```
import numpy as np
import pandas as pd

df=pd.read_csv("/StudentPerformance.csv")

from sklearn.datasets import load_boston
boston_dataset = load_boston()
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	75	2020	2
1	77	74	75	61	2019	1
2	61	60	79	69	2022	1
3	66	62	69	80	2018	2
4	95	66	60	61	2021	1
5	71	79	90	71	2018	1
6	65	70	63	74	2020	1
7	63	79	71	72	2019	1
8	68	60	63	70	2020	1
9	64	89	64	84	2019	2
10	72	68	62	66	2022	2
11	67	80	74	77	2018	2
12	71	68	78	74	2021	1
13	60	70	78	77	2018	2
14	75	64	68	85	2020	3
15	66	71	63	67	2019	1
16	67	66	86	80	2020	2
17	73	64	63	64	2019	1
18	73	61	76	82	2022	2
19	97	80	60	72	2018	1
20	66	69	98	76	2021	2
21	79	75	75	69	2018	1
22	77	77	70	64	2020	1
23	78	73	63	80	2019	2
24	76	69	68	65	2020	1
25	61	65	71	75	2019	2
26	61	74	76	80	2022	2
27	68	90	76	69	2018	1
28	76	65	68	61	2021	1
29	64	76	62	72	2018	1

```
df.isnull()
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False
9	False	False	False	False	False	False
10	False	False	False	False	False	False
11	False	False	False	False	False	False
12	False	False	False	False	False	False
13	False	False	False	False	False	False
14	False	False	False	False	False	False
15	False	False	False	False	False	False
16	False	False	False	False	False	False
17	False	False	False	False	False	False
18	False	False	False	False	False	False

```
series=pd.isnull(df["math score"])
df[series]
```

	math score	reading score	writing score	placement score	club join year	placement offer count

```
df.notnull()
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
5	True	True	True	True	True	True

```
series1=pd.notnull(df["math score"])
df[series1]
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	75	2020	2
1	77	74	75	61	2019	1
2	61	60	79	69	2022	1
3	66	62	69	80	2018	2
4	95	66	60	61	2021	1
5	71	79	90	71	2018	1
6	65	70	63	74	2020	1
7	63	79	71	72	2019	1
8	68	60	63	70	2020	1
9	64	89	64	84	2019	2
10	72	68	62	66	2022	2
11	67	80	74	77	2018	2
12	71	68	78	74	2021	1
13	60	70	78	77	2018	2
14	75	64	68	85	2020	3
15	66	71	63	67	2019	1
16	67	66	86	80	2020	2
17	73	64	63	64	2019	1
18	73	61	76	82	2022	2
19	97	80	60	72	2018	1
20	66	69	98	76	2021	2
21	79	75	75	69	2018	1
22	77	77	70	64	2020	1
23	78	73	63	80	2019	2
24	76	69	68	65	2020	1
25	61	65	71	75	2019	2
26	61	74	76	80	2022	2
27	68	90	76	69	2018	1
28	76	65	68	61	2021	1
29	64	76	62	72	2018	1

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['placement score']=le.fit_transform(df['placement score'])
newdf=df
df
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1
16	67	66	86	13	2020	2
17	73	64	63	1	2019	1
18	73	61	76	14	2022	2
19	97	80	60	8	2018	1
20	66	69	98	11	2021	2
21	79	75	75	5	2018	1
22	77	77	70	1	2020	1
23	78	73	63	13	2019	2
24	76	69	68	2	2020	1
25	61	65	71	10	2019	2

```
ndf=df
ndf.fillna(0)
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2

```
data = df
data['math score']=data['math score'].fillna(data['math score'].mean())
data['math score']=data['math score'].fillna(data['math score'].median())
data['math score']=data['math score'].fillna(data['math score'].std())
data['math score']=data['math score'].fillna(data['math score'].min())
data['math score']=data['math score'].fillna(data['math score'].max())
data
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1
16	67	66	86	13	2020	2
17	73	64	63	1	2019	1
18	73	61	76	14	2022	2
19	97	80	60	8	2018	1
20	66	69	98	11	2021	2
21	79	75	75	5	2018	1
22	77	77	70	1	2020	1
23	78	73	63	13	2019	2
24	76	69	68	2	2020	1
25	61	65	71	10	2019	2
26	61	74	76	13	2022	2
27	68	90	76	5	2018	1
28	76	65	68	0	2021	1
29	64	76	62	8	2018	1

```
m_v=df['math score'].mean()
df['math score'].fillna(value=m_v,inplace=True)
df
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1
16	67	66	86	13	2020	2
17	73	64	63	1	2019	1
18	73	61	76	14	2022	2
19	97	80	60	8	2018	1
20	66	69	98	11	2021	2
21	79	75	75	5	2018	1
22	77	77	70	1	2020	1
23	78	73	63	13	2019	2
24	76	69	68	2	2020	1
25	61	65	71	10	2019	2
26	61	74	76	13	2022	2
27	68	90	76	5	2018	1
28	76	65	68	0	2021	1
29	64	76	62	8	2018	1

```
ndf.replace(to_replace=np.nan,value=-99)
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1

```
ndf.dropna()
```

```
ndf.dropna(how='all')
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1
16	67	66	86	13	2020	2
17	73	64	63	1	2019	1
18	73	61	76	14	2022	2
19	97	80	60	8	2018	1
20	66	69	98	11	2021	2
21	79	75	75	5	2018	1
22	77	77	70	1	2020	1
23	78	73	63	13	2019	2
24	76	69	68	2	2020	1
25	61	65	71	10	2019	2
26	61	74	76	13	2022	2
27	68	90	76	5	2018	1
28	76	65	68	0	2021	1
29	64	76	62	8	2018	1

```
ndf.dropna(axis =1)
```

	math score	reading score	writing score	placement score	club join year	placement offer count
0	61	78	76	10	2020	2
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1
13	60	70	78	12	2018	2
14	75	64	68	16	2020	3
15	66	71	63	4	2019	1

```
new_data=ndf.dropna(axis=0,how='any')
new_data
```

	math score	reading score	writing score	placement score	club join year	placement offer count
1	77	74	75	0	2019	1
2	61	60	79	5	2022	1
3	66	62	69	13	2018	2
4	95	66	60	0	2021	1
5	71	79	90	7	2018	1
6	65	70	63	9	2020	1
7	63	79	71	8	2019	1
8	68	60	63	6	2020	1
9	64	89	64	15	2019	2
10	72	68	62	3	2022	2
11	67	80	74	12	2018	2
12	71	68	78	9	2021	1