In [3]:
```
print(df)
```

```
     gender race/ethnicity parental level of education         lunch  \
0    female        group B           bachelor's degree      standard
1    female        group C                some college      standard
2    female        group B             master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E             master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test preparation course  math score  reading score  writing score
0                      none          72             72             74
1                 completed          69             90             88
2                      none          90             95             93
3                      none          47             57             44
4                      none          76             78             75
..                      ...         ...            ...            ...
995               completed          88             99             95
996                    none          62             55             55
997               completed          59             71             65
998               completed          68             78             77
999                    none          77             86             86
```

In [4]:
```
df.head(10)
```

Out[4]:

	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
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50

In [6]:
```
print(df.describe())
```

```
       math score  reading score  writing score
count   1000.00000    1000.000000    1000.000000
mean      66.08900      69.169000      68.054000
std       15.16308      14.600192      15.195657
min        0.00000      17.000000      10.000000
25%       57.00000      59.000000      57.750000
50%       66.00000      70.000000      69.000000
75%       77.00000      79.000000      79.000000
max      100.00000     100.000000     100.000000
```

In [7]:
```
df.dtypes
```

Out[7]:
```
gender                         object
race/ethnicity                 object
parental level of education    object
lunch                          object
test preparation course        object
math score                      int64
reading score                   int64
writing score                   int64
dtype: object
```

In [8]:
```
df.dropna(axis=1)
```

Out[8]:

	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
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95

```
In [9]: y=df.iloc[ : ,0:1]

In [10]: print(y)
```

```
     gender
0    female
1    female
2    female
3      male
4      male
..      ...
995  female
996    male
997  female
998  female
999  female

[1000 rows x 1 columns]
```

In [11]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
y = le.fit_transform(y)
print(y)
```

```
[0 0 0 1 1 0 0 1 1 0 1 1 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 0 0 0 1 1 1 0
 0 0 1 1 0 0 1 0 1 0 0 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 0 0 1 1 0 1 0 0 1 0 1
 1 1 1 1 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 0 1 1 0 0 0 1 1 0 1 1 0 0 1 0 0 0
 1 1 0 0 1 0 0 0 0 0 1 0 1 1 0 1 1 1 0 1 1 1 0 1 1 1 1 0 1 0 0 0 1 1 0 1 1
 0 1 1 1 0 1 1 0 0 1 0 1 1 0 1 1 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 1
 1 1 1 1 0 0 1 0 1 0 1 1 1 0 0 0 0 1 0 1 1 1 1 0 0 1 1 0 1 1 1 0 0 1 1 0 1
 0 1 0 0 0 1 1 0 1 1 0 1 1 1 1 0 1 1 1 0 0 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 0
 0 0 1 0 0 1 1 0 0 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 0 0 1 1 0 1 1 1 1 1 0 1 1
 1 1 1 1 1 1 0 1 0 1 1 1 0 0 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 0 1 1 1 0 1 1 1
 1 0 0 1 1 0 0 1 0 0 1 1 0 1 0 1 1 0 1 0 0 0 0 1 0 1 0 0 1 0 0 1 1 1 1 0 0
 1 0 1 0 0 1 0 0 0 1 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 1 0 0 1 1 0 1 0 0 0 0 1
 0 0 1 0 1 1 1 0 1 1 1 1 1 0 0 0 0 1 0 1 1 1 1 1 0 1 0 1 1 1 1 1 1 0 0 0 0
 1 0 1 1 1 1 0 0 0 1 0 1 0 1 0 1 1 1 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 1 0 1 1
 0 1 1 0 1 1 0 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 1 1 1 0 0 0 0 0
 0 0 1 0 1 1 1 1 1 0 0 0 0 0 1 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 0 1 1
 0 0 1 0 1 0 0 1 0 1 1 0 0 1 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 1 0 0 0 1 1 0 0 0 0 0 1 1 1 0 0 0 0 1 0 1 0 0 0 0 1 1 1 0 1 1 1 1 0 1 1 1
 0 1 1 0 0 1 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 1 0 1 1 1 0 0 1 0
 0 0 1 1 0 1 0 0 0 0 0 0 1 1 0 1 1 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1
 0 0 1 1 1 1 0 1 0 0 1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 1 0 1 0 1 0 1 1 0 1 1
 1 0 0 0 1 1 1 1 0 1 1 1 1 0 1 0 1 1 0 1 0 0 1 0 1 0 0 1 0 1 1 1 0 0 1 0 0
 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1
 0 1 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1 1 1 0 1 1
```

In [12]:
```
print(df['race/ethnicity'].value_counts())
```

```
group C    319
group D    262
group B    190
group E    140
group A     89
Name: race/ethnicity, dtype: int64
```

In [13]:
```
df_lunch = pd.get_dummies(df['lunch'])
df_new = pd.concat([df, df_lunch], axis=1)
print(df_new)
```

```
     gender race/ethnicity parental level of education         lunch  \
0    female        group B           bachelor's degree      standard
1    female        group C                some college      standard
2    female        group B             master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E             master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test preparation course  math score  reading score  writing score  \
0                      none          72             72             74
1                 completed          69             90             88
2                      none          90             95             93
3                      none          47             57             44
4                      none          76             78             75
..                      ...         ...            ...            ...
995               completed          88             99             95
996                    none          62             55             55
997               completed          59             71             65
```