# 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 air pollution data (2012-2021).
* Use **time** (month/year) as a feature for regression.
* Build a **linear regression model** for predicting future pollution levels.
* Evaluate the model performance with metrics like **mean squared error (MSE)**.

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_squared\_error

df = pd.read\_csv("C:\\Users\\exam\\Downloads\\archive (13)\\cinemaTicket\_Ref.csv")

df['date'] = pd.to\_datetime(df['date'])

daily\_df = df.groupby('date')['total\_sales'].sum().reset\_index()

daily\_df = daily\_df.sort\_values('date')

daily\_df['Time'] = (daily\_df['date'] - daily\_df['date'].min()).dt.days

X = daily\_df[['Time']]

y = daily\_df['total\_sales']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, shuffle=False, test\_size=0.2)

model = LinearRegression()

model.fit(X\_train, y\_train)

y\_pred = model.predict(X\_test)

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

plt.plot(daily\_df['date'].iloc[-len(y\_test):], y\_test, label='Actual')

plt.plot(daily\_df['date'].iloc[-len(y\_test):], y\_pred, label='Predicted', linestyle='--')

plt.xlabel('Date')

plt.ylabel('Total Sales')

plt.title('Cinema Sales Forecast - Linear Regression')

plt.legend()

plt.grid(True)

plt.tight\_layout()

plt.show()

print("Mean Squared Error:", mean\_squared\_error(y\_test, y\_pred))

# OUTPUT:

# 

**RESULT:**

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