

Session7-Pandas(Missing Values)

DAwithPython S-7
Training Clarusway
Pear Deck - January 27, 2022 at 7:38PM

Part 1 - Summary

Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1



Use this space to take notes:

Slide 2



Use this space to take notes:

Slide 3

▶ Table of Contents



- ▶ What is Missing Value?
- ▶ Types of Missing Values
- ▶ Handling with Missing Values
- ▶ Some Useful Methods



Use this space to take notes:

Slide 4

Your Response

You Chose

I've completed the pre-class content?

True

False

Students choose an option

Pear Deck Interactive Slide
Do not remove this bar

- **True**

Other Choices

- False

Use this space to take notes:

Slide 5

► What is Missing Value? ►

- ▶ Missing data is common in many different areas of data science and machine learning.
- ▶ Unfortunately, it can be challenging to handle effectively, and often there is no best solution.



5

Use this space to take notes:

Slide 6

Your Response

You Chose

- **Not a Number**

Other Choices

- Not any Value

► What is Missing Value?

What does the **NaN** stand for?



- Nothing Available
- No Available Number

Students choose an option

Peer Deck Interactive Slide
Do not remove this slide

Use this space to take notes:

Slide 7

► What is Missing Value?

- ▶ Used Audi A-3 Car prices (£) in UK

index	car_price
1	22.000
2	24.000
3	NaN
4	28.000
5	NaN

No value (**car_price**)
• **NaN** :Not a Number

Such values are called
missing values

Use this space to take notes:

Slide 8

► What is Missing Value?



- ▶ Missing data occurs because of variety of reasons, including:
 - Manual data entry techniques,
 - Equipment faults,
 - Wrong measurements.

8

Use this space to take notes:

Slide 9

► Types of Missing Values?



- ▶ Missing completely at random (MCAR)
- ▶ Missing at random (MAR)
- ▶ Missing not at random (MNAR)
- ▶ Structurally missing

9

Use this space to take notes:

Slide 10

► Types of Missing Values ➤

- ▶ Missing completely at random (MCAR)
 - Follow no discernable pattern
 - Cannot be predicted from the remaining known variables
 - Example: data generated explicitly at random or survey data using a random subset of questions from a pre-defined list.

10

Use this space to take notes:

Slide 11

► Types of Missing Values ➤

- ▶ Missing at random (MAR)
 - Errors with recording the data correctly
 - Can roughly be interpolated from the remaining values to a reasonable degree of accuracy.
 - Example: A sensor that misses a particular minute's measurement

11

Use this space to take notes:

Slide 12

► Types of Missing Values ➤

- ▶ Missing not at random (MNAR)
 - Why the data is missing is known
 - Can not effectively be inferred or predicted
 - Example: people in a certain age/income bracket refuse to answer how many vehicles or houses they own

12

Use this space to take notes:

Slide 13

► Types of Missing Values ➤

- ▶ Structurally missing
 - The missing data is missing for an apparent reason.
 - Mechanism that caused the missing data is easily inferred
 - Example: a survey that asks for income from employment would have missing values for those who do not have a job

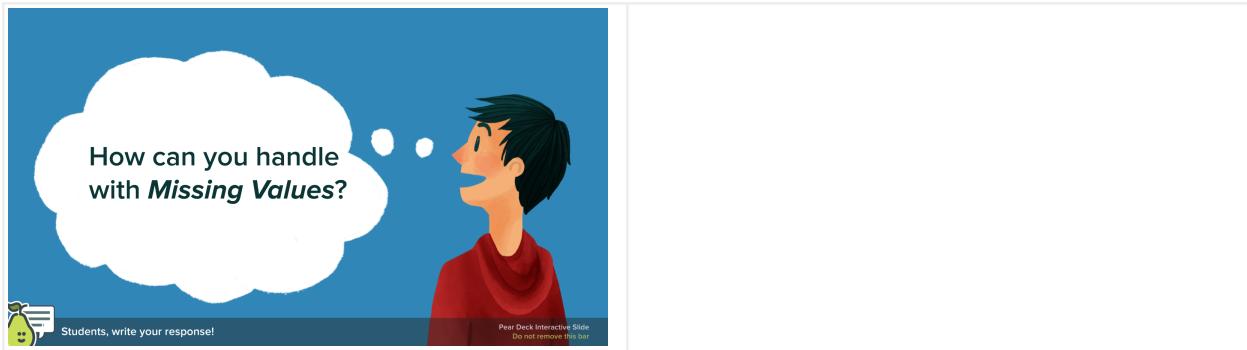
13

Use this space to take notes:

Slide 14

Your Response

Answer 1:
fill replace



Use this space to take notes:

Slide 15

► Handling with Missing Values ➤

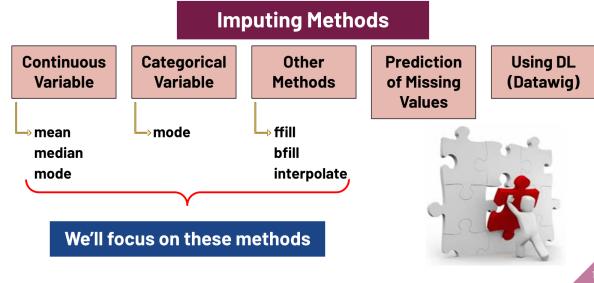
- ▶ Remove the missing data instances. (This method should be acceptable if there are few missing values and you have a lot of data.)
- ▶ Imputation methods. (This is a common approach it allows most models to function as usual without any modifications.)
- ▶ Keep the missing values and use a model which incorporates them. (This method limits the models available.)

15

Use this space to take notes:

Slide 16

► Handling with Missing Values ➞



Use this space to take notes:

Slide 17

► Handling with Missing Values ➞

There are several methods for handling with Missing Values.

index	car_price	model
1	22.000	2012
2	18.000	2005
3	NaN	2005
4	28.000	2012
5	NaN	2012

We should consider the group (**model**) of the missing values (**prices**)

17

Use this space to take notes:

Slide 18

► Handling with Missing Values ➞

The most important point when handling with missing value

Domain
Knowledge

18

Use this space to take notes:

Slide 19

► Some Useful Methods ➞

- isnull()
- isna()
- notnull()
- notna()
- drop()
- dropna()
- any()
- all()
- fillna()
- where()
- map()
- replace()
- interpolate()

19

Use this space to take notes:

Slide 20

Your Response

Draw lines to match the attributes/methods to their definitions:

<code>df.unique()</code>	Used for imputing in a missing data.
<code>df.dropna()</code>	Sort by the values along either axis.
<code>df.value_counts()</code>	Return unique values of Series object.
	Return object with labels on given axis omitted where alternately any or all of the data are missing.
	Count distinct observations over requested axis.
	Apply a function along an axis of the DataFrame.
<code>df.apply()</code>	
<code>df.sort_values()</code>	
<code>df.fillna()</code>	

 Students, draw anywhere on this slide!

Pear Deck Interactive Slide
Do not remove this bar

 Students, draw anywhere on this slide!

Pear Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 21

Data Analysis with Python



let's start the
hands-on phase

21

Use this space to take notes:

Slide 22

Your Response

Did you find this lesson interesting and challenging?

Too hard Just right Too easy

Students, drag the icon!

Pear Deck
Peer Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 23

THANKS!

Any questions?

You can find us at:

steve_w@clarusway.com
michael_g@clarusway.com



23

Use this space to take notes: