

# Session10-Pandas(Handling with Outliers)

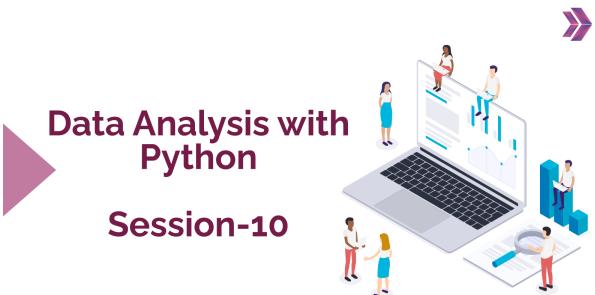
Pear Deck Session  
Training Clarusway  
Pear Deck - December 12, 2022 at 7:50PM

## Part 1 - Summary

Use this space to summarize your thoughts on the lesson

## Part 2 - Responses

Slide 1



The slide features a large purple play button icon on the left. To its right, the text "Data Analysis with Python" is written in a large, bold, maroon font, with "Session-10" in a slightly smaller font below it. To the right of the text is a 3D illustration of a laptop screen displaying various charts and graphs. Several small human figures are interacting with the laptop and a bar chart, suggesting a collaborative data analysis environment. A small CLARUSWAY logo is visible at the bottom left of the slide area.

Use this space to take notes:

## Slide 2



Use this space to take notes:

## Slide 3

### ▶ Table of Contents

- ▶ What is the Outliers?
- ▶ Detecting Outliers
- ▶ Handling with Outliers
- ▶ Some Useful Methods



3

Use this space to take notes:

## Slide 4

## Your Response

I've completed the pre-class content?

**True** **False**

Pear Deck

Students choose an option

Pear Deck Interactive Slide  
Do not remove this bar

Use this space to take notes:

You Chose

- **False**

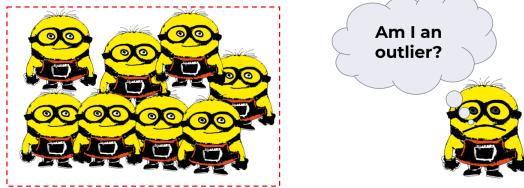
Other Choices

- True

## Slide 5

### ► What is the Outlier?

- ▶ Outliers can be unusually and extremely different from most of the data points existing in our sample.



Use this space to take notes:

## Slide 6

### ► What is the Outlier?

- ▶ Outliers can create biased results while calculating the stats of the data due to its extreme nature, thereby affecting further statistical/ML models.

Index	car_price
1	22.000
2	24.000
3	1050
4	28.000
5	149.000

The abnormal values of given variable (**car\_price**)

Such values are called **outliers**

Use this space to take notes:

## Slide 7

### ► What is the Outlier?

#### Causes of Outliers

- ▶ Data entries errors
- ▶ Measurement errors or instrument errors
- ▶ Sampling errors
- ▶ Data processing errors
- ▶ Natural novelties in data

Use this space to take notes:

## Slide 8

### ► What is the Outlier?

#### Types of Outliers

##### Univariate Outliers

- ▶ generally referred to as extreme points on a variable

##### Multivariate Outliers

- ▶ generally combination of unusual data points for **two or more variables**

An assumption of many multivariate statistical analysis, such as Multiple linear regression, is that there are no multivariate outliers.



Link(s) on this slide:

- [https://en.wikiversity.org/wiki/Multivariate\\_statistics](https://en.wikiversity.org/wiki/Multivariate_statistics)

Use this space to take notes:

## Slide 9

### ► Detecting Outliers

#### Methods for Detecting Outliers

##### Graphs

- ▶ Scatter plot
- ▶ Box plot
- ▶ Histogram

##### InterQuartile range (IQR) technique

##### Statistical Tests

- ▶ Grubbs' test
- ▶ Chi-square test
- ▶ Dixon's Q test

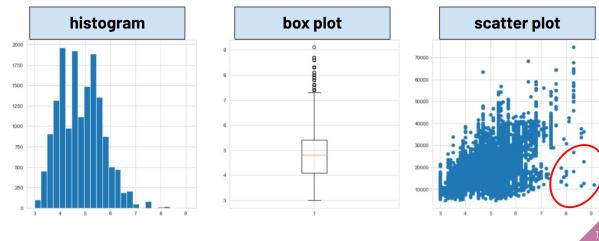


Use this space to take notes:

## Slide 10

### ▶ Detecting Outliers

#### Graphs

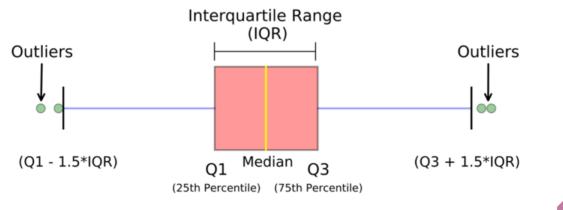


Use this space to take notes:

## Slide 11

### ▶ Detecting Outliers

#### InterQuartile range (IQR) technique



Use this space to take notes:

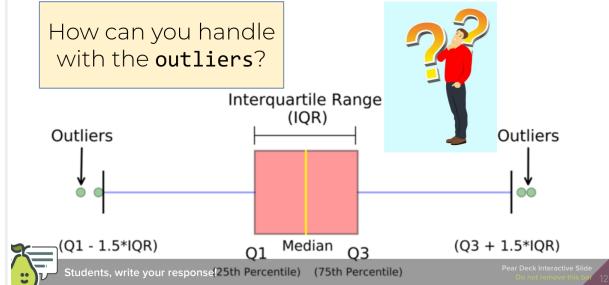
## Slide 12

## Your Response

## Slide 12

## Your Response

### ► Handling with Outliers



Use this space to take notes:

## Slide 13

### ► Handling with Outliers

#### Methods for Handling Outliers

- ▶ Removing the outliers.
- ▶ Limitation the outliers. (winsorize)
- ▶ Data transformation. (log, square root, exponentiating)
- ▶ Replacing the outliers. (mean, median, mode)
- ▶ Using different analysis methods. (statistical/nonparametric tests)
- ▶ Valuing the outliers. (valid reason for the outlier to exist)

Use this space to take notes:

## Slide 14

### ► Handling with Outliers

#### Guideline for Handling Outliers



If the outlier in question is:

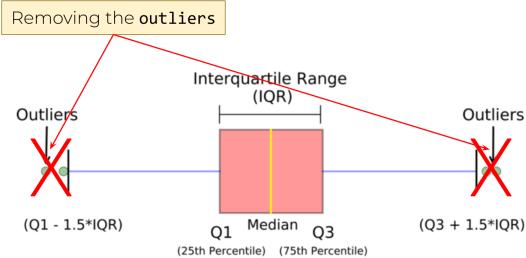
- ▶ A measurement error or data entry error, correct the error if possible. If you can't fix it, remove that observation because you know it's incorrect.
- ▶ Not a part of the population you are studying (i.e., unusual properties or conditions), you can legitimately remove the outlier.
- ▶ A natural part of the population you are studying, you should not remove it.

14

Use this space to take notes:

## Slide 15

### ► Handling with Outliers

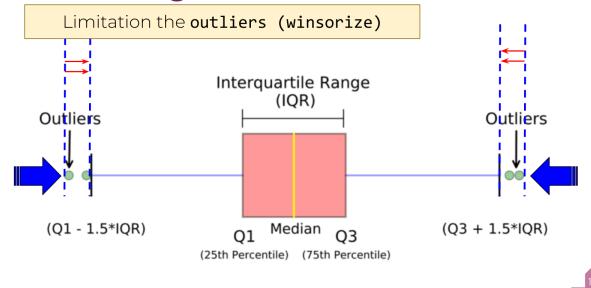


15

Use this space to take notes:

## Slide 16

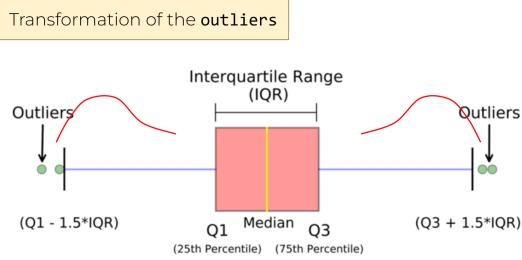
### ▶ Handling with Outliers



Use this space to take notes:

## Slide 17

### ▶ Handling with Outliers



Use this space to take notes:

## Slide 18

## Your Response

## Slide 18

## Your Response

**Make connections**  
What are the advantages and disadvantages of dropping & limiting the outliers?

Students, write your response!

Do a research and find your answers.

Peer Deck Interactive Slide  
Do not remove this bar

Use this space to take notes:

## Slide 19

### ► Some Useful Methods



- quantile()
- winsorize()
- log()



Use this space to take notes:

## Slide 20

### Data Analysis with Python



let's start the  
hands-on phase

20

Use this space to take notes:

## Slide 21

### Your Response

Did you find this lesson interesting and challenging?



Use this space to take notes:

# THANKS!

Any questions?

You can find us at:  
► [#questions-answers@Slack](mailto:#questions-answers@Slack)



22

Use this space to take notes: