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

In [3]:
```
df = pd.read_csv('StudentsPerformance.csv')
```

In [4]:
```
df.head(n=5)
```

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	72.0	72.0	74.0
1	female	group C	some college	standard	completed	69.0	90.0	88.0
2	female	group B	master's degree	standard	none	90.0	95.0	93.0
3	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	male	group C	some college	standard	none	76.0	78.0	75.0

In [5]:
```
df.tail(n=5)
```

Out[5]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
995	female	group E	master's degree	standard	completed	88.0	99.0	95.0
996	male	group C	high school	free/reduced	none	62.0	55.0	55.0
997	female	group C	high school	free/reduced	completed	59.0	71.0	65.0
998	female	group D	some college	standard	completed	68.0	78.0	77.0
999	female	group D	some college	free/reduced	none	77.0	86.0	86.0

In [6]:
```
df.columns
```

Out[6]:
```
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
       'test preparation course', 'math score', 'reading score',
       'writing score'],
      dtype='object')
```

In [7]:
```
df.describe(include='all')
```

Out[7]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
count	1000	1000	1000	1000	1000	998.000000	996.000000	998.000000
unique	2	5	6	2	2	NaN	NaN	NaN

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
top	female	group C	some college	standard	none	NaN	NaN	NaN
freq	518	319	226	645	642	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	66.093186	69.186747	68.046092
std	NaN	NaN	NaN	NaN	NaN	15.177728	14.623079	15.208213
min	NaN	NaN	NaN	NaN	NaN	0.000000	17.000000	10.000000
25%	NaN	NaN	NaN	NaN	NaN	57.000000	59.000000	57.250000
50%	NaN	NaN	NaN	NaN	NaN	66.000000	70.000000	69.000000
75%	NaN	NaN	NaN	NaN	NaN	77.000000	79.250000	79.000000
max	NaN	NaN	NaN	NaN	NaN	100.000000	100.000000	100.000000

In [8]:
```
df.dtypes
```

Out[8]:
```
gender                         object
race/ethnicity                 object
parental level of education    object
lunch                          object
test preparation course        object
math score                    float64
reading score                 float64
writing score                 float64
dtype: object
```

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

Out[9]:

	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
...	...	...	...	...	...	...	...	...
995	False	False	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False

1000 rows × 8 columns

In [10]:
```
df.isnull().any()
```

Out[10]:
```
gender                         False
race/ethnicity                 False
parental level of education    False
lunch                          False
test preparation course        False
math score                      True
reading score                   True
writing score                   True
dtype: bool
```

In [11]:
```
df.isnull().sum().sum()
```

Out[11]:
```
8
```

In [12]:
```
df.isnull().sum(axis = 1)
```

Out[12]:
```
0      0
1      0
2      0
3      0
4      0
      ..
995    0
996    0
997    0
998    0
999    0
Length: 1000, dtype: int64
```

In [13]:
```
df.size
```

Out[13]:
```
8000
```

In [14]:
```
df.ndim
```

Out[14]:
```
2
```

In [15]:
```
df.dtypes
```

Out[15]:
```
gender                         object
race/ethnicity                 object
parental level of education    object
lunch                          object
test preparation course        object
math score                    float64
reading score                 float64
writing score                 float64
dtype: object
```

In [16]:
```
dummies=pd.get_dummies(df.gender)
merged = pd.concat([df, dummies], axis = 'columns')
merged.drop(['gender'], axis = 'columns')
```

Out[16]:

	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	female	male
0	group B	bachelor's degree	standard	none	72.0	72.0	74.0	1	0
1	group C	some college	standard	completed	69.0	90.0	88.0	1	0
2	group B	master's degree	standard	none	90.0	95.0	93.0	1	0
3	group A	associate's degree	free/reduced	none	47.0	57.0	44.0	0	1
4	group C	some college	standard	none	76.0	78.0	75.0	0	1
...	...	...	...	...	...	...	...	...	...
995	group E	master's degree	standard	completed	88.0	99.0	95.0	1	0
996	group C	high school	free/reduced	none	62.0	55.0	55.0	0	1
997	group C	high school	free/reduced	completed	59.0	71.0	65.0	1	0
998	group D	some college	standard	completed	68.0	78.0	77.0	1	0
999	group D	some college	free/reduced	none	77.0	86.0	86.0	1	0

1000 rows × 9 columns

In [17]:
```
merged
```

Out[17]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	female	r
0	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0	1	
1	female	group C	some college	standard	completed	69.0	90.0	88.0	1	
2	female	group B	master's degree	standard	none	90.0	95.0	93.0	1	
3	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0	0	
4	male	group C	some college	standard	none	76.0	78.0	75.0	0	
...	...	...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88.0	99.0	95.0	1	
996	male	group C	high school	free/reduced	none	62.0	55.0	55.0	0	

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	female	r
997	female	group C	high school	free/reduced	completed	59.0	71.0	65.0	1	
998	female	group D	some college	standard	completed	68.0	78.0	77.0	1	
999	female	group D	some college	free/reduced	none	77.0	86.0	86.0	1	

1000 rows × 10 columns

In [ ]: