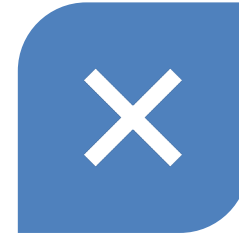# Outliers in time series data

Contents

OUTLIERS IN TIME SERIES

WHY ARE OUTLIERS A PROBLEM FOR FORECASTING?
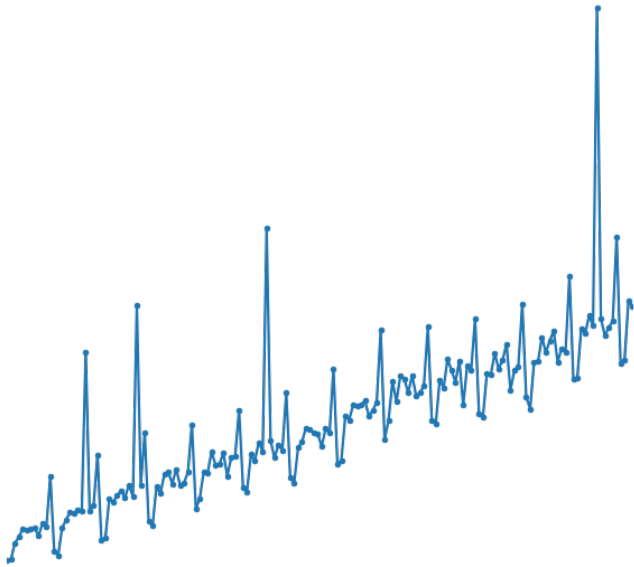
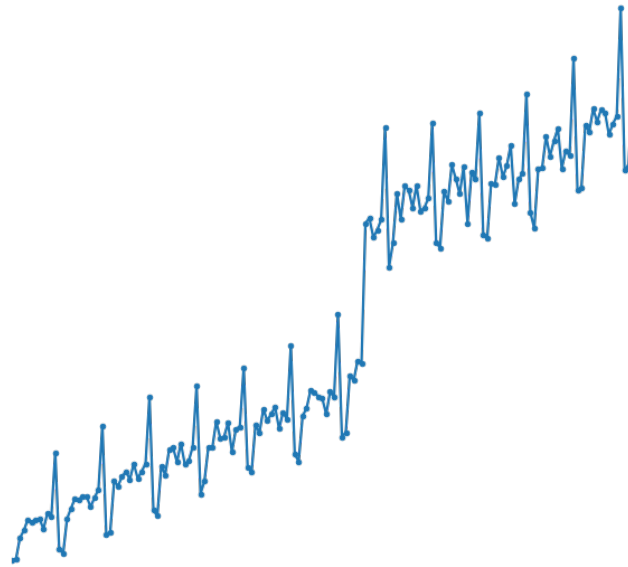METHODS TO IDENTIFY OUTLIERS

METHODS TO HANDLE OUTLIERS

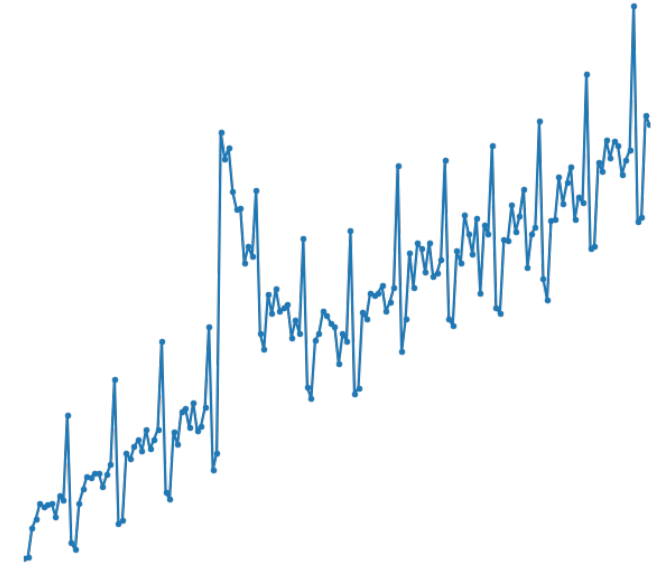PRACTICAL TIPS ON APPLYING THESE METHODS

Outliers in time series data

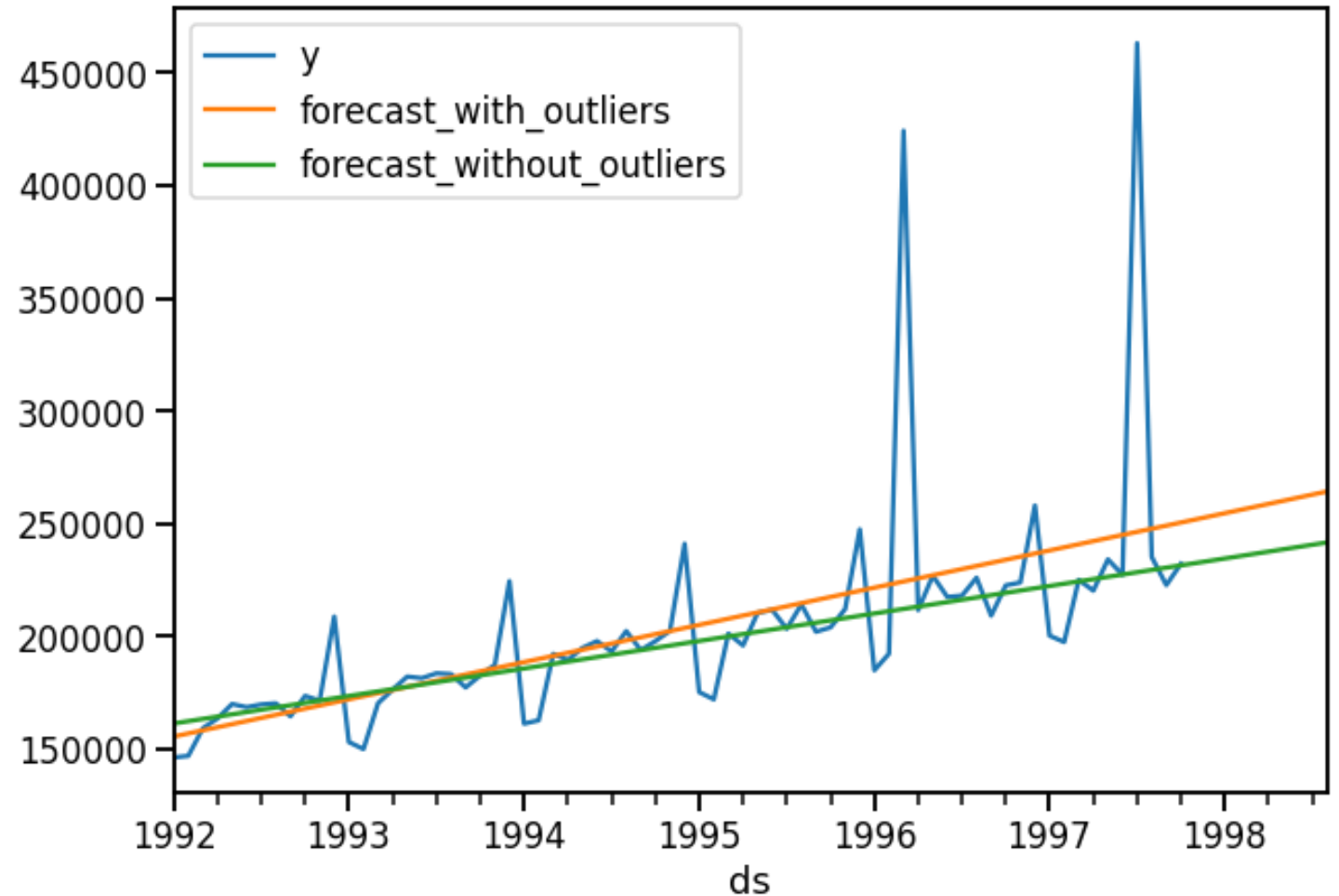Point outliers

Level shift outlier

Transient shift outlier

Subsequence outlier *

* Blázquez-García, Ane, et al. "A review on outlier/anomaly detection in time series data." arXiv preprint arXiv:2002.04236 (2020)

Why are outliers a problem?

Outliers can bias models which result in worse forecasts

Outlier identification methods

VISUALLY INSPECT
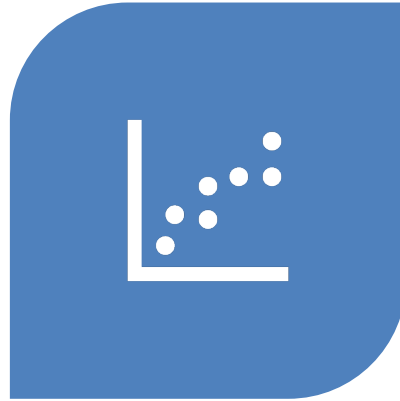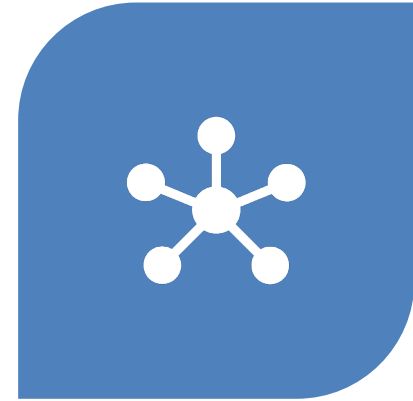
ESTIMATION METHODS

DENSITY BASED METHODS

Outlier identification methods

- Plot the time series and visually inspect
- Estimation methods
- Density based methods

Outlier identification methods

- Plot the time series and visually inspect
- Estimation methods
- Density based methods

Is the actual value significantly different from an expected value?

$$|y_t - \hat{y}_t| > \delta$$

Outlier identification methods

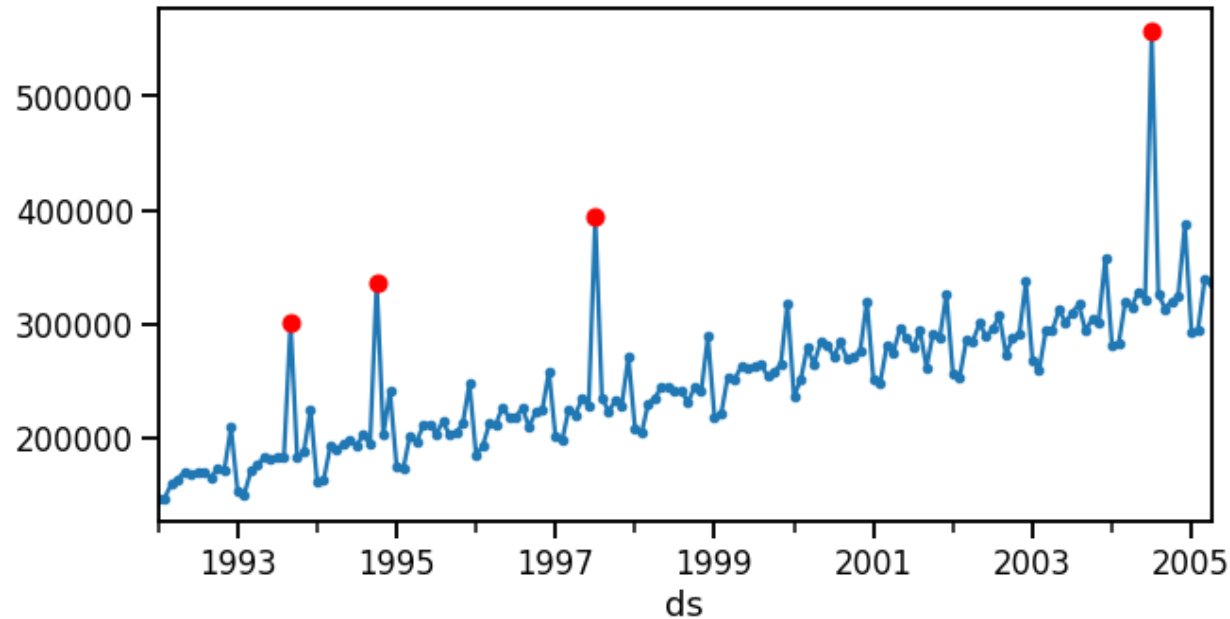- Plot the time series and visually inspect
- Estimation methods
- Density based methods

Does an observation have a small number of neighbouring data points?

Methods for handling outliers

1. Model the outlier (e.g., add a feature such as is_public_holiday)
2. Treat outlier as missing data and impute the value

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	70
2020-02-15	30
2020-02-16	25

Methods for handling outliers

1. Model the outlier (e.g., add a feature such as is_public_holiday)
2. Treat outlier as missing data and impute the value

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	70
2020-02-15	30
2020-02-16	25

Methods for handling outliers

1. **Model the outlier (e.g., add a feature such as is_public_holiday)**
2. Treat outlier as missing data and impute the value

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	70
2020-02-15	30
2020-02-16	25

Methods for handling outliers

1. **Model the outlier (e.g., add a feature such as is_public_holiday)**
2. Treat outlier as missing data and impute the value

Date	Sales	is_valentines _day
2020-02-12	23	0
2020-02-13	30	0
2020-02-14	70	1
2020-02-15	30	0
2020-02-16	25	0

Methods for handling outliers

1. Model the outlier (e.g., add a feature such as is_public_holiday)

2. **Treat outlier as missing data and impute the value**

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	70
2020-02-15	30
2020-02-16	25

Methods for handling outliers

1. Model the outlier (e.g., add a feature such as is_public_holiday)

2. **Treat outlier as missing data and impute the value**

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	NaN
2020-02-15	30
2020-02-16	25

Methods for handling outliers

1. Model the outlier (e.g., add a feature such as is_public_holiday)

2. **Treat outlier as missing data and impute the value**

Date	Sales
2020-02-12	23
2020-02-13	30
2020-02-14	30
2020-02-15	30
2020-02-16	25

Practical tips

Plot data first and understand nature of outliers

If outlier is meaningful try to model it through a feature

If outlier is an error then consider imputing it

Summary

Outliers are data points that are very different to rest of the data

Outliers can bias models

Outliers can be automatically identified using various methods

Outliers should be handled depending on their cause