



TIME SERIES PREPROCESSING



– Hunter



QUIZ

X What is seasonal decomposition



TOPICS

- X Impute Missing Values
- X Impute Outliers
- X Time Series Decomposition



IMPUTE MISSING DATA

IMPUTE MISSING DATA



Why are Missing Values Special in Time Series Data?

- X Human errors, i.e., forgot input, input wrong values.
- X System errors, i.e., service failure, database error.

Solutions:

- Deletion: eliminating rows that contain missing values
- Constant Imputation: substitutes all missing values with a constant
- Last Observation Carried Forward (LOCF) and Next Observation Carried Backward (NOCB)
- Mean/Median/Mode Imputation
- Rolling Statistics Imputation: substitutes missing values with a rolling statistic over a specified window period
- Linear Interpolation: replaced based on a linear equation derived from the available values
- Spline Interpolation: replaced based on a non-linear equation
- K-Nearest Neighbors (KNN) Imputation: computationally expensive
- STL Decomposition for Time Series: breaks down the time series into trend, seasonality, and residuals, then imputes missing values in the residuals before reassembling the components



2.

IMPUTE OUTLIERS

DETECT OUTLIERS



- X Just plot it
- X Box plots
- X The Z-score
- X Inter-Quartile Range
- X Anomaly detection

IMPUTE OUTLIERS



X Treat outlier as a missing value and apply missing value imputation methods.



3.

TIME SERIES DECOMPOSITION

TIME SERIES DECOMPOSITION



Trend, Seasonality, Residual

Level: The average value in the series.

Trend: The increasing or decreasing value in the series.

Seasonality: The repeating short-term cycle in the series.

Noise: The random variation in the series.

Additive Model: $y(t) = \text{Level} + \text{Trend} + \text{Seasonality} + \text{Noise}$

Multiplicative Model: $y(t) = \text{Level} * \text{Trend} * \text{Seasonality} * \text{Noise}$



RECOMMENDED READINGS

<https://medium.com/@aaabulkhair/data-imputation-demystified-time-series-data-69bc9c798cb7>

<https://baotramduong.medium.com/dealing-with-outliers-imputation-096f0b60e348>

<https://medium.com/@ba13026/cleaning-outliers-in-conductance-timeseries-from-molecular-dynamics-5ebce9f312f4>

<https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>

https://medium.com/@tushar_aggarwal/master-the-power-of-seasonal-decomposition-of-time-series-stl-unveiling-the-essence-of-time-26c19a910314

<https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data>