In [1]:
```python
import numpy as np
import pandas as pd
df = pd.read_csv("StudentsPerformanceTest1.csv")
df
```

Out[1]:

	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	male	5	77	89.0	55.0	0	NaN

In [2]:
```python
df.describe()
```

Out[2]:

	writing score	Placement Score	placement offer count
count	8.000000	8.000000	9.000000
mean	77.000000	62.125000	1.777778
std	13.416408	23.295846	1.201850
min	52.000000	12.000000	0.000000
25%	72.250000	53.500000	1.000000
50%	76.500000	72.000000	2.000000
75%	88.250000	78.000000	2.000000
max	93.000000	81.000000	4.000000

In [3]:
```python
df.isnull().any().any()
```

Out[3]: True

In [4]:
```python
df.isnull()
```

Out[4]:

	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

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
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 [5]:
```python
series = pd.isnull(df["math score"])
df[series]
```

Out[5]:

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

In [6]:
```python
df.notnull()    #check null value in dataset
```

Out[6]:

	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 [7]:
```python
series1 = pd.notnull(df["math score"])
df[series1]
```

Out[7]:

	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

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
6	male	12	44	52.0	12.0	2	Nashik
8	male	5	77	89.0	55.0	0	NaN

In [8]:
```python
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
newdf=df
df
```

Out[8]:

	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	1	5	77	89.0	55.0	0	NaN

In [19]:
```python
missing_values = ["Na", "na"]
```

In [23]:
```python
df=pd.read_csv("StudentsPerformanceTest1.csv",  na_values     = missing_values)
df
```

Out[23]:

	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	male	5.0	77.0	89.0	55.0	0	NaN

check missing value in pandas

In [24]:
```python
ndf = df
ndf.fillna(0)
```

Out[24]:

	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	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	male	5.0	77.0	89.0	55.0	0	0

filling missing values using mean, medianand standard deviation of that column.

In [26]:
```python
df['math score'] = df['math score'].fillna(df['math score'].mean())
df["math score"] = df["math score"].fillna(df["math score"].median())
df["math score"] = df["math score"].fillna(df["math score"].std())
```

In [28]:
```python
df["math score"] = df["math score"].fillna(df["math score"].min())

df["math score"] = df["math score"].fillna(df["math score"].max())
```

In [29]:
```python
m_v=df['math score'].mean()
df['math score'].fillna(value=m_v, inplace=True)
df
```

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
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

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
8	male	5.000000	77.0	89.0	55.0	0	NaN

In [30]:
```
ndf.replace(to_replace = np.nan, value = -99)
```

Out[30]:

	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	male	5.000000	77.0	89.0	55.0	0	-99

In [31]:
```
df
```

Out[31]:

	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	male	5.000000	77.0	89.0	55.0	0	NaN

In [32]:
```
ndf
```

Out[32]:

	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

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
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	male	5.000000	77.0	89.0	55.0	0	NaN

In [33]:
```python
ndf.dropna()
```

Out[33]:

	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 [34]:
```python
ndf.dropna(how = 'all')
```

Out[34]:

	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	male	5.000000	77.0	89.0	55.0	0	NaN

In [35]:
```python
ndf.dropna(axis = 1)
```

Out[35]:

	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

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

In [36]:
```python
new_data = ndf.dropna(axis = 0, how ='any')
```

In [37]:
```python
new_data
```

Out[37]:

	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 [ ]: