

Overview

Missing data

Contents



What is missing data?



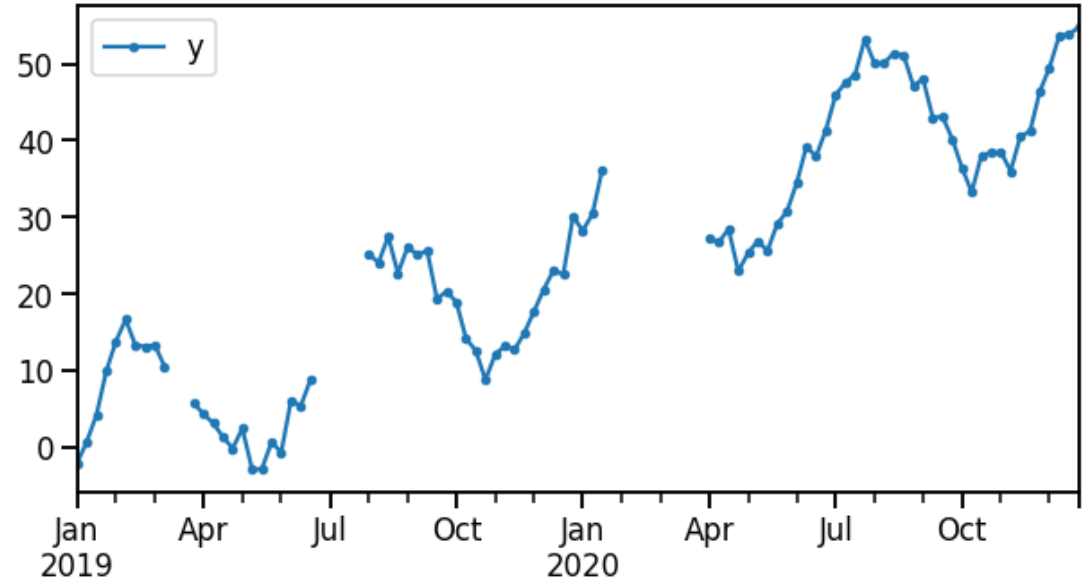
Why is missing data a problem for forecasting?



How to handle missing data

What is missing data?

Date	Sales
2020-01-01	3
2020-01-02	10
2020-01-03	23
2020-01-04	nan
2020-01-05	nan
2020-01-06	nan
2020-01-07	58
2020-01-08	5

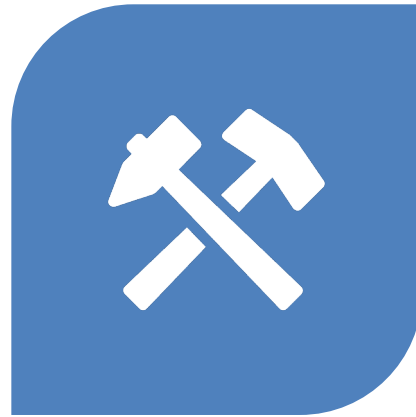


- **Missing data is the lack of values at certain time points**
- Missing at random (e.g., sensor malfunction, clerical error)
- Missing not at random (e.g., public holiday)

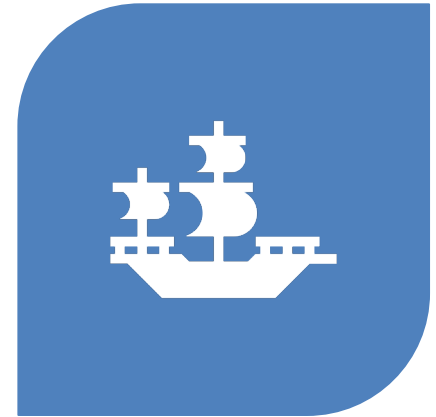
Why is missing data a problem?



MODELLING



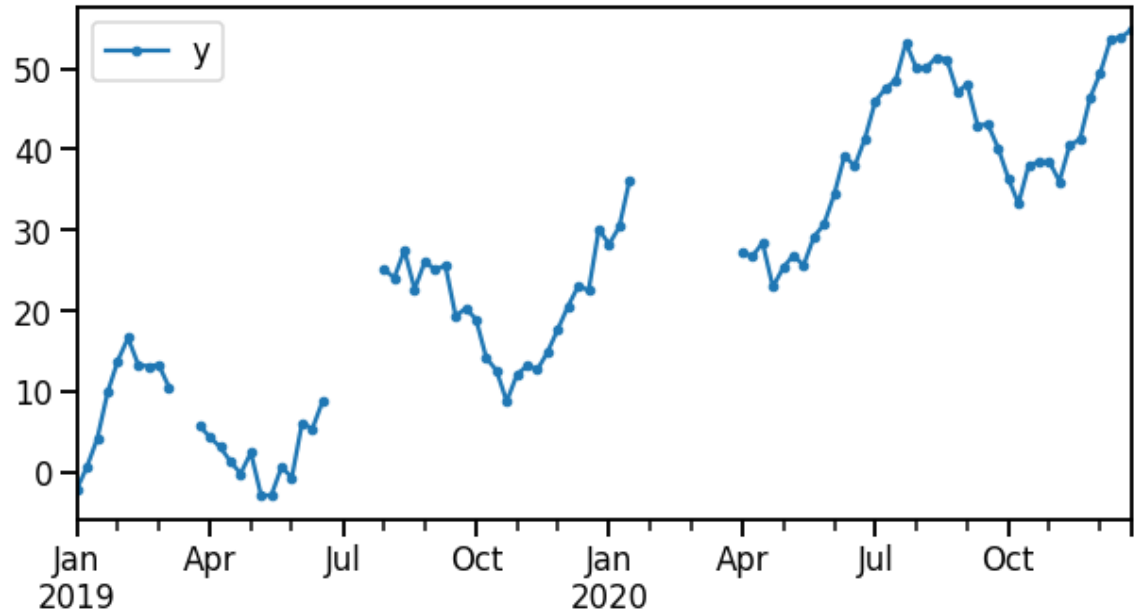
FEATURE ENGINEERING



EXPLORATORY DATA
ANALYSIS

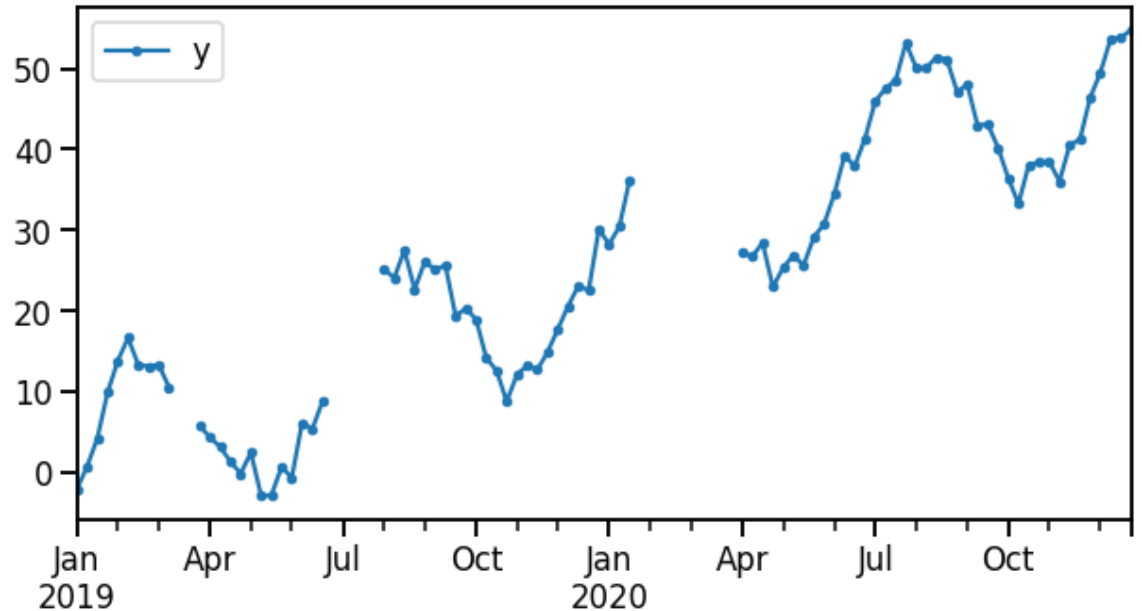
Missing data impacts modelling

- **When training:** some models cannot handle missing values in the training data
- **When forecasting:** some models require values from the past as an input, these values can't be missing



Missing data impacts feature engineering

- Many time series feature engineering methods can't handle missing data
- Example: Lagged value of the target. Like the sales this time last year $\text{lag_sales}_t = \text{sales}_{t-12}$
- If the value of the target 12 months ago is missing then the feature is also missing



Missing data impacts EDA

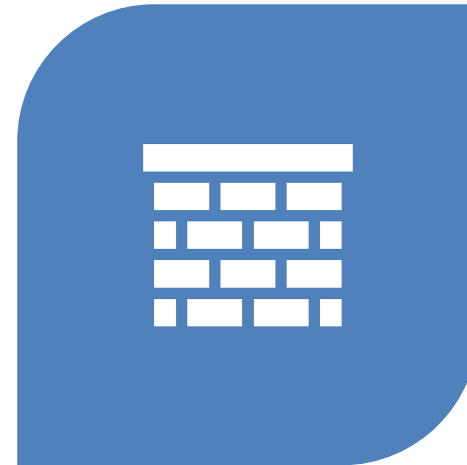
- Need to choose how to handle missing data when computing summary statistics
- Example: Compute the mean sales by week.
- Missing data means we should think about:
 - If data is missing not at random this can bias the analysis
 - If data was incidentally missing on high sales days then we underestimate the value

Date	Page views	Sales
2020-01-01	54	3
2020-01-02	67	10
2020-01-03	89	23
2020-01-04	134	nan
2020-01-05	1000	nan
2020-01-06	100	nan
2020-01-07	130	58
2020-01-08	60	5

Solutions



IMPUTE MISSING DATA



USE FORECASTING METHODS
ROBUST TO MISSING DATA

Imputation methods for time series

1. Forward filling (aka last observation carried forward)
2. Backward filling (aka next observation carried backwards)
3. Linear interpolation
4. Spline interpolation
5. Seasonal decomposition and interpolation