**EX:No.1 221501016**

**21/01/25**

**Program to implement time series data for import library, load data, Pre-processing and visualising**

**Aim :** To implement time series data for importing, loading , preprocessing and visualising data.

**Implementation :**

import pandas as pd

import matplotlib.pyplot as plt

**Loading the dataset**

file\_path = '/mnt/data/sales-of-shampoo-over-a-three-ye.csv'

data = pd.read\_csv(file\_path)

data.columns = ['Month', 'Sales']

**Cleaning the dataset**

# Convert 'Month' to a datetime object

data['Month'] = pd.to\_datetime(data['Month'], format='%d-%b', errors='coerce')

# Drop rows with invalid or missing dates

data = data.dropna(subset=['Month'])

**Preprocessing the dataset**

# Sort the data by date

data = data.sort\_values('Month')

# Reset the index after sorting

data = data.reset\_index(drop=True)

# Handle missing values in 'Sales' (if any)

data['Sales'] = data['Sales'].fillna(data['Sales'].mean())

# Save the cleaned data for future use

data.to\_csv('/mnt/data/cleaned\_sales\_data.csv', index=False)

print("Data loading, cleaning, and preprocessing completed.")

**Visualizing the dataset**

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

plt.plot(data['Month'], data['Sales'], marker='o', linestyle='-', color='b', label='Sales')

plt.title('Sales of Shampoo Over Time', fontsize=16)

plt.xlabel('Month', fontsize=12)

plt.ylabel('Sales', fontsize=12)

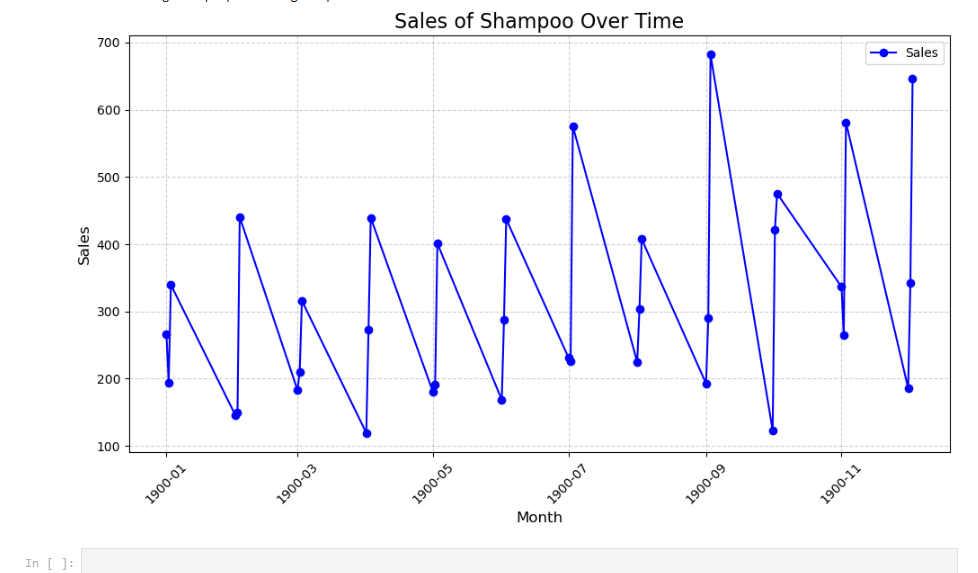
plt.grid(True, linestyle='--', alpha=0.6)

plt.legend()

plt.xticks(rotation=45, fontsize=10)

plt.tight\_layout()

plt.show()

**Output :**

**Result:** Thus importing, loading, pre-processing and visualising time series data has been successful.