Working with Missing Data in Pandas

Missing Data can occur when no information is provided for one or more items or

for a whole unit. Missing Data is a very big problem in a real-life scenarios. Missing

Data can also refer to as NA(Not Available) values in pandas. In DataFrame

sometimes many datasets simply arrive with missing data, either because it exists

and was not collected or it never existed

In Pandas missing data is represented by two value:

 None: None is a Python singleton object that is often used for missing data in

Python code.

 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

there are several useful functions for detecting, removing, and replacing null values

in Pandas DataFrame :

 isnull()

 notnull()

 dropna()

 fillna()

 replace()

 interpolate()

Program

#Import Neccesary Libraries

import pandas as pd

import numpy as np

#Create A DATAFRAME with a column Heading Test1,Test2,Test3 and row heading a,b,d,f,h

df=pd.DataFrame(np.random.randn(5,3),index=["A","C","E","G","H"],columns=["TEST1","TEST2","TEST3"])

print(df)

#Using Reindexing, WE Have to create a DataFrame with missing values

df=df.reindex(["A","B","C","D","E","F","G"])

df

#push a Data

print(df.fillna(method='pad'))

print(df.fillna(method='backfill'))