**DATA CLEANING AND PREPARATION - PANDAS**

**Ex.No. :4a 230901039**

**Date :28-02-2025**

**AIM:**

To do data cleaning and preparation using dataframe.

**SOFTWARE USED:**

Jupyter notebook

**DESCRIPTION:**

1. Handling the missing data using pandas data frame.
2. Drop missing values using dropna() function.
3. Fill the missing values using fillna() function.
4. Replace the missing values using replace() function with scalar values.
5. Fill the missing values from forward and backward values through pad or ffill or bfill functions.

**PROGRAM:**

**import** pandas **as** pd

**import** numpy **as** np

df**=**pd**.**DataFrame(np**.**random**.**randn(4,3),index**=**['a','b','d','f'],columns**=**['one','two','three'])

df**=**df**.**reindex(['a','b','c','d','e','f'])

print("Original dataframe with NaN\n",df)

print("\n")

print("Droped Dataframe\n",df**.**dropna())

print("\n")

print("NaN replaced with '1':\n")

print(df**.**fillna(1))

**import** pandas **as** pd

**import** numpy **as** np

df**=**pd**.**DataFrame(np**.**random**.**randn(4,3),index**=**['a','b','d','f'],columns**=**['one','two','three'])

df**=**df**.**reindex(['a','b','c','d','e','f'])

print("NaN with True Fill:\n")

print(df['one']**.**isnull())

**import** pandas **as** pd

**import** numpy **as** np

df**=**pd**.**DataFrame(np**.**random**.**randn(4,3),index**=**['a','b','d','f'],columns**=**['one','two','three'])

df**=**df**.**reindex(['a','b','c','d','e','f'])

print("NaN filled with background values:\n")

print(df**.**bfill())*#backfill*

**import** pandas **as** pd

**import** numpy **as** np

df**=**pd**.**DataFrame(np**.**random**.**randn(4,3),index**=**['a','b','d','f'],columns**=**['one','two','three'])

df**=**df**.**reindex(['a','b','c','d','e','f'])

print("NaN filled with foreground values:\n")

print(df**.**ffill())

df**=**pd**.**DataFrame({'one':[10,20,80,40,50],'two':[60,70,80,0,10]})

print("Original datframe:\n",df)

print("\nDataframe with replaced values:\n")

print(df**.**replace({10:5,80:30}))

**OUTPUT:**

Original dataframe with NaN

one two three

a -0.831753 0.295343 -0.202132

b -0.261707 0.900825 -0.475135

c NaN NaN NaN

d -1.398241 0.211079 -0.492954

e NaN NaN NaN

f -0.247728 -1.195857 0.182530

Droped Dataframe

one two three

a -0.831753 0.295343 -0.202132

b -0.261707 0.900825 -0.475135

d -1.398241 0.211079 -0.492954

f -0.247728 -1.195857 0.182530

NaN replaced with '1':

one two three

a -0.831753 0.295343 -0.202132

b -0.261707 0.900825 -0.475135

c 1.000000 1.000000 1.000000

d -1.398241 0.211079 -0.492954

e 1.000000 1.000000 1.000000

f -0.247728 -1.195857 0.182530

NaN with True Fill:

a False

b False

c True

d False

e True

f False

Name: one, dtype: bool

NaN filled with background values:

one two three

a 1.146313 0.671651 -0.442012

b -0.990522 -1.165514 2.004817

c -1.724100 1.099339 1.445647

d -1.724100 1.099339 1.445647

e -0.531620 0.564976 2.266509

f -0.531620 0.564976 2.266509

NaN filled with foreground values:

one two three

a 0.673582 0.050569 -1.039491

b -1.260888 0.213760 -0.241731

c -1.260888 0.213760 -0.241731

d -0.791382 -1.598773 -0.184952

e -0.791382 -1.598773 -0.184952

f 0.091394 1.006954 0.269622

Original datframe:

one two

0 10 60

1 20 70

2 80 80

3 40 0

4 50 10

Dataframe with replaced values:

one two

0 5 60

1 20 70

2 30 30

3 40 0

4 50 5

**Result:**

Thus data cleaning and preparation was done using pandas.