EX:No.1 221501046

**Program to implement time series data for import library, load data,Preprocessing and visualising**

**Aim:**

Write a program to implement time series data for import library, load data, Preprocessing and visualising.

**Algorithm:**

1. Import Libraries: Import pandas, numpy, matplotlib, seaborn, MinMaxScaler, and Colab's file upload module.
2. Upload & Read Data: Upload AirPassengers.csv and read it using pd.read_csv().
3. Rename Columns: Rename columns for easier reference.
4. Datetime Conversion: Convert the date column to datetime format.
5. Index Setting: Set the date column as the DataFrame index.
6. Set Frequency: Define the frequency as monthly for consistent time series handling.

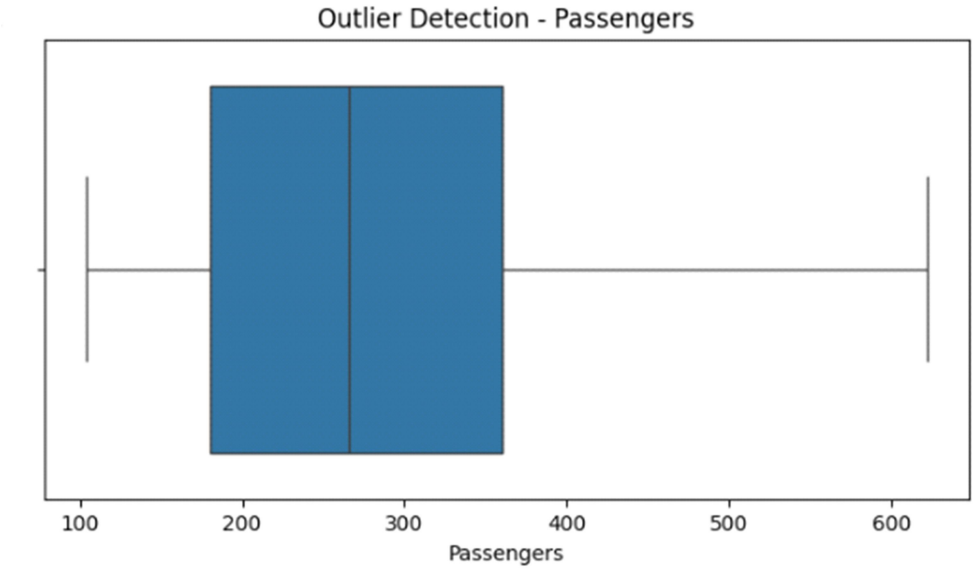
**Code:**

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from google.colab import files

from sklearn.preprocessing import MinMaxScaler

uploaded = files.upload()

df = pd.read_csv('AirPassengers.csv')

df.columns = ['Month', 'Passengers']

df['Month'] = pd.to_datetime(df['Month'])
```

```
df.set_index('Month', inplace=True)

df = df.asfreq('MS')

print("\nMissing Values:\n", df.isnull().sum())

df.fillna(method='ffill', inplace=True)

duplicates = df.duplicated().sum()

print(f"\nDuplicate rows: {duplicates}")

print("\nData types:\n", df.dtypes)

print("\nSummary Statistics:")

print(df.describe())

plt.figure(figsize=(8, 4))

sns.boxplot(x=df['Passengers'])

plt.title("Outlier Detection - Passengers")

plt.show()

scaler = MinMaxScaler()

df['Normalized_Passengers'] = scaler.fit_transform(df[['Passengers']])

print("\nFinal Cleaned Data Sample:")

print(df.head())
```

**Output:**

```
Duplicate rows: 26

Data types:
 Passengers    int64
dtype: object

Summary Statistics:
       Passengers
count  144.000000
mean   280.298611
std    119.966317
min    104.000000
25%    180.000000
50%    265.500000
75%    360.500000
max    622.000000
```




```
Final Cleaned Data Sample:
            Passengers  Normalized_Passengers
Month
1949-01-01         112               0.015444
1949-02-01         118               0.027027
1949-03-01         132               0.054054
1949-04-01         129               0.048263
1949-05-01         121               0.032819
```

**Result:**

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