In [3]:
```
1 import pandas as pd
```

In [4]:
```
1 import seaborn as sns
```

In [5]:
```
1 import numpy as py
```

In [17]:
```
1 df= pd.read_csv("C:/Users/Welcome/Music/Book1.csv")
```

In [19]:
```
1 df
```

Out[19]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
0	60.0	63.0	76.0	95.0	2021.0	3	female
1	75.0	70.0	64.0	85.0	2020.0	3	male
2	74.0	NaN	55.0	91.0	2020.0	3	male
3	68.0	76.0	78.0	97.0	2020.0	3	female
4	94.0	67.0	71.0	93.0	2020.0	3	male
5	70.0	64.0	80.0	98.0	2018.0	3	female
6	61.0	78.0	92.0	94.0	NaN	3	male
7	61.0	74.0	78.0	80.0	2021.0	2	male
8	64.0	76.0	79.0	76.0	2019.0	2	male
9	65.0	95.0	75.0	90.0	2020.0	3	female
10	66.0	76.0	67.0	100.0	2019.0	1	male
11	84.0	67.0	71.0	92.0	2020.0	3	male
12	69.0	66.0	70.0	NaN	2021.0	3	female
13	74.0	NaN	65.0	80.0	2021.0	2	male
14	74.0	63.0	72.0	96.0	2018.0	3	male
15	76.0	64.0	80.0	96.0	2020.0	3	male
16	60.0	64.0	54.0	91.0	2021.0	3	female
17	77.0	70.0	72.0	99.0	2020.0	3	male
18	67.0	NaN	64.0	87.0	2018.0	3	female
19	71.0	53.0	78.0	75.0	2018.0	2	female
20	58.0	65.0	NaN	92.0	2019.0	3	female
21	68.0	63.0	62.0	94.0	2021.0	3	male
22	77.0	63.0	68.0	97.0	2021.0	3	female
23	80.0	64.0	86.0	85.0	2018.0	3	female
24	NaN	63.0	67.0	83.0	2018.0	1	male
25	68.0	67.0	73.0	88.0	2019.0	3	female
26	76.0	64.0	68.0	96.0	2021.0	3	female
27	92.0	96.0	61.0	83.0	2018.0	2	male
28	60.0	68.0	59.0	93.0	2020.0	3	male

In [20]:
```
df.isnull()
```

Out[20]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	True	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False
6	False	False	False	False	True	False	False
7	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False
12	False	False	False	True	False	False	False
13	False	True	False	False	False	False	False
14	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False
18	False	True	False	False	False	False	False
19	False	False	False	False	False	False	False
20	False	False	True	False	False	False	False
21	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False
24	True	False	False	False	False	False	False
25	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False

In [21]:
```
series = pd.isnull(df["math score"])
df[series]
```

Out[21]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
24	NaN	63.0	67.0	83.0	2018.0	1	male

In [22]:
```
1 df.notnull()
```

Out[22]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
0	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True
2	True	False	True	True	True	True	True
3	True	True	True	True	True	True	True
4	True	True	True	True	True	True	True
5	True	True	True	True	True	True	True
6	True	True	True	True	False	True	True
7	True	True	True	True	True	True	True
8	True	True	True	True	True	True	True
9	True	True	True	True	True	True	True
10	True	True	True	True	True	True	True
11	True	True	True	True	True	True	True
12	True	True	True	False	True	True	True
13	True	False	True	True	True	True	True
14	True	True	True	True	True	True	True
15	True	True	True	True	True	True	True
16	True	True	True	True	True	True	True
17	True	True	True	True	True	True	True
18	True	False	True	True	True	True	True
19	True	True	True	True	True	True	True
20	True	True	False	True	True	True	True
21	True	True	True	True	True	True	True
22	True	True	True	True	True	True	True
23	True	True	True	True	True	True	True
24	False	True	True	True	True	True	True
25	True	True	True	True	True	True	True
26	True	True	True	True	True	True	True
27	True	True	True	True	True	True	True
28	True	True	True	True	True	True	True

In [23]:
```
series1 = pd.notnull(df["math score"])
df[series1]
```

Out[23]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
0	60.0	63.0	76.0	95.0	2021.0	3	female
1	75.0	70.0	64.0	85.0	2020.0	3	male
2	74.0	NaN	55.0	91.0	2020.0	3	male
3	68.0	76.0	78.0	97.0	2020.0	3	female
4	94.0	67.0	71.0	93.0	2020.0	3	male
5	70.0	64.0	80.0	98.0	2018.0	3	female
6	61.0	78.0	92.0	94.0	NaN	3	male
7	61.0	74.0	78.0	80.0	2021.0	2	male
8	64.0	76.0	79.0	76.0	2019.0	2	male
9	65.0	95.0	75.0	90.0	2020.0	3	female
10	66.0	76.0	67.0	100.0	2019.0	1	male
11	84.0	67.0	71.0	92.0	2020.0	3	male
12	69.0	66.0	70.0	NaN	2021.0	3	female
13	74.0	NaN	65.0	80.0	2021.0	2	male
14	74.0	63.0	72.0	96.0	2018.0	3	male
15	76.0	64.0	80.0	96.0	2020.0	3	male
16	60.0	64.0	54.0	91.0	2021.0	3	female
17	77.0	70.0	72.0	99.0	2020.0	3	male
18	67.0	NaN	64.0	87.0	2018.0	3	female
19	71.0	53.0	78.0	75.0	2018.0	2	female
20	58.0	65.0	NaN	92.0	2019.0	3	female
21	68.0	63.0	62.0	94.0	2021.0	3	male
22	77.0	63.0	68.0	97.0	2021.0	3	female
23	80.0	64.0	86.0	85.0	2018.0	3	female
25	68.0	67.0	73.0	88.0	2019.0	3	female
26	76.0	64.0	68.0	96.0	2021.0	3	female
27	92.0	96.0	61.0	83.0	2018.0	2	male
28	60.0	68.0	59.0	93.0	2020.0	3	male

In [25]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
newdf = df
df
```

16	60.0	64.0	54.0	91.0	2021.0	3	0
17	77.0	70.0	72.0	99.0	2020.0	3	1
18	67.0	NaN	64.0	87.0	2018.0	3	0
19	71.0	53.0	78.0	75.0	2018.0	2	0
20	58.0	65.0	NaN	92.0	2019.0	3	0
21	68.0	63.0	62.0	94.0	2021.0	3	1
22	77.0	63.0	68.0	97.0	2021.0	3	0
23	80.0	64.0	86.0	85.0	2018.0	3	0
24	NaN	63.0	67.0	83.0	2018.0	1	1
25	68.0	67.0	73.0	88.0	2019.0	3	0
26	76.0	64.0	68.0	96.0	2021.0	3	0
27	92.0	96.0	61.0	83.0	2018.0	2	1
28	60.0	68.0	59.0	93.0	2020.0	3	1

In [26]:
```
missing_values = ["Na","na"]
df = pd.read_csv("C:/Users/Welcome/Music/Book1.csv", na_values = missing_values)
df
```

Out[26]:

	math score	reading score	writing score	placememt score	club join year	placement offer count	gender
0	60.0	63.0	76.0	95.0	2021.0	3	female
1	75.0	70.0	64.0	85.0	2020.0	3	male
2	74.0	NaN	55.0	91.0	2020.0	3	male
3	68.0	76.0	78.0	97.0	2020.0	3	female
4	94.0	67.0	71.0	93.0	2020.0	3	male
5	70.0	64.0	80.0	98.0	2018.0	3	female
6	61.0	78.0	92.0	94.0	NaN	3	male
7	61.0	74.0	78.0	80.0	2021.0	2	male
8	64.0	76.0	79.0	76.0	2019.0	2	male
9	65.0	95.0	75.0	90.0	2020.0	3	female
10	66.0	76.0	67.0	100.0	2019.0	1	male
11	84.0	67.0	71.0	92.0	2020.0	3	male
12	69.0	66.0	70.0	NaN	2021.0	3	female
13	74.0	NaN	65.0	80.0	2021.0	2	male
14	74.0	63.0	72.0	96.0	2018.0	3	male
15	76.0	64.0	80.0	96.0	2020.0	3	male
16	60.0	64.0	54.0	91.0	2021.0	3	female
17	77.0	70.0	72.0	99.0	2020.0	3	male
18	67.0	NaN	64.0	87.0	2018.0	3	female
19	71.0	53.0	78.0	75.0	2018.0	2	female
20	58.0	65.0	NaN	92.0	2019.0	3	female
21	68.0	63.0	62.0	94.0	2021.0	3	male
22	77.0	63.0	68.0	97.0	2021.0	3	female
23	80.0	64.0	86.0	85.0	2018.0	3	female
24	NaN	63.0	67.0	83.0	2018.0	1	male
25	68.0	67.0	73.0	88.0	2019.0	3	female
26	76.0	64.0	68.0	96.0	2021.0	3	female
27	92.0	96.0	61.0	83.0	2018.0	2	male
28	60.0	68.0	59.0	93.0	2020.0	3	male

In [ ]:
```
```