**Implement Programs For Time Series Data Cleaning, Loading, And Handling Time Series Data And Pre-Processing Techniques**

**EX.No:1 DATE: 25/01/2**

**AIM:**

To clean, preprocess, and visualize electricity production data, focusing on trend analysis and handling missing values.

**ALGORITHM:**

1. Import libraries and load dataset from a CSV file.
2. Preview data and handle missing values by filling price with the median.
3. Convert month to datetime format and set all years to 2022.
4. Sort data by month and remove outliers using the Z-score method.
5. Set month as index and display summary statistics of cleaned data.
6. Plot shampoo sales over time using a line chart and show cleaned data.

**CODE:**

import pandas as pd

import matplotlib.pyplot as plt

from scipy import stats

file\_path = 'C:\\Users\\Lenovo\\Downloads\\1.csv'

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

print("Initial Data Preview:")

print(data.head())

missing\_values = data.isnull().sum()

print("\nMissing values:")

print(missing\_values)

data['price'].fillna(data['price'].median(), inplace=True)

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

data['month'] = data['month'].apply(lambda x: x.replace(year=2022))

data.sort\_values('month', inplace=True)

z\_scores = stats.zscore(data['price'])

outliers = (z\_scores > 3) | (z\_scores < -3)

data = data[~outliers]

data.set\_index('month', inplace=True)

print("\nSummary Statistics of Cleaned Data:")

print(data.describe())

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

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

plt.title('Shampoo Sales Over 3 Years')

plt.xlabel('Month')

plt.ylabel('Price')

plt.legend()

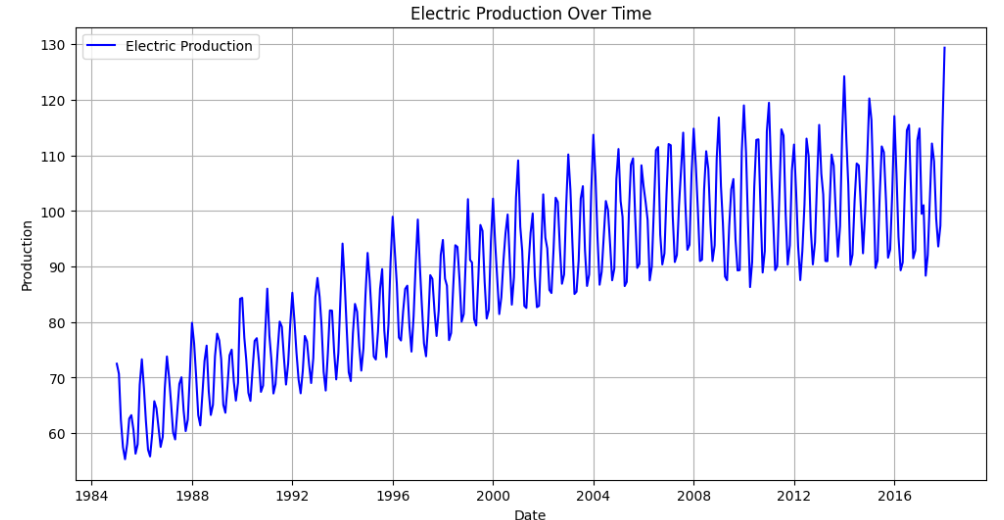
plt.grid(True)

plt.show()

print("\nCleaned Data Preview:")

print(data.head())

**OUTPUT:**

****

**RESULT:**

Thus the program has been completed and verified successfully.