| **EXP:1**  **23/01/2025** | **DATA PRE-PROCESSING TECHNIQUES** |
| --- | --- |

**AIM:**

To implement a program for time series data cleaning,loading and handling time series data & preprocessing techniques.

**PROCEDURE:**

**1) import the libraries**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**2) Load the dataset**

**file\_path = "monthly-beer.csv"**

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

**3) Check the first few rows**

**print(data.head(25))**

**4) Convert the date column (if exists) to a datetime type**

**data['Month'] = pd.to\_datetime(data['Month']) # Adjust column name if different**

**5) Calculate Q1 (25th percentile) and Q3 (75th percentile) for the 'Monthly beer production' column**

**Q1 = data['Monthly beer production'].quantile(0.25)**

**Q3 = data['Monthly beer production'].quantile(0.75)**

**6) Calculate IQR (Interquartile Range)**

**IQR = Q3 - Q1**

**7) Determine the lower and upper bounds**

**lower\_bound = Q1 - 1.5 \* IQR**

**upper\_bound = Q3 + 1.5 \* IQR**

**8) Identify outliers**

**outliers = data[(data['Monthly beer production'] < lower\_bound) | (data['Monthly beer production'] > upper\_bound)]**

**9) Print outliers if any**

**if not outliers.empty:**

**print("Outliers detected:")**

**print(outliers)**

**else:**

**print("No outliers detected.")**

**10) Plot the data with outliers highlighted**

**plt.figure(figsize=(25, 4))**

**plt.plot(data['Month'], data['Monthly beer production'], label='Beer Production')**

**plt.scatter(outliers['Month'], outliers['Monthly beer production'], color='red', label='Outliers', zorder=5)**

**plt.title('Monthly Beer Production with Outliers')**

**plt.xlabel('Date')**

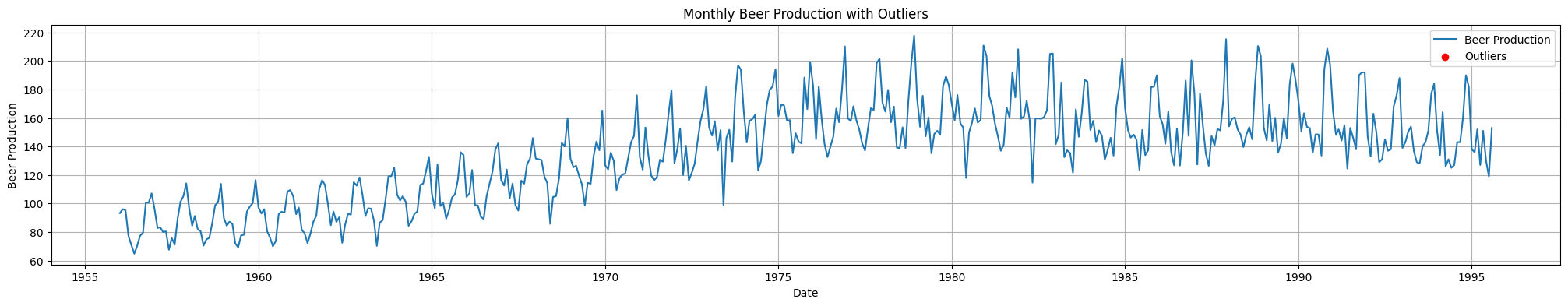
**plt.ylabel('Beer Production')**

**plt.legend()**

**plt.grid()**

**plt.show()**

**OUTPUT:**

****

**RESULT:**

Thus the program has been executed successfully.