```
In [16]:   #Name: Dongare Shweta Santosh

           #Roll No: 15
           #Practical no: 02
           #Academic year: 2024-25
```

```
In [2]:   import pandas as pd
```

```
In [3]:   import numpy as np
```

```
In [4]:   df=pd.read_csv("/home/jaihind/Desktop/StudentPerformance.csv")
```

```
In [5]:   df
```

Out[5]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72	72	74.0	78.0	1	Pune
1	female	69	90	88.0	NaN	2	na
2	female	90	95	93.0	74.0	2	Nashik
3	male	47	57	NaN	78.0	1	Na
4	male	na	78	75.0	81.0	3	Pune
5	female	71	Na	78.0	70.0	4	na
6	male	12	44	52.0	12.0	2	Nashik
7	male	NaN	65	67.0	49.0	1	Pune
8	female	5	77	89.0	55.0	0	NaN

```
In [6]:   df.isnull()
```

Out[6]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	False	False	False	False	False	False	False
1	False	False	False	False	True	False	False
2	False	False	False	False	False	False	False
3	False	False	False	True	False	False	False
4	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False
7	False	True	False	False	False	False	False
8	False	False	False	False	False	False	True

```
In [7]: series=pd.isnull(df["math score"])
        df[series]
```

Out[7]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
7	male	NaN	65	67.0	49.0	1	Pune

```
In [8]: import pandas as pd
        import numpy as np
```

```
In [9]: df.notnull()
```

Out[9]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	True	True	True	True	True	True	True
1	True	True	True	True	False	True	True
2	True	True	True	True	True	True	True
3	True	True	True	False	True	True	True
4	True	True	True	True	True	True	True
5	True	True	True	True	True	True	True
6	True	True	True	True	True	True	True
7	True	False	True	True	True	True	True
8	True	True	True	True	True	True	False

```
In [10]: series=pd.notnull(df["math score"])
         df[series]
```

Out[10]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72	72	74.0	78.0	1	Pune
1	female	69	90	88.0	NaN	2	na
2	female	90	95	93.0	74.0	2	Nashik
3	male	47	57	NaN	78.0	1	Na
4	male	na	78	75.0	81.0	3	Pune
5	female	71	Na	78.0	70.0	4	na
6	male	12	44	52.0	12.0	2	Nashik
8	female	5	77	89.0	55.0	0	NaN

```
In [11]: from sklearn.preprocessing import LabelEncoder
```

```
In [12]: from sklearn.preprocessing import LabelEncoder
```

```
In [13]: le=LabelEncoder()
```

In [14]: ```python
df['gender']=le.fit_transform(df['gender'])
```

In [15]: ```python
newdf=df
df
```

Out[15]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	0	72	72	74.0	78.0	1	Pune
1	0	69	90	88.0	NaN	2	na
2	0	90	95	93.0	74.0	2	Nashik
3	1	47	57	NaN	78.0	1	Na
4	1	na	78	75.0	81.0	3	Pune
5	0	71	Na	78.0	70.0	4	na
6	1	12	44	52.0	12.0	2	Nashik
7	1	NaN	65	67.0	49.0	1	Pune
8	0	5	77	89.0	55.0	0	NaN

In [16]: ```python
missing_values=["Na","na"]
```

In [17]: ```python
df=pd.read_csv("/home/jaihind/Desktop/StudentPerformance.csv",na_values=missing_values)
```

In [18]: ```python
df
```

Out[18]:

| | gender | math score | reading score | writing score | placement score | placement offer count | Region |
|---|---|---|---|---|---|---|---|
| 0 | female | 72.0 | 72.0 | 74.0 | 78.0 | 1 | Pune |
| 1 | female | 69.0 | 90.0 | 88.0 | NaN | 2 | NaN |
| 2 | female | 90.0 | 95.0 | 93.0 | 74.0 | 2 | Nashik |
| 3 | male | 47.0 | 57.0 | NaN | 78.0 | 1 | NaN |
| 4 | male | NaN | 78.0 | 75.0 | 81.0 | 3 | Pune |
| 5 | female | 71.0 | NaN | 78.0 | 70.0 | 4 | NaN |
| 6 | male | 12.0 | 44.0 | 52.0 | 12.0 | 2 | Nashik |
| 7 | male | NaN | 65.0 | 67.0 | 49.0 | 1 | Pune |
| 8 | female | 5.0 | 77.0 | 89.0 | 55.0 | 0 | NaN |

In [19]: ```python
ndf=df
```

```
In [20]: ndf.fillna(0)
```

Out[20]:

	gender	math	reading score	writing score	placement score	placement offer count	Region
0	female	72.0	72.0	74.0	78.0	1	Pune
1	female	69.0	90.0	88.0	0.0	2	0
2	female	90.0	95.0	93.0	74.0	2	Nashik
3	male	47.0	57.0	0.0	78.0	1	0
4	male	0.0	78.0	75.0	81.0	3	Pune
5	female	71.0	0.0	78.0	70.0	4	0
6	male	12.0	44.0	52.0	12.0	2	Nashik
7	male	0.0	65.0	67.0	49.0	1	Pune
8	female	5.0	77.0	89.0	55.0	0	0

```
In [21]: m_v=df['math score'].mean()
```

```
In [22]: df['math score'].fillna(value=m_v, inplace=True)
```

```
In [23]: df
```

Out[23]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72.000000	72.0	74.0	78.0	1	Pune
1	female	69.000000	90.0	88.0	NaN	2	NaN
2	female	90.000000	95.0	93.0	74.0	2	Nashik
3	male	47.000000	57.0	NaN	78.0	1	NaN
4	male	52.285714	78.0	75.0	81.0	3	Pune
5	female	71.000000	NaN	78.0	70.0	4	NaN
6	male	12.000000	44.0	52.0	12.0	2	Nashik
7	male	52.285714	65.0	67.0	49.0	1	Pune
8	female	5.000000	77.0	89.0	55.0	0	NaN

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

Out[24]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72.000000	72.0	74.0	78.0	1	Pune
1	female	69.000000	90.0	88.0	-99.0	2	-99
2	female	90.000000	95.0	93.0	74.0	2	Nashik
3	male	47.000000	57.0	-99.0	78.0	1	-99
4	male	52.285714	78.0	75.0	81.0	3	Pune
5	female	71.000000	-99.0	78.0	70.0	4	-99
6	male	12.000000	44.0	52.0	12.0	2	Nashik
7	male	52.285714	65.0	67.0	49.0	1	Pune
8	female	5.000000	77.0	89.0	55.0	0	-99

In [25]: `ndf.dropna()`

Out[25]:

	gender	math	reading score	writing score	placement score	placement offer count	Region
0	female	72.000000	72.0	74.0	78.0	1	Pune
2	female	90.000000	95.0	93.0	74.0	2	Nashik
4	male	52.285714	78.0	75.0	81.0	3	Pune
6	male	12.000000	44.0	52.0	12.0	2	Nashik
7	male	52.285714	65.0	67.0	49.0	1	Pune

In [26]: `ndf.dropna(how='all')`

Out[26]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72.000000	72.0	74.0	78.0	1	Pune
1	female	69.000000	90.0	88.0	NaN	2	NaN
2	female	90.000000	95.0	93.0	74.0	2	Nashik
3	male	47.000000	57.0	NaN	78.0	1	NaN
4	male	52.285714	78.0	75.0	81.0	3	Pune
5	female	71.000000	NaN	78.0	70.0	4	NaN
6	male	12.000000	44.0	52.0	12.0	2	Nashik
7	male	52.285714	65.0	67.0	49.0	1	Pune
8	female	5.000000	77.0	89.0	55.0	0	NaN

In [27]: `ndf.dropna(axis=1)`

Out[27]:

	gender	math score	placement offer count
0	female	72.000000	1
1	female	69.000000	2
2	female	90.000000	2
3	male	47.000000	1
4	male	52.285714	3
5	female	71.000000	4
6	male	12.000000	2
7	male	52.285714	1
8	female	5.000000	0

In [28]: `new_data=ndf.dropna(axis=0, how='any')`

In [29]: `new_data`

Out[29]:

	gender	math score	reading score	writing score	placement score	placement offer count	Region
0	female	72.000000	72.0	74.0	78.0	1	Pune
2	female	90.000000	95.0	93.0	74.0	2	Nashik
4	male	52.285714	78.0	75.0	81.0	3	Pune
6	male	12.000000	44.0	52.0	12.0	2	Nashik
7	male	52.285714	65.0	67.0	49.0	1	Pune

In [ ]: