HW2 (Basic analytics using Numpy & Pandas)

Please note that you are only allowed to use general "Python", "Numpy", and "Pandas" commands (as covered in lectures 1 - 8) to complete tasks 1 – 10. The use other libraries is not permitted.

Also, keep in mind that optimizing for fewer resources (such as memory and CPU usage) and shorter computation times will earn you some bonus points.

Collaborate with your previous groups. No assignments will be accepted after the due dates.

Ensure your .ipynb file contains adequate comments using markdown cells. Refrain from using AI-driven programming tools. With practice, you can outperform AI !

Part A: Due Date 2023.11.07 (Midnight)

Step 1: Data Loading

Download your data from https://polybox.ethz.ch/index.php/s/x5HZ7weUnuZwgDf.

Step 2: Basic analytics

Perform the following analytics on each Series (and any plausible pair of Series, where necessary).

1- Data description: Enumerate all statistics (minimum, maximum, quartiles,…) for every series;
2- Data cleansing: Investigate missing values and fill gaps in the series using suitable methods;
3- Outliers detection: Identify and report the count and respective values of outliers in each series;
4- Normality test: Analyze and present the degree of normality in the series, supporting your findings with quantitative metrics;
5- Log-transformation: Implement log-transformation and compare the modified data with the original set;
6- Standardization: standardize all series and contrast the standardized series mutually to extract insights regarding the interrelation of all variables;
7- Correlation and covariance: obtain linear and monotonic correlation and covariance values between all plausible pairs of series;
8- Autocorrelation and cross-correlation: ascertain the auto- and cross-correlation measurements for all series and their pairs.

Step 3: Advanced analytics

9- Stationary: discuss the stationary level of the series with an appropriate approach;
10- Trend: discuss the trend of the series with an appropriate approach;
11- Seasonality: discuss the seasonality of the series with an appropriate approach; and
12- anomaly detection: detect the anomalies of the series with an appropriate approach.

Part B: Due Date 2023.11.14 (Midnight)

Revisit HW1, this time using Pandas and Numpy., Compare the results (in terms of accuracy, execution time, resource consumption, etc.) with your original solutions.