ITF22 – Time Series Data Analysis & Preprocessing Assignment 5.1

1. Objective

This assignment focuses on working with sensor-based time series data. You will find an existing publicly available dataset, load it into a Python environment, preprocess the data, and describe it.

2. Dataset Selection

You should find a time series dataset from a public source (e.g., Kaggle, UCI Repository, Hugging Face Datasets, or other open sources). The dataset should come from sensors such as:

- Temperature or humidity sensors
- Environmental or air-quality sensors
- Motion sensors (accelerometer/gyroscope)
- Physiological or biometric sensors
- Weather or climate monitoring systems

3. Dataset Description

You must provide a short description of your dataset, either in your notebook or a separate document. Include:

- **Dataset source** (where you got it)
- **Type of sensor** (what measurements it records)
- **Collected variables** (the columns or measurements in the data, e.g., temperature, heart rate, x/y/z acceleration)
- **Number of samples** (how many rows of data)

4. Data Loading and Preprocessing

When working with time series data, you should perform the following steps:

1. **Parse timestamps correctly**
   Convert the time-related column (usually a string) into a proper datetime format. This allows Python to understand the order of events and perform time-based operations.
2. **Sort the data in chronological order**
   Arrange the dataset so the earliest timestamp comes first. Many time series calculations require data to be in the correct order.
3. **Handle missing values in a reasonable way**
   Fill in or deal with gaps in the data. You can use methods like forward fill, backward fill, interpolation, or remove rows with missing values. Handling missing data ensures that calculations and statistical analyses are accurate.

4. **Perform basic cleaning if needed**
   Fix simple issues in the dataset, such as removing duplicate rows, correcting impossible values, or renaming columns for clarity. This makes the dataset consistent and ready for analysis.

Include a short explanation of why each preprocessing step is important.

5. Submission Requirements

You must submit:

- A notebook containing your code and outputs
- A short description (in the notebook or separate document) covering:
  - Dataset description
  - Preprocessing steps