Time Series Analysis - Assignment 1: Introduction

Objective: The objective of this assignment is to introduce students to time series data by exploring a dataset, decomposing it into its components, and handling missing values using different strategies.

Requirements:

1. **Dataset Visualization:**
   - Load the time series dataset.
   - Ensure the dataset has a date column that is parsed correctly as a datetime index.
   - Plot the time series data using appropriate visualization techniques (line chart, scatter plot, etc).
   - Add labels, titles, and a legend to the chart.
2. **Statistical Analysis:**
   - Compute and report statistical values of the time series (ex mean, variance, min and max values)
3. **Introduce Missing Values:** ^0°%
   - Randomly remove a subset of data points from the dataset to simulate missing values.
   - Visualize the dataset after introducing missing values to observe the impact.
4. **Handling Missing Values:**
   - Implement backward/forward fill, moving average and linear interpolation strategies to handle missing values.
   - Apply each strategy separately and visualize the results.
5. **Time Series Decomposition:**
   - Decompose the time series into its components: trend, seasonality, and residuais.
   - Use the additive and multiplicative decomposition models
   - Display the decomposed components in separate subplots. Which is the most appropriate and why?
6. **Creation of new basic synthetic time series**
   - After time series decomposition, generate new residual component
   - Recompose the signal using the same Trend, Seasonality and the new residual component
   - Plot the new time series. How does it look compared to the original one?
7. **Implementation Details:**
   - You can use Python and the following libraries: `pandas`, `numpy`, `matplotlib`, `seaborn`, and `statsmodels`.
   - Ensure the notebook is well-structured with proper markdown explanations.
   - Include inline comments for clarity in the code.

Deliverables:

- A Jupyter Notebook (`.ipynb`) containing:
    - Dataset visualization
    - Time series decomposition
    - Handling missing values with both strategies
    - New synthetic TS
    - Clear explanations (use markdown components for explanations) and visualizations

Evaluation Criteria:

- Correctness and clarity of the implementation
- Quality of visualizations and explanations
- Proper handling requirements
- Code readability and explanations

Deadline:
- Submit your notebook on moodle until 16 march 23:59 for full grade
- Allowed delay: 23 march 23:59 with 2 points penalisation

Datasets:

You can use one of the following datasets for your assignment:
- https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data
- https://www.kaggle.com/datasets/kingki19/semarang-daily-climate-data-2020-2023