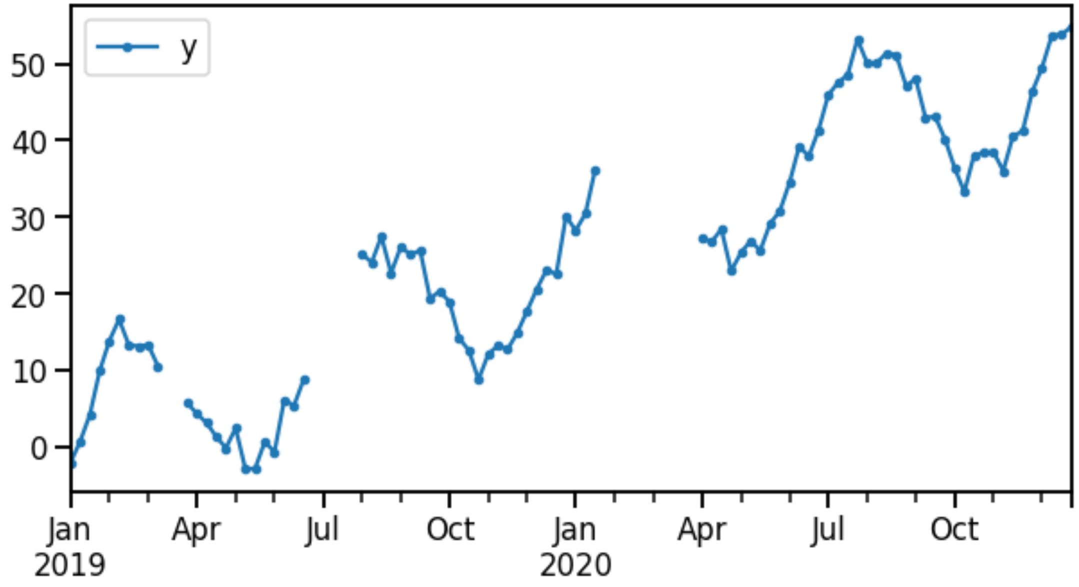
Summary

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

Solutions

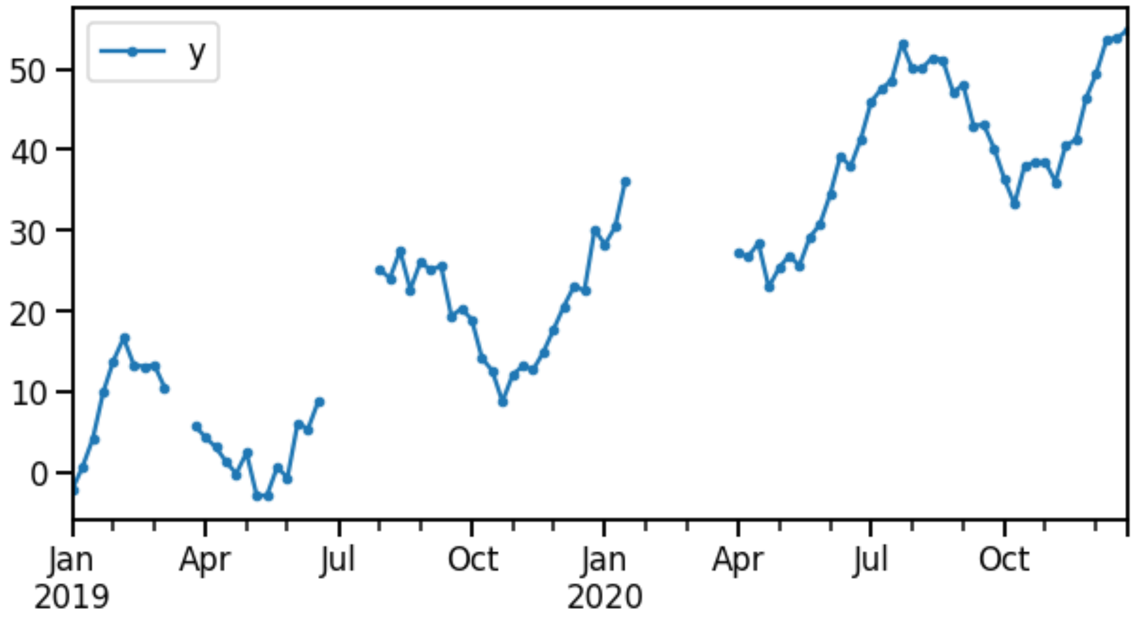
IMPUTE MISSING DATA

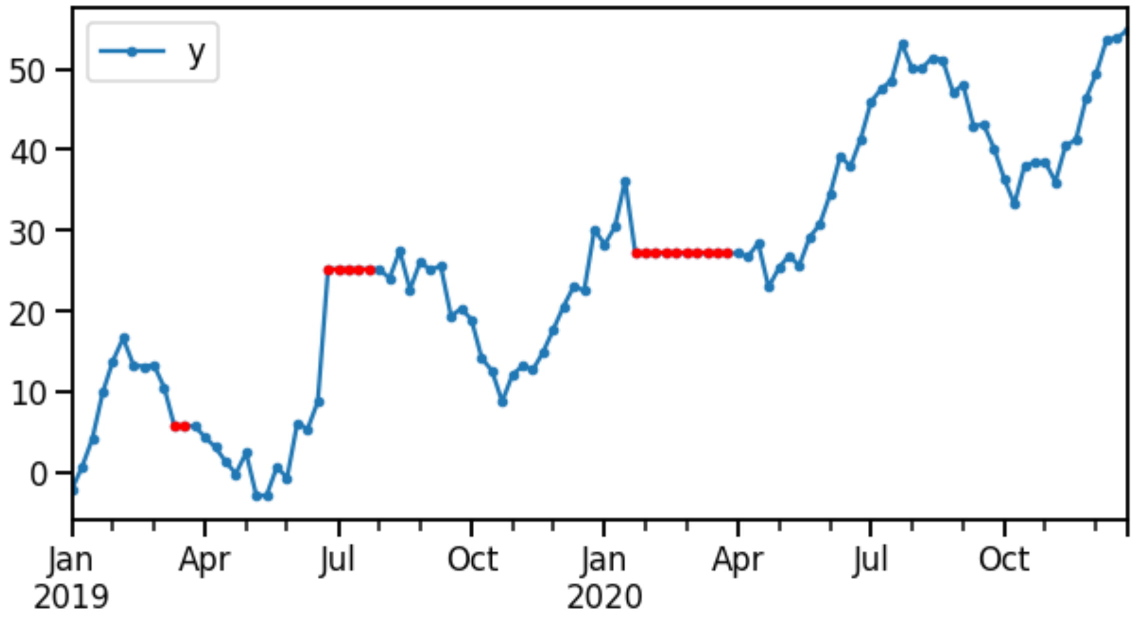
USE FORECASTING METHODS ROBUST TO MISSING DATA

Imputation methods for time series

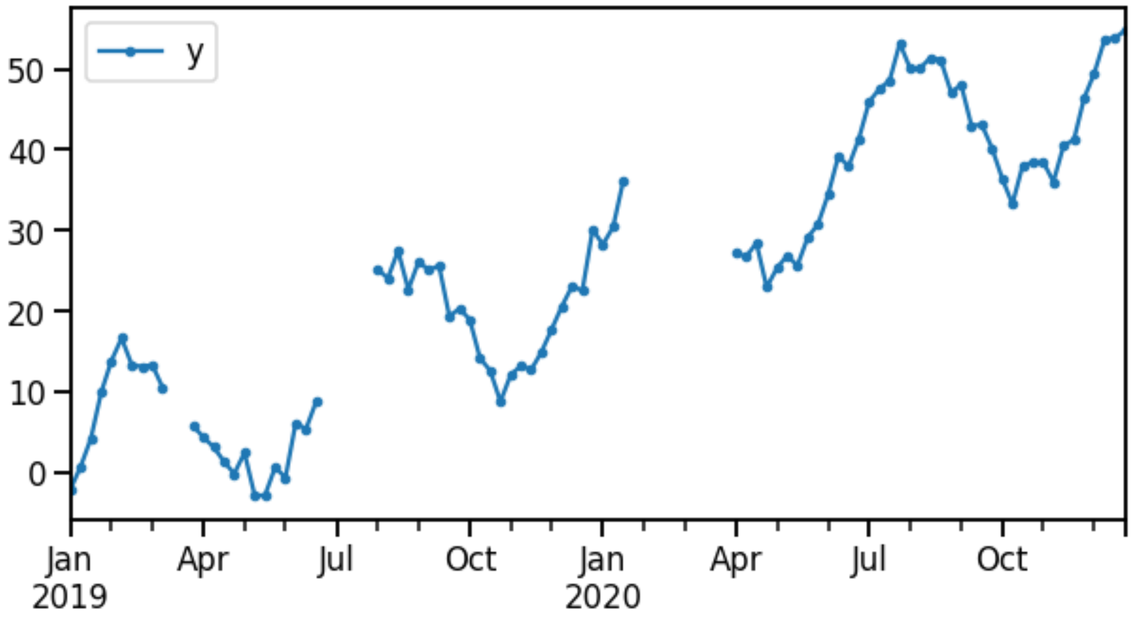
1. Forward filling (aka last observation carried forward)
2. Backward filling (aka next observation carried backwards)
3. Linear interpolation
4. Spline interpolation
5. Seasonal decomposition and interpolation

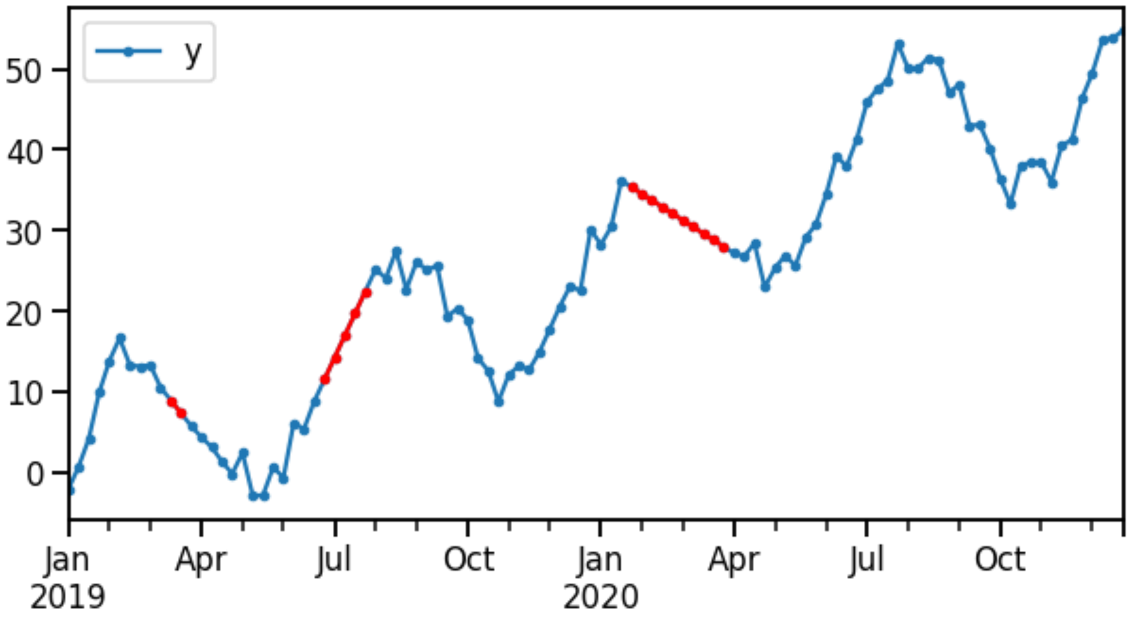
Which method to use depends on the time series and the size of the gaps



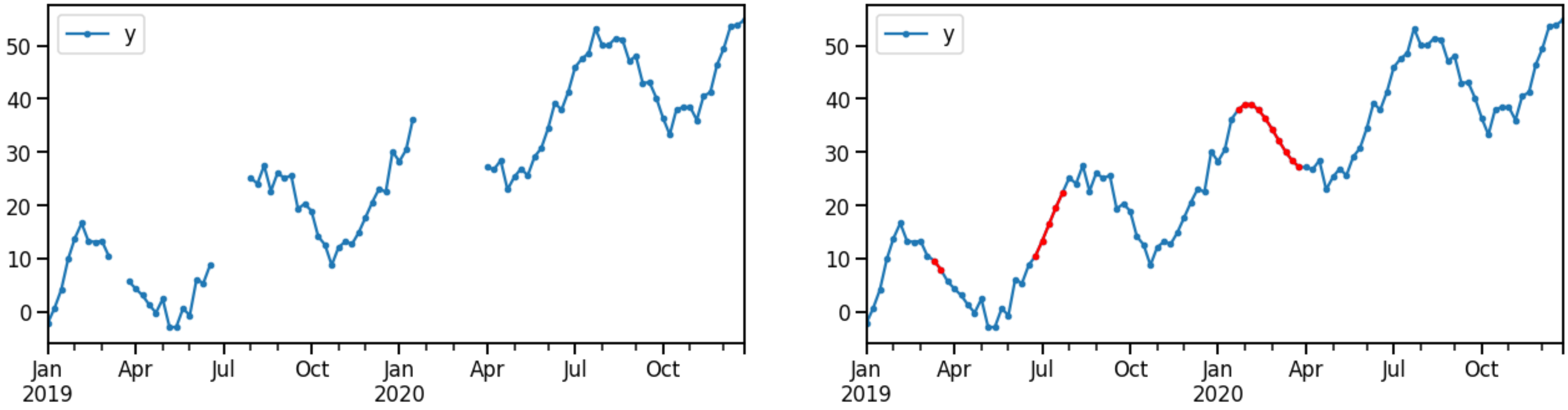


Which method to use depends on the time series and the size of the gaps

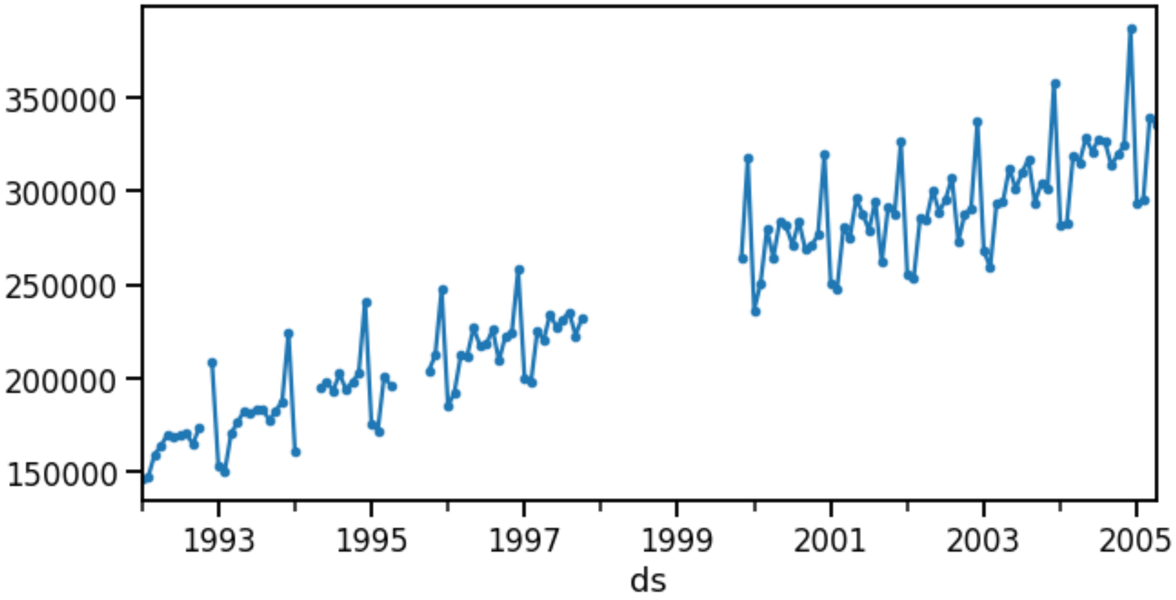


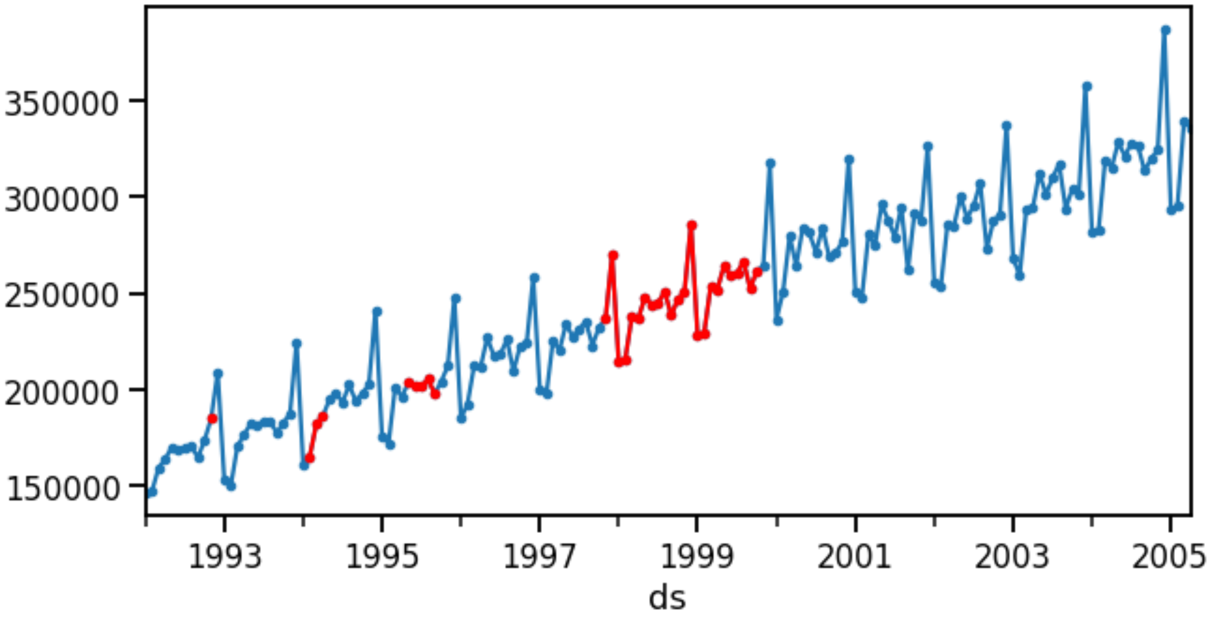


Which method to use depends on the time series and the size of the gaps



Which method to use depends on the time series and the size of the gaps





Practical tips

- Consider how the method will distort the time series
  - Does the method distort seasonality or long term trends?
  - Does the method create artificial jumps in the data?
- Small gaps: forward fill or linear interpolation
- Larger gaps: consider structure of time series
  - No trend or seasonality: Forward fill, linear interpolation
  - Strong trend and no seasonality: linear interpolation
  - Strong seasonality: Seasonal decomposition and interpolation
- Sense check time series plots after interpolation

Methods shown here are for time series: can be feature or target

Date	y	temperature	marketing
2015-01-01	9	26	0
2015-01-02			
2015-01-03	18	23	1
2015-01-04	27	26	0
2015-01-05	15	25	0
2015-01-06	7	24	0

- Features which take discrete values need to be handled carefully so that they are not imputed with nonsensical values