

In [92]:

```
import pandas as pd
```

In [93]:

```
df=pd.read_csv(r"C:\Users\GAYATRI\Downloads\archive (10)\StudentsPerformance.csv")
```

In [94]:

```
df      #displaying data
```

Out[94]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74.0
1	female	group C	some college	standard	completed	69	90	88.0
2	female	group B	master's degree	standard	none	90	95	93.0
3	male	group A	associate's degree	free/reduced	none	47	57	44.0
4	male	group C	some college	standard	none	76	66	75.0
5	female	group B	associate's degree	standard	none	71	83	78.0
6	female	group B	some college	standard	completed	88	95	92.0
7	male	group B	some college	free/reduced	none	40	43	39.0
8	male	group D	high school	free/reduced	completed	64	64	67.0
9	female	group B	high school	free/reduced	none	38	60	NaN
10	male	group C	associate's degree	standard	none	58	54	52.0
11	male	group D	associate's degree	standard	none	40	52	43.0
12	female	group B	high school	standard	none	65	81	73.0
13	male	group A	some college	standard	completed	78	72	70.0
14	female	group A	master's degree	standard	none	50	53	58.0
15	female	group C	some high school	standard	none	69	75	78.0
16	male	group C	high school	standard	none	88	89	86.0
17	female	group B	some high school	free/reduced	none	18	32	28.0
18	male	group C	master's degree	free/reduced	completed	46	42	46.0
19	female	group C	associate's degree	free/reduced	none	54	58	61.0
20	male	group D	high school	standard	none	66	69	63.0
21	female	group B	some college	free/reduced	completed	65	75	70.0
22	male	group D	some college	standard	none	44	54	53.0
23	female	group C	some high school	standard	none	69	73	73.0
24	male	group D	bachelor's degree	free/reduced	completed	74	71	80.0

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
25	male	group A	master's degree	free/reduced	none	73	74	72.0
26	male	group B	some college	standard	none	69	54	55.0
27	female	group C	bachelor's degree	standard	none	67	69	75.0
28	male	group C	high school	standard	none	70	70	65.0
29	female	group D	master's degree	standard	none	62	70	75.0
...
69	female	group C	associate's degree	standard	none	39	64	57.0
70	female	group D	some college	free/reduced	completed	58	63	73.0
71	male	group D	some college	standard	completed	63	55	63.0
72	female	group A	associate's degree	free/reduced	none	41	51	NaN
73	male	group C	some high school	free/reduced	none	61	57	56.0
74	male	group C	some high school	standard	none	49	49	41.0
75	male	group B	associate's degree	free/reduced	none	44	41	38.0
76	male	group E	some high school	standard	none	30	26	22.0
77	male	group A	bachelor's degree	standard	completed	80	78	81.0
78	female	group D	some high school	standard	completed	61	74	72.0
79	female	group E	master's degree	standard	none	62	68	68.0
80	female	group B	associate's degree	standard	none	47	49	50.0
81	male	group B	high school	free/reduced	none	49	45	45.0
82	male	group A	some college	free/reduced	completed	50	47	54.0
83	male	group E	associate's degree	standard	none	72	64	63.0
84	male	group D	high school	free/reduced	none	42	39	34.0
85	female	group C	some college	standard	none	73	80	82.0
86	female	group C	some college	free/reduced	none	76	83	88.0
87	female	group D	associate's degree	standard	none	71	71	74.0
88	female	group A	some college	standard	none	58	70	67.0

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
89	female	group D	some high school	standard	none	73	86	82.0
90	female	group C	bachelor's degree	standard	none	65	72	NaN
91	male	group C	high school	free/reduced	none	27	34	36.0
92	male	group C	high school	standard	none	71	79	NaN
93	male	group C	associate's degree	free/reduced	completed	43	66	50.0
94	female	group B	some college	standard	none	79	86	92.0
95	male	group C	associate's degree	free/reduced	completed	78	81	82.0
96	male	group B	some high school	standard	completed	65	66	NaN
97	female	group E	some college	standard	completed	63	72	70.0
98	female	group D	some college	free/reduced	none	58	67	NaN

99 rows × 8 columns

In [81]:

```
df.head()
```

Out[81]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74.0
1	female	group C	some college	standard	completed	69	90	88.0
2	female	group B	master's degree	standard	none	90	95	93.0
3	male	group A	associate's degree	free/reduced	none	47	57	44.0
4	male	group C	some college	standard	none	76	66	75.0

In [82]:

```
df.tail()
```

Out[82]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
94	female	group B	some college	standard	none	79	86	92.0
95	male	group C	associate's degree	free/reduced	completed	78	81	82.0
96	male	group B	some high school	standard	completed	65	66	62.0
97	female	group E	some college	standard	completed	63	72	70.0
98	female	group D	some college	free/reduced	none	58	67	62.0

In [83]:

```
df.isnull()          #checking the null values
```

Out[83]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	True
10	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	False	False
18	False	False	False	False	False	False	False	True
19	False	False	False	False	False	False	False	False
20	False	False	False	False	False	False	False	False
21	False	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False
23	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False
25	False	False	False	False	False	False	False	False
26	False	False	False	False	False	False	False	False
27	False	False	False	False	False	False	False	False
28	False	False	False	False	False	False	False	False
29	False	False	False	False	False	False	False	False
...
69	False	False	False	False	False	False	False	False
70	False	False	False	False	False	False	False	False
71	False	False	False	False	False	False	False	False
72	False	False	False	False	False	False	False	True
73	False	False	False	False	False	False	False	False

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
74	False	False	False	False	False	False	False	False
75	False	False	False	False	False	False	False	False
76	False	False	False	False	False	False	False	False
77	False	False	False	False	False	False	False	False
78	False	False	False	False	False	False	False	False
79	False	False	False	False	False	False	False	False
80	False	False	False	False	False	False	False	False
81	False	False	False	False	False	False	False	False
82	False	False	False	False	False	False	False	False
83	False	False	False	False	False	False	False	False
84	False	False	False	False	False	False	False	False
85	False	False	False	False	False	False	False	False
86	False	False	False	False	False	False	False	False
87	False	False	False	False	False	False	False	False
88	False	False	False	False	False	False	False	False
89	False	False	False	False	False	False	False	False
90	False	False	False	False	False	False	False	True
91	False	False	False	False	False	False	False	False
92	False	False	False	False	False	False	False	True
93	False	False	False	False	False	False	False	False
94	False	False	False	False	False	False	False	False
95	False	False	False	False	False	False	False	False
96	False	False	False	False	False	False	False	False
97	False	False	False	False	False	False	False	False
98	False	False	False	False	False	False	False	False

99 rows × 8 columns

In [84]:

```
df.isnull().sum()
```

Out[84]:

```
gender                0
race/ethnicity        0
parental level of education  0
lunch                 0
test preparation course  0
math score            0
reading score         0
writing score         5
dtype: int64
```

In [85]:

```
df.describe()
```

Out[85]:

	math score	reading score	writing score
count	99.000000	99.000000	94.000000
mean	60.616162	64.626263	63.563830
std	16.057745	15.744507	16.898532
min	0.000000	17.000000	10.000000
25%	50.000000	55.000000	54.000000
50%	62.000000	66.000000	65.000000
75%	71.000000	74.000000	75.000000
max	97.000000	95.000000	93.000000

In [86]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 8 columns):
gender                99 non-null object
race/ethnicity        99 non-null object
parental level of education  99 non-null object
lunch                 99 non-null object
test preparation course  99 non-null object
math score            99 non-null int64
reading score         99 non-null int64
writing score         94 non-null float64
dtypes: float64(1), int64(2), object(5)
memory usage: 4.3+ KB
```

In [95]:

```
df['writing score'] = df['writing score'].fillna(0)
df.tail()
```

Out[95]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
94	female	group B	some college	standard	none	79	86	92.0
95	male	group C	associate's degree	free/reduced	completed	78	81	82.0
96	male	group B	some high school	standard	completed	65	66	0.0
97	female	group E	some college	standard	completed	63	72	70.0
98	female	group D	some college	free/reduced	none	58	67	0.0

In [96]:

```
grouped=df.groupby("gender") #using group by function
```

In [97]:

```
grouped.first()
```

Out[97]:

	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
gender							
female	group B	bachelor's degree	standard	none	72	72	74.0
male	group A	associate's degree	free/reduced	none	47	57	44.0

In [99]:

```
non_duplicate = df[~df.duplicated('math score')]
non_duplicate.head()           #non-duplicate values
```

Out[99]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74.0
1	female	group C	some college	standard	completed	69	90	88.0
2	female	group B	master's degree	standard	none	90	95	93.0
3	male	group A	associate's degree	free/reduced	none	47	57	44.0
4	male	group C	some college	standard	none	76	66	75.0

In [100]:

```
df['lunch'] = df['lunch'].map({'standard': 0,
'free/reduced':1 }).astype(float)
df.head()           #reshaped data
```

Out[100]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	0.0	none	72	72	74.0
1	female	group C	some college	0.0	completed	69	90	88.0
2	female	group B	master's degree	0.0	none	90	95	93.0
3	male	group A	associate's degree	1.0	none	47	57	44.0
4	male	group C	some college	0.0	none	76	66	75.0