**Branch: - Computer Science and Engineering Class: - III Year**

**Subject: - Big Data Analytics Lab Sem: - VI**

**Teacher Manual**

**PRACTICAL NO. 7**

**Aim:** Write a program to implement Time Series Analysis.

**Software Requirement**: Jupyter

**Theory:**

**Time series analysis:**

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.

What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data.

Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms. Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow’s weather report to future years of climate change.

**Program:**

|  |
| --- |
| import pandas as pd import numpy as np  import matplotlib.pyplot as plt |

|  |
| --- |
| df = pd.read\_csv('AirPassengers.csv') |

|  |
| --- |
| df.head() *# read the first five rows* |

|  |
| --- |
| df.isnull().sum() *# return the number of missing values in dataset* |

|  |
| --- |
| df.dtypes *# look at the data types for each column* |

|  |
| --- |
| df['Month']= pd.to\_datetime(df['Month']) *#method helps to convert string Date time into Python Date time object.*  df.head()*# read the first five rows* |

* After conversion again check data types of each column

|  |
| --- |
| df.dtypes *# look at the data types for each column* |

|  |
| --- |
| df['Month'].dt.year.head() *#use Series.dt.year attribute to return the year of the datetime in the underlying data of the given Series object here we use df[‘Month’] as Series object.* |

|  |
| --- |
| df= df.set\_index('Month') *#* *To set the Month column as the index*  df.head() *# read the first five rows* |

|  |
| --- |
| df.plot(grid='on') *# to plot DataFrame into grid format* |

|  |
| --- |
| from datetime import datetime *#* *Import the datetime module*  *# datetime() class requires three parameters to create a date: year, month, day.*  start\_date= datetime(1959,1,1) *#here we create object as start\_date*  end\_date= datetime(1960,12,1) *#here we create object as send\_date*  df[(start\_date <= df.index) & (df.index <= end\_date)].plot(grid='on') *# slice any range of dates.* |

**Result:**