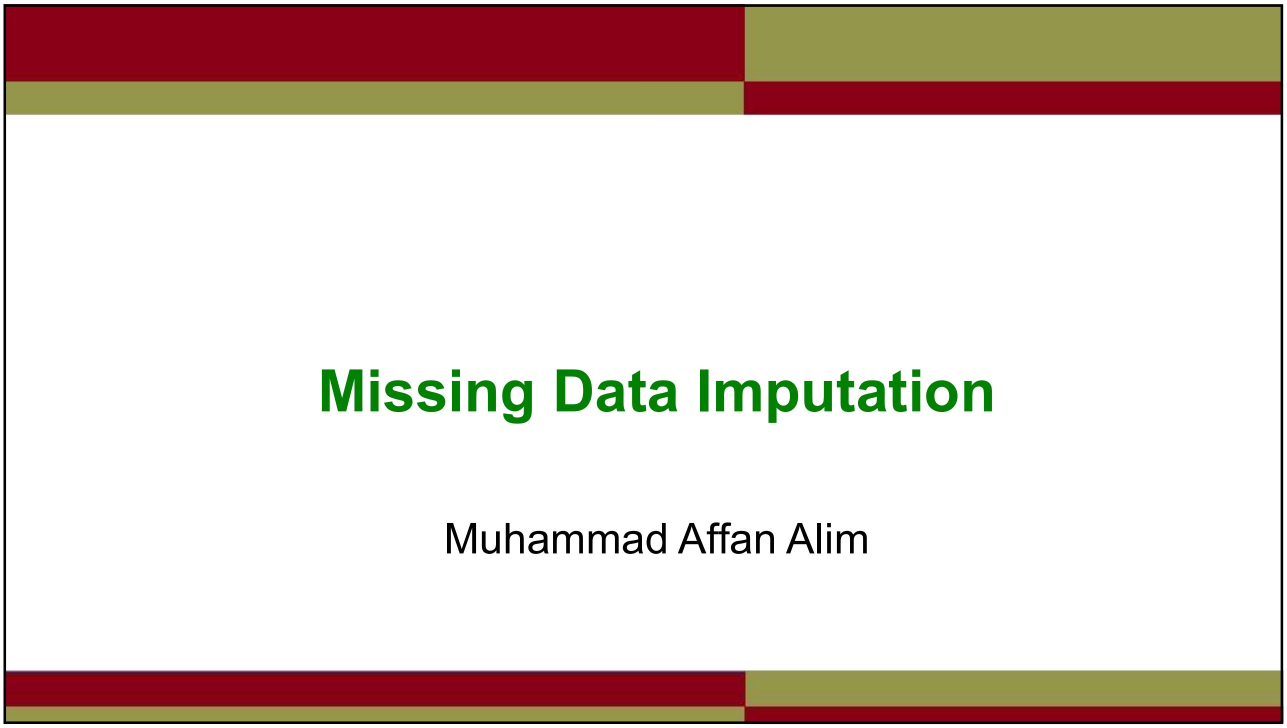
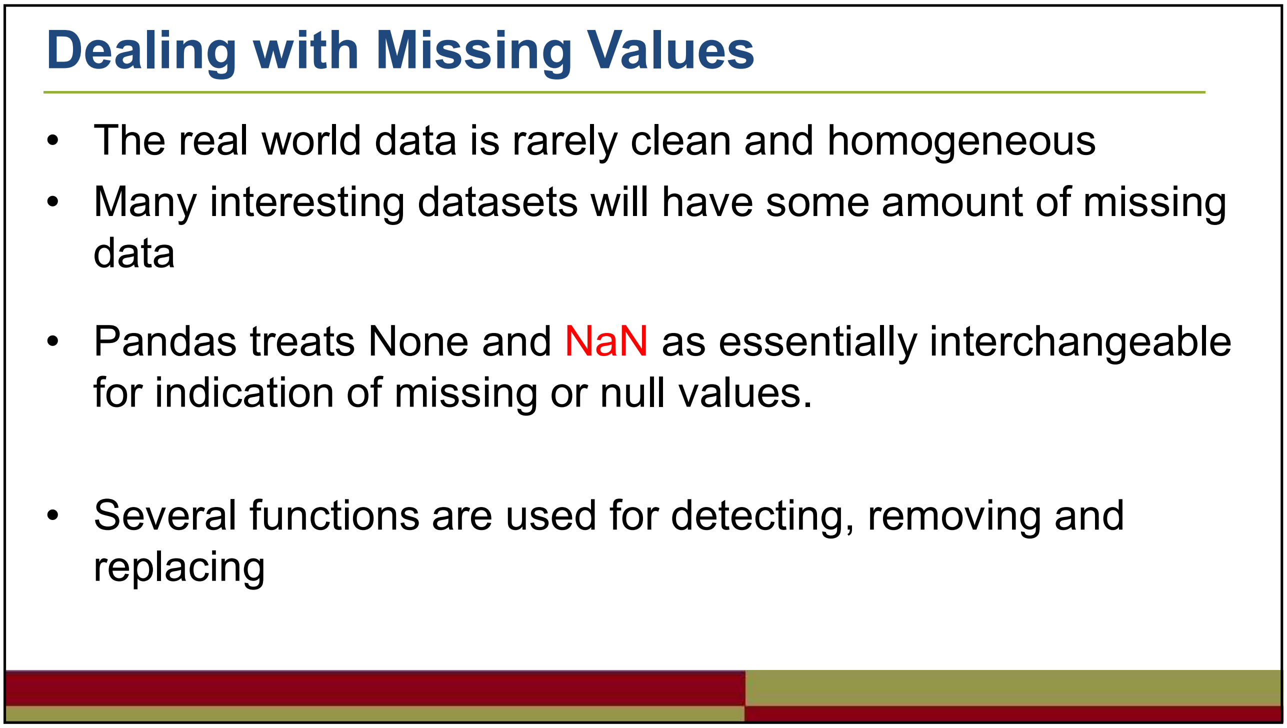


Missing Data Imputation

Muhammad Affan Alim



Dealing with Missing Values

- The real world data is rarely clean and homogeneous
- Many interesting datasets will have some amount of missing data
- Pandas treats None and NaN as essentially interchangeable for indication of missing or null values.
- Several functions are used for detecting, removing and replacing

Dealing with Missing Values

- In Pandas missing data is represented by three values:

1. None: None is a Python singleton object that is often used for missing data in Python code.

2. NaN : NaN (an acronym for Not a Number), is a special floating-point value recognized by all systems that use the standard IEEE floating-point representation

3. Na: Not available

Dealing with Missing Values

- Pandas treat None, NaN and Na as essentially interchangeable for indicating missing or null values.
- To facilitate this convention, there are several useful functions for detecting, removing, and replacing null values in Pandas DataFrame :

  isnull(), notnull(), dropna(),

  fillna(), replace(), interpolate()

Dealing with Missing Values

- In this article we are using CSV file, employee.csv
- Checking for missing values using isnull() and notnull()
- In order to check missing values in Pandas DataFrame, we use a function isnull() and notnull().

Dealing with Missing Values

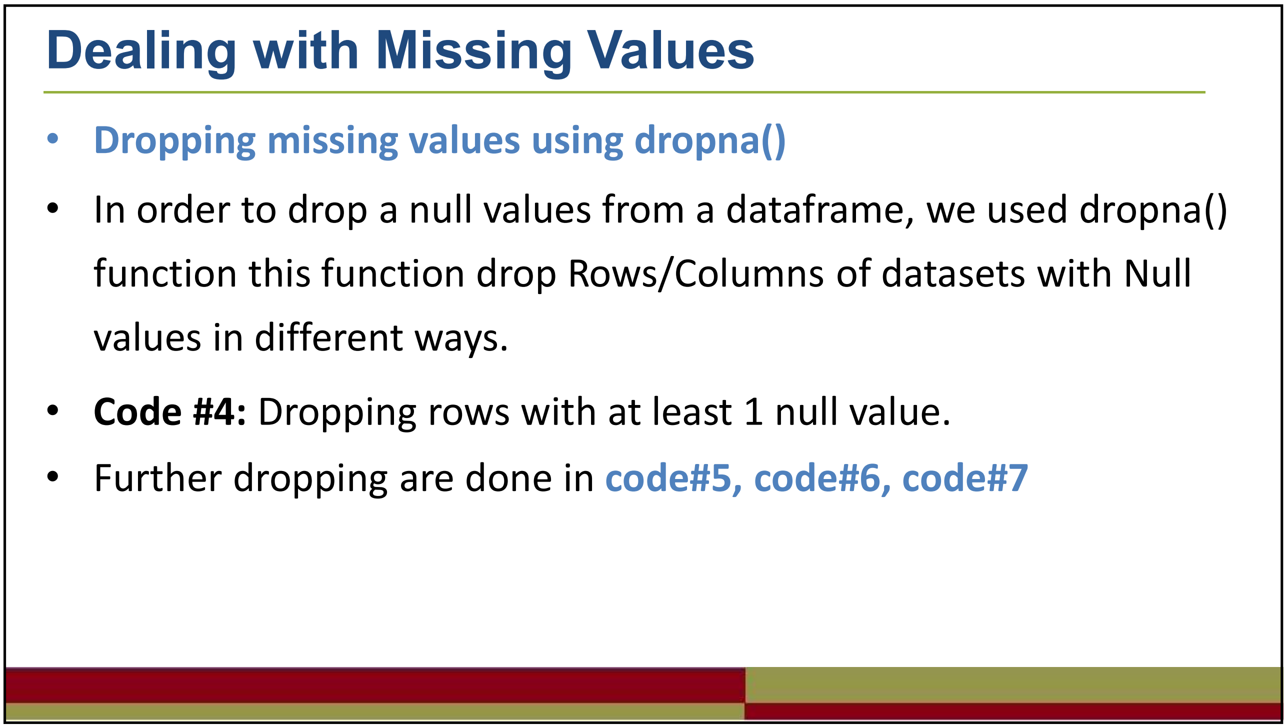
- Both function help in checking whether a value is NaN or not.
- These function can also be used in Pandas Series in order to find null values in a series.

Dealing with Missing Values

- Checking for missing values using isnull()
- In order to check null values in Pandas DataFrame, we use isnull() function this function return dataframe of Boolean values which are True for NaN values. Code #1: BDA-8 jupyter notebook

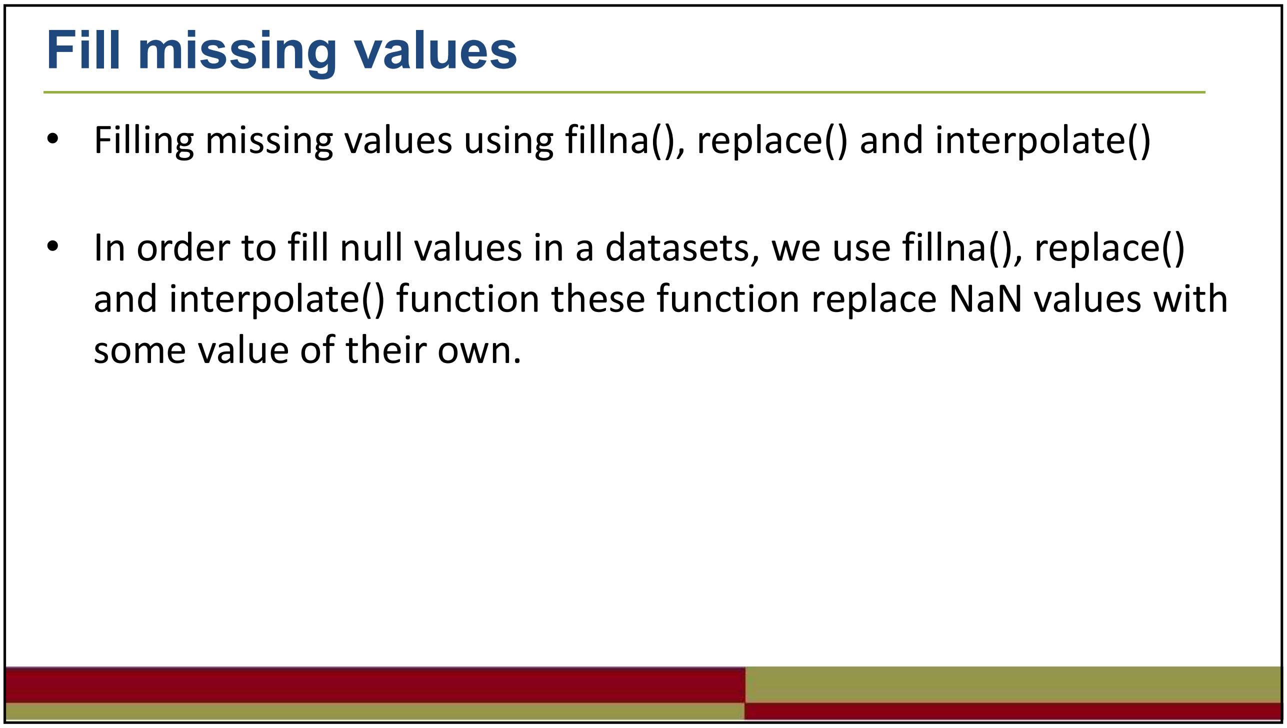
Dealing with Missing Values

- **Checking for missing values using notnull()**
- In order to check null values in Pandas Dataframe, we use notnull() function this function return Dataframe of Boolean values which are False for NaN values. **Code #3:**



Dealing with Missing Values

- Dropping missing values using dropna()
- In order to drop a null values from a dataframe, we used dropna() function this function drop Rows/Columns of datasets with Null values in different ways.
- **Code #4:** Dropping rows with at least 1 null value.
- Further dropping are done in **code#5, code#6, code#7**



Fill missing values

- Filling missing values using fillna(), replace() and interpolate()
- In order to fill null values in a datasets, we use fillna(), replace() and interpolate() function these function replace NaN values with some value of their own.

Fill missing values

- All these function help in filling a null values in datasets of a DataFrame.
- Interpolate() function is basically used to fill NA values in the dataframe but it uses various interpolation technique to fill the missing values rather than hard-coding the value.
- **Code #9**: Filling null values with a single value
- **Code#10, Code#11, Code#12, Code#13, and Code#14**

Fill missing values with aggregate functions

- Check the code# 15 and code#16