Lab 5
Time Series Profiling

GOAL

To explore the time series in order to identify the data characteristics and the required preparation tasks to apply next.
To prepare the data according to the issues identified in the exploration task

WORK TO BE DONE

To explore the data, following the five perspectives:

- Dimensionality;
- Granularity;
- Distribution;
- Seasonality;
- Stationarity.

To prepare the data, following the needs identified before, with respect to:

- Scaling.
- Aggregation.
- Differentiation.
- Smoothing.

Per each preparation task:

1. Pick the dataset and split it into two datasets: one for training and another for testing.
2. For each preparation task, choose three different parameterizations and apply them to the training dataset. Then apply the adequate transformation to the testing set, when necessary.
3. Train a Persistence and a Linear Regression model for each resulting dataset.
4. Compare the performance of the resulting models.
5. Select the dataset that yields the best performance, comparing MSE, MAE and R2.
6. Pick this dataset to proceed to the next preparation task.

REPORT TO DELIVER

PDF file only with charts and tables. No analysis or justification is needed at this point.
Suggested charts per dataset:

Profiling:

Dimensionality:

Original time series.

Granularity:

Two additional granularities

Distribution:

Component studies for different granularities.

Autocorrelation lag-plots at different granularities.

Autocorrelation correlograms at different granularities.

Stationarity:

Augmented Dickey-Fuller test result for the different granularities.

Preparation:

Persistence and linear regression models' predictions over the testing dataset.

Performance of Persistence and Linear Regression models for each approach applied.

Identification of the best dataset and corresponding approach (best approach).

Good work!!!