**Exp no: 3 Develop a linear regression model for forecasting time series data.**

**Date: 18/2/25**

**Aim:**

The aim of this analysis is to apply linear regression to model the trend in the Air Passengers dataset.

**Objectives:**

The primary objective of this analysis is to apply linear regression to model the trend in the Air Passengers dataset and evaluate its effectiveness in capturing the underlying growth pattern. Additionally, the analysis aims to visualize the original data alongside the fitted trend to assess the model's performance. A secondary objective is to decompose the time series into its trend, seasonal, and residual components to gain deeper insights into the dataset's structure. This analysis will provide a foundation for more advanced time series forecasting techniques and contribute to understanding the dataset's behavior over time.

**Background/Scope:**

The "Air Passengers" dataset is a well-known time series dataset that records the monthly totals of international airline passengers from 1949 to 1960. It is widely used in time series analysis to study trends, seasonality, and forecasting. The dataset exhibits a clear upward trend and strong seasonal patterns, making it an ideal candidate for exploring linear regression in the context of time series. Linear regression can be applied to model the trend component, which helps in understanding the underlying growth pattern of passenger numbers over time. This analysis is particularly useful for forecasting future passenger counts and making data-driven decisions in the aviation industry.

**Steps for Time Series Sales Data Preprocessing:**

**Step 1: Load the Dataset**

Load the dataset from a local CSV file and display the first few rows to understand its structure.

import pandas as pd

# Load the dataset

df = pd.read\_csv('/content/AirPassengers.csv')

# Display the first few rows of the dataset

print(df.head())

**Step 2: Preprocess the Data**

Convert the Month column to a datetime object and set it as the index for time series analysis.

# Convert the 'Month' column to datetime

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

# Set 'Month' as the index

df.set\_index('Month', inplace=True)

# Display the first few rows of the dataset

print(df.head())

**Step 3: Visualize the Time Series Data**

Plot the time series data to visualize the trend of airline passengers over time.

import matplotlib.pyplot as plt

# Plot the time series

plt.figure(figsize=(10, 6))

plt.plot(df.index, df['#Passengers'], label='Passengers')

plt.title('Monthly International Airline Passengers (1949-1960)')

plt.xlabel('Date')

plt.ylabel('Number of Passengers')

plt.legend()

plt.show()

**Step 4: Perform Linear regression**

We can use linear regression to model the trend in the time series. We’ll use the LinearRegression class from sklearn.linear\_model.

from sklearn.linear\_model import LinearRegression

import numpy as np

# Create a time variable

df['Time'] = np.arange(len(df))

# Define the independent variable (Time) and dependent variable (Passengers)

X = df[['Time']]

y = df['#Passengers']

# Fit the linear regression model

model = LinearRegression()

model.fit(X, y)

# Predict the trend

df['Trend'] = model.predict(X)

# original data vs fitted trend

plt.figure(figsize=(10, 6))

plt.plot(df.index, df['#Passengers'], label='Passengers')

plt.plot(df.index, df['Trend'], label='Trend', linestyle='--')

plt.title('Monthly International Airline Passengers with Linear Trend')

plt.xlabel('Date')

plt.ylabel('Number of Passengers')

plt.legend()

plt.show()

**Step 5: Interpret the Results**

Analyze the plots to understand the trend, seasonality, and residuals in the dataset. The trend component shows the long-term increase in passengers, the seasonal component reveals periodic fluctuations, and the residuals represent the noise or irregularity in the data.

**Output :**





**Result:**

Thus linear regression has been implemented for time series dataset successfully.