PYTHHON PROGRAMMING

Lab-27 Answers

HAREESHA H M

AF0364330

1.Write a Pandas program to detect missing values of a given DataFrame.

Input: df = pd.DataFrame({'ord\_no':[70001,np.nan,70002,70004,np.nan,70005,np.nan,70010,70003,70012,np.na n,70013], 'purch\_amt':[150.5,270.65,65.26,110.5,948.5,2400.6,5760,1983.43,2480.4,250.45, 75.29,3045.6], 'ord\_date': ['2012-10-05','2012-09-10',np.nan,'2012-08-17','2012-09-10','2012-07-27','2012-09-10' ,'2012-10-10','2012-10-10','2012-06-27','2012-08-17','2012-04-25'], 'customer\_id':[3002,3001,3001,3003,3002,3001,3001,3004,3003,3002,3001,3001], 'salesman\_id':[5002,5003,5001,np.nan,5002,5001,5001,np.nan,5003,5002,5003,np.n an]})

Code:

import pandas as pd #importing pandas as pd.

import numpy as np #importing numpy as np.

# Creating the DataFrame.

df = pd.DataFrame({

    'ord\_no': [70001, np.nan, 70002, 70004, np.nan, 70005, np.nan, 70010, 70003, 70012, np.nan, 70013],

    'purch\_amt': [150.5, 270.65, 65.26, 110.5, 948.5, 2400.6, 5760, 1983.43, 2480.4, 250.45, 75.29, 3045.6],

    'ord\_date': ['2012-10-05', '2012-09-10', np.nan, '2012-08-17', '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', '2012-04-25'],

    'customer\_id': [3002, 3001, 3001, 3003, 3002, 3001, 3001, 3004, 3003, 3002, 3001, 3001],

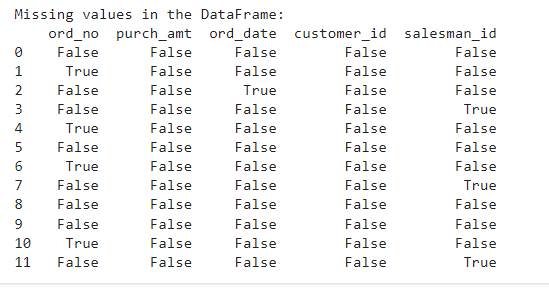
    'salesman\_id': [5002, 5003, 5001, np.nan, 5002, 5001, 5001, np.nan, 5003, 5002, 5003, np.nan]

})

missing\_values = df.isna() # Detect missing values.

print("Missing values in the DataFrame:") #printing the missing values.

print(missing\_values)

Output:

2.Write a Pandas program to drop the rows where at least one element is missing in a given DataFrame.

Input: df = pd.DataFrame({ 'ord\_no':[70001,np.nan,70002,70004,np.nan,70005,np.nan,70010,70003,70012,np.na n,70013], 'purch\_amt':[150.5,270.65,65.26,110.5,948.5,2400.6,5760,1983.43,2480.4,250.45, 75.29,3045.6], 'ord\_date': ['2012-10-05','2012-09-10',np.nan,'2012-08-17','2012-09-10','2012-07-27','2012-09-10' ,'2012-10-10','2012-10-10','2012-06-27','2012-08-17','2012-04-25'], 'customer\_id':[3002,3001,3001,3003,3002,3001,3001,3004,3003,3002,3001,3001],

'salesman\_id':[5002,5003,5001,np.nan,5002,5001,5001,np.nan,5003,5002,5003,np.n an]})

Code:

import pandas as pd #importing pandas as pd.

import numpy as np #importing numpy as np.

# Creating the DataFrame.

df = pd.DataFrame({

    'ord\_no': [70001, np.nan, 70002, 70004, np.nan, 70005, np.nan, 70010, 70003, 70012, np.nan, 70013],

    'purch\_amt': [150.5, 270.65, 65.26, 110.5, 948.5, 2400.6, 5760, 1983.43, 2480.4, 250.45, 75.29, 3045.6],

    'ord\_date': ['2012-10-05', '2012-09-10', np.nan, '2012-08-17', '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', '2012-04-25'],

    'customer\_id': [3002, 3001, 3001, 3003, 3002, 3001, 3001, 3004, 3003, 3002, 3001, 3001],

    'salesman\_id': [5002, 5003, 5001, np.nan, 5002, 5001, 5001, np.nan, 5003, 5002, 5003, np.nan]

})

df\_cleaned = df.dropna() # Drop rows with missing values.

print("DataFrame after dropping rows with missing values:")

print(df\_cleaned)

Output:

DataFrame after dropping rows with missing values:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001.0 150.50 2012-10-05 3002 5002.0

5 70005.0 2400.60 2012-07-27 3001 5001.0

8 70003.0 2480.40 2012-10-10 3003 5003.0

9 70012.0 250.45 2012-06-27 3002 5002.0

3.Write a Pandas program to drop the rows where all elements are missing in a given DataFrame.

df = pd.DataFrame({ 'ord\_no':[np.nan,np.nan,70002,70004,np.nan,70005,np.nan,70010,70003,70012,np.n an,70013], 'purch\_amt':[np.nan,270.65,65.26,110.5,948.5,2400.6,5760,1983.43,2480.4,250.45, 75.29,3045.6], 'ord\_date': [np.nan,'2012-09-10',np.nan,'2012-08-17','2012-09-10','2012-07-27','2012-09-10','201 2-10-10','2012-10-10','2012-06-27','2012-08-17','2012-04-25'], 'customer\_id':[np.nan,3001,3001,3003,3002,3001,3001,3004,3003,3002,3001,3001]})

Code:

import pandas as pd #importing pandas as pd.

import numpy as np #importing numpy as np.

# Creating the DataFrame.

df = pd.DataFrame({

    'ord\_no': [np.nan, np.nan, 70002, 70004, np.nan, 70005, np.nan, 70010, 70003, 70012, np.nan, 70013],

    'purch\_amt': [np.nan, 270.65, 65.26, 110.5, 948.5, 2400.6, 5760, 1983.43, 2480.4, 250.45, 75.29, 3045.6],

    'ord\_date': [np.nan, '2012-09-10', np.nan, '2012-08-17', '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', '2012-04-25'],

    'customer\_id': [np.nan, 3001, 3001, 3003, 3002, 3001, 3001, 3004, 3003, 3002, 3001, 3001]

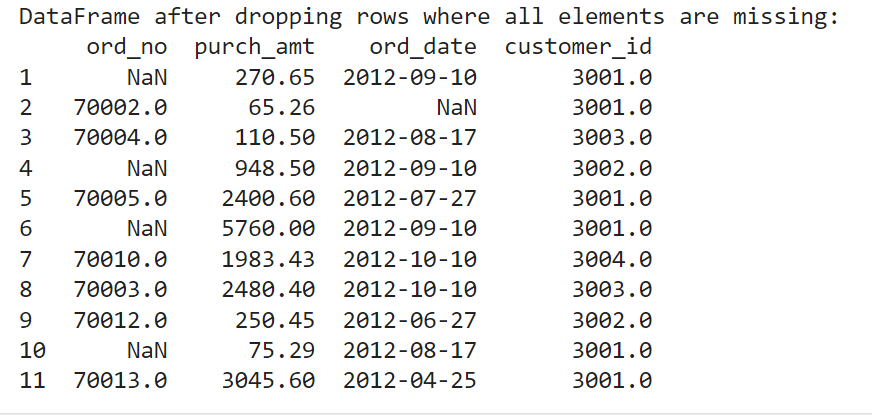
})

df\_cleaned = df.dropna(how='all')# Drop rows where all elements are missing.

print("DataFrame after dropping rows where all elements are missing:")

print(df\_cleaned)

Output:



4.Write a Pandas program to drop those rows from a given DataFrame in which specific columns have missing values. Input:

df = pd.DataFrame({ 'ord\_no':[np.nan,np.nan,70002,np.nan,np.nan,70005,np.nan,70010,70003,70012,np.n an,np.nan], 'purch\_amt':[np.nan,270.65,65.26,np.nan,948.5,2400.6,5760,1983.43,2480.4,250.45, 75.29,np.nan], 'ord\_date': [np.nan,'2012-09-10',np.nan,np.nan,'2012-09-10','2012-07-27','2012-09-10','2012-10- 10','2012-10-10','2012-06-27','2012-08-17',np.nan], 'customer\_id':[np.nan,3001,3001,np.nan,3002,3001,3001,3004,3003,3002,3001,np.na n]})

Code:

import pandas as pd #importing pandas as pd.

import numpy as np #importing numpy as np.

# Creating the DataFrame

df = pd.DataFrame({

    'ord\_no': [np.nan, np.nan, 70002, np.nan, np.nan, 70005, np.nan, 70010, 70003, 70012, np.nan, np.nan],

    'purch\_amt': [np.nan, 270.65, 65.26, np.nan, 948.5, 2400.6, 5760, 1983.43, 2480.4, 250.45, 75.29, np.nan],

    'ord\_date': [np.nan, '2012-09-10', np.nan, np.nan, '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', np.nan],

    'customer\_id': [np.nan, 3001, 3001, np.nan, 3002, 3001, 3001, 3004, 3003, 3002, 3001, np.nan]

})

columns\_to\_check = ['ord\_no', 'purch\_amt', 'ord\_date', 'customer\_id'] # Drop rows where specific columns have missing values.

df\_cleaned = df.dropna(subset=columns\_to\_check)

print("DataFrame after dropping rows with missing values in specific columns:")

print(df\_cleaned)

Output:

DataFrame after dropping rows with missing values in specific columns:

ord\_no purch\_amt ord\_date customer\_id

5 70005.0 2400.60 2012-07-27 3001.0

7 70010.0 1983.43 2012-10-10 3004.0

8 70003.0 2480.40 2012-10-10 3003.0

9 70012.0 250.45 2012-06-27 3002.0