SUMANTH S
AF0363570

LAB: PANDAS IO

Lab1: 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.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.n an]})

```
import pandas as pd
# Given 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]
})
# Detecting missing values
missing_values = df.isnull()
print(missing_values)
```

```
    ord_no  purch_amt  ord_date  customer_id  salesman_id
0    False      False     False        False        False
1     True      False     False        False        False
2    False      False      True        False        False
3    False      False     False        False         True
4     True      False     False        False        False
5    False      False     False        False        False
6     True      False     False        False        False
7    False      False     False        False         True
8    False      False     False        False        False
9    False      False     False        False        False
10    True      False     False        False        False
11   False      False     False        False         True
```

Lab2: 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]})

```
import pandas as pd

# Given 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]
})

# Dropping rows with missing values
df_cleaned = df.dropna()
print(df_cleaned)
```

```
    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
```

Lab3: 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','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]})

```
import pandas as pd
# Define 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']
})
# Drop rows where all elements are missing
df_cleaned = df.dropna(how='all')
# Print the cleaned DataFrame
print(df_cleaned)
```

```
     ord_no  purch_amt    ord_date
1       NaN     270.65  2012-09-10
2   70002.0      65.26         NaN
3   70004.0     110.50  2012-08-17
4       NaN     948.50  2012-09-10
5   70005.0    2400.60  2012-07-27
6       NaN    5760.00  2012-09-10
7   70010.0    1983.43  2012-10-10
8   70003.0    2480.40  2012-10-10
9   70012.0     250.45  2012-06-27
10      NaN      75.29  2012-08-17
11  70013.0    3045.60  2012-04-25
```

Lab4: 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]})

```
import pandas as pd
# Define 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', '2012-04-25']
})
# Drop rows where specific columns have missing values
df_cleaned = df.dropna(subset=['ord_no', 'purch_amt'])
# Print the cleaned DataFrame
print(df_cleaned)
```

```
    ord_no  purch_amt    ord_date
2  70002.0      65.26         NaN
5  70005.0    2400.60  2012-07-27
7  70010.0    1983.43  2012-10-10
8  70003.0    2480.40  2012-10-10
9  70012.0     250.45  2012-06-27
```