**1.** Write a Pandas program to detect missing values of a given DataFrame. Display True or False.

### **Test Data:**

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	$7000\overline{1.0}$	150.50	2012-10-05	3002	$500\overline{2.0}$
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0
10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

**2.** Write a Pandas program to identify the column(s) of a given DataFrame which have at least one missing value.

#### **Test Data:**

	ord no	purch amt	ord date	customer id	salesman id
0	$7000\overline{1}.0$	$15\overline{0.50}$	2012-10-05	3002	$500\overline{2}.0$
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0
10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

**3.** Write a Pandas program to count the number of missing values in each column of a given DataFrame.

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	$7000\overline{1.0}$	150.50	2012-10-05	3002	$500\overline{2.0}$
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0

70004.0	110.50	2012-08-17	3003	NaN
NaN	948.50	2012-09-10	3002	5002.0
70005.0	2400.60	2012-07-27	3001	5001.0
NaN	5760.00	2012-09-10	3001	5001.0
70010.0	1983.43	2012-10-10	3004	NaN
70003.0	2480.40	2012-10-10	3003	5003.0
70012.0	250.45	2012-06-27	3002	5002.0
NaN	75.29	2012-08-17	3001	5003.0
70013.0	3045.60	2012-04-25	3001	NaN
	NaN 70005.0 NaN 70010.0 70003.0 70012.0 NaN	NaN 948.50 70005.0 2400.60 NaN 5760.00 70010.0 1983.43 70003.0 2480.40 70012.0 250.45 NaN 75.29	NaN948.502012-09-1070005.02400.602012-07-27NaN5760.002012-09-1070010.01983.432012-10-1070003.02480.402012-10-1070012.0250.452012-06-27NaN75.292012-08-17	NaN       948.50       2012-09-10       3002         70005.0       2400.60       2012-07-27       3001         NaN       5760.00       2012-09-10       3001         70010.0       1983.43       2012-10-10       3004         70003.0       2480.40       2012-10-10       3003         70012.0       250.45       2012-06-27       3002         NaN       75.29       2012-08-17       3001

### **4.** Write a Pandas program to find and replace the missing values in a given DataFrame which do not have any valuable information.

#### **Test Data:**

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	70001	150.5	?	3002	5002
1	NaN	270.65	2012-09-10	3001	5003
2	70002	65.26	NaN	3001	3
3	70004	110.5	2012-08-17	3003	5001
4	NaN	948.5	2012-09-10	3002	NaN
5	70005	2400.6	2012-07-27	3001	5002
6		5760	2012-09-10	3001	5001
7	70010	?	2012-10-10	3004	?
8	70003	12.43	2012-10-10		5003
9	70012	2480.4	2012-06-27	3002	5002
10	NaN	250.45	2012-08-17	3001	5003
11	70013	3045.6	2012-04-25	3001	

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

	ord no	purch amt	ord date	customer id	salesman id
0	$7000\overline{1.0}$	$15\overline{0.50}$	2012-10-05	3 <del>0</del> 02	$500\overline{2}.0$
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0

10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

# **6.** Write a Pandas program to drop the columns where at least one element is missing in a given DataFrame.

### **Test Data:**

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	70001.0	150.50	2012-10-05	3002	5002.0
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0
10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

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

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	70004.0	110.50	2012-08-17	3003.0
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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
10	NaN	75.29	2012-08-17	3001.0
11	70013.0	3045.60	2012-04-25	3001.0

**8.** Write a Pandas program to keep the rows with at least 2 NaN values in a given DataFrame.

#### Test Data:

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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
10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

**9.** Write a Pandas program to drop those rows from a given DataFrame in which specific columns have missing values.

### Test Data:

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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
10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

**10.** Write a Pandas program to keep the valid entries of a given DataFrame. **Test Data**:

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0

3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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
10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

### **11.** Write a Pandas program to calculate the total number of missing values in a DataFrame.

### **Test Data:**

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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
10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

# **12.** Write a Pandas program to replace NaNs with a single constant value in specified columns in a DataFrame.

	ord_no	purch_amt	ord_date	customer_id
0	$\overline{\mathtt{N}}\mathtt{a}\mathtt{N}$	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	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

10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

# **13.** Write a Pandas program to replace NaNs with the value from the previous row or the next row in a given DataFrame.

### **Test Data:**

<del></del>	purch_amt	sale_amt	ord_date	customer_id
salesman_id 0 70001.0 5002.0	150.50	10.50	2012-10-05	3002
1 NaN 5003.0	NaN	20.65	2012-09-10	3001
2 70002.0 5001.0	65.26	NaN	NaN	3001
3 70004.0 NaN	110.50	11.50	2012-08-17	3003
4 NaN 5002.0	948.50	98.50	2012-09-10	3002
5 70005.0 5001.0	NaN	NaN	2012-07-27	3001
6 NaN 5001.0	5760.00	57.00	2012-09-10	3001
	1983.43	19.43	2012-10-10	3004
8 70003.0 5003.0	NaN	NaN	2012-10-10	3003
	250.45	25.45	2012-06-27	3002
10 NaN 5003.0	75.29	75.29	2012-08-17	3001
11 70013.0 NaN	3045.60	35.60	2012-04-25	3001

### **14.** Write a Pandas program to replace NaNs with median or mean of the specified columns in a given DataFrame.

ord_no	purch_amt	sale_amt	ord_date	customer_id
salesman_id				
$0 7000\overline{1.0}$	150.50	10.50	2012-10-05	3002
5002.0				
1 NaN	NaN	20.65	2012-09-10	3001
5003.0				

2 70002.0	65.26	NaN	NaN	3001
5001.0				
3 70004.0	110.50	11.50	2012-08-17	3003
NaN				
4 NaN	948.50	98.50	2012-09-10	3002
5002.0				
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0				
6 NaN	5760.00	57.00	2012-09-10	3001
5001.0				
7 70010.0	1983.43	19.43	2012-10-10	3004
NaN				
8 70003.0	NaN	NaN	2012-10-10	3003
5003.0				
9 70012.0	250.45	25.45	2012-06-27	3002
5002.0				
10 NaN	75.29	75.29	2012-08-17	3001
5003.0				
11 70013.0	3045.60	35.60	2012-04-25	3001
NaN				

**15.** Write a Pandas program to interpolate the missing values using the Linear Interpolation method in a given DataFrame.

From Wikipedia, in mathematics, linear interpolation is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points.

ord_no	purch_amt	sale_amt	ord_date	customer_id
salesman_id 0 70001.0 5002.0	150.50	10.50	2012-10-05	3002
1 NaN 5003.0	NaN	20.65	2012-09-10	3001
2 70002.0 5001.0	65.26	NaN	NaN	3001
3 70004.0 NaN	110.50	11.50	2012-08-17	3003
4 NaN 5002.0	948.50	98.50	2012-09-10	3002
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0 6 NaN 5001.0	5760.00	57.00	2012-09-10	3001

7	70010.0	1983.43	19.43	2012-10-10	3004
NaN					
8	70003.0	NaN	NaN	2012-10-10	3003
500	3.0				
9	70012.0	250.45	25.45	2012-06-27	3002
500	2.0				
10	NaN	75.29	75.29	2012-08-17	3001
500	3.0				
11	70013.0	3045.60	35.60	2012-04-25	3001
NaN					

### **16.** Write a Pandas program to count the number of missing values of a specified column in a given DataFrame.

### **Test Data:**

_	purch_amt	sale_amt	ord_date	customer_id
salesman_id	150 50	10 50	0010 10 05	2000
0 70001.0 5002.0	150.50	10.50	2012-10-05	3002
1 NaN	NaN	20.65	2012-09-10	3001
5003.0				
2 70002.0	65.26	NaN	NaN	3001
5001.0 3 70004.0	110 50	11 50	2012 00 17	3003
NaN	110.50	11.50	2012-06-17	3003
4 NaN	948.50	98.50	2012-09-10	3002
5002.0				
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0				
6 NaN	5760.00	57.00	2012-09-10	3001
5001.0 7 70010.0	1983.43	19 /3	2012-10-10	3004
NaN	1909.45	17.43	2012 10 10	3004
8 70003.0	NaN	NaN	2012-10-10	3003
5003.0				
9 70012.0	250.45	25.45	2012-06-27	3002
5002.0				
10 NaN	75.29	75.29	2012-08-17	3001
5003.0				
11 70013.0	3045.60	35.60	2012-04-25	3001
NaN				

**17.** Write a Pandas program to count the missing values in a given DataFrame. **Test Data:** 

	2012-10-05 2012-09-10	
20.65	2012-09-10	3001
20.65	2012-09-10	3001
		3001
		0.0.01
NaN	NaN	3001
44 50	0010 00 15	0.000
11.50	2012-08-17	3003
00 50	2012 00 10	2000
98.50	2012-09-10	3002
$N \supset N$	2012-07-27	3001
ivaiv	2012 07 27	3001
57 00	2012-09-10	3001
37.00	2012 09 10	3001
19.43	2012-10-10	3004
<b>13</b> ( 13		0001
NaN	2012-10-10	3003
25.45	2012-06-27	3002
75.29	2012-08-17	3001
35.60	2012-04-25	3001
	11.50 98.50 NaN 57.00 19.43 NaN 25.45 75.29	NaN NaN  11.50 2012-08-17  98.50 2012-09-10  NaN 2012-07-27  57.00 2012-09-10  19.43 2012-10-10  NaN 2012-10-10  25.45 2012-06-27

### **18.** Write a Pandas program to find the Indexes of missing values in a given DataFrame.

ord_no	purch_amt	sale_amt	ord_date	customer_id
salesman id				
$0 7000\overline{1.0}$	150.50	10.50	2012-10-05	3002
5002.0				
1 NaN	NaN	20.65	2012-09-10	3001
5003.0				
2 70002.0	65.26	NaN	NaN	3001
5001.0				
3 70004.0	110.50	11.50	2012-08-17	3003
NaN				
4 NaN	948.50	98.50	2012-09-10	3002
5002.0				
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0				

6	NaN	5760.00	57.00	2012-09-10	3001
500	1.0				
7	70010.0	1983.43	19.43	2012-10-10	3004
NaN					
8	70003.0	NaN	NaN	2012-10-10	3003
5003.0					
9	70012.0	250.45	25.45	2012-06-27	3002
500	2.0				
10	NaN	75.29	75.29	2012-08-17	3001
500	3.0				
11	70013.0	3045.60	35.60	2012-04-25	3001
NaN					

# **19.** Write a Pandas program to replace the missing values with the most frequent values present in each column of a given DataFrame.

ord_no	purch_amt	sale_amt	ord_date	customer_id
salesman id				
$0 7000\overline{1.0}$	150.50	10.50	2012-10-05	3002
5002.0				
	NaN	20 65	2012-09-10	3001
5003.0	1,011	20.00	2012 03 10	0001
2 70002.0	65.26	NaN	NaN	3001
5001.0	03.20	nan	Ivalv	3001
	110 50	11 50	0010 00 17	2002
	110.50	11.50	2012-08-17	3003
NaN				
4 NaN	948.50	98.50	2012-09-10	3002
5002.0				
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0				
6 NaN	5760.00	57.00	2012-09-10	3001
5001.0				
7 70010.0	1983.43	10 /3	2012-10-10	3004
NaN	1703.43	17.43	2012 10 10	3004
	NT - NT	NT - NT	0010 10 10	2002
8 70003.0	NaN	Nan	2012-10-10	3003
5003.0				
9 70012.0	250.45	25.45	2012-06-27	3002
5002.0				
10 NaN	75.29	75.29	2012-08-17	3001
5003.0				
	3045 60	35 60	2012-04-25	3001
NaN	J045.00	33.00	2012 04 20	5001
INAIN				

**20.** Write a Pandas program to create a hitmap for more information about the distribution of missing values in a given DataFrame.

ord_no	purch_amt	sale_amt	ord_date	customer_id
salesman_id				
0 70001.0	150.50	10.50	2012-10-05	3002
5002.0				
1 NaN	NaN	20.65	2012-09-10	3001
5003.0				
2 70002.0	65.26	NaN	NaN	3001
5001.0				
3 70004.0	110.50	11.50	2012-08-17	3003
NaN				
4 NaN	948.50	98.50	2012-09-10	3002
5002.0				
5 70005.0	NaN	NaN	2012-07-27	3001
5001.0	IVAIV	IVAIV	2012 07 27	3001
6 NaN	5760 00	57.00	2012-09-10	3001
5001.0	3700.00	37.00	2012 05 10	3001
7 70010.0	1983.43	19.43	2012-10-10	3004
	1903.43	19.43	2012-10-10	3004
NaN	NI - NI	NI - NI	2012 10 10	2002
8 70003.0	NaN	NaN	2012-10-10	3003
5003.0	050 45	05.45	0010 06 07	2000
9 70012.0	250.45	25.45	2012-06-27	3002
5002.0				
10 NaN	75.29	75.29	2012-08-17	3001
5003.0				
11 70013.0	3045.60	35.60	2012-04-25	3001
NaN				