

# Importing Libraries

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

## Upload Data

```
In [2]: data=pd.read_csv("Student.csv", sep=",")
```

```
In [3]: data.head()
```

```
Out[3]:
```

	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

```
In [4]: data.tail()
```

```
Out[4]:
```

	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	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

```
In [10]: data.shape
```

```
Out[10]: (1000, 8)
```

```
In [5]: data.describe()
```

```
Out[5]:
```

	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 [6]: data.dtypes
```

```
Out[6]: 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]: data.size
```

```
Out[8]: 8000
```

```
In [11]: data.columns
```

```
Out[11]: Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
               'test preparation course', 'math score', 'reading score',
               'writing score'],
              dtype='object')
```

# Checking Unique Value

```
In [12]: data.nunique()
```

```
Out[12]: gender                2
race/ethnicity                5
parental level of education    6
lunch                        2
test preparation course        2
math score                    81
reading score                  72
writing score                  77
dtype: int64
```

```
In [14]: data['gender'].unique()
```

```
Out[14]: array(['female', 'male'], dtype=object)
```

```
In [15]: data['race/ethnicity'].unique()
```

```
Out[15]: array(['group B', 'group C', 'group A', 'group D', 'group E'],
dtype=object)
```

## Cleaning The Data

```
In [16]: data.isnull().sum()
```

```
Out[16]: gender                0
race/ethnicity                0
parental level of education    0
lunch                        0
test preparation course        0
math score                    0
reading score                  0
writing score                  0
dtype: int64
```

## Remove Columns These Are Not Important Value For Any Valuation

```
In [17]: df = data.drop(['race/ethnicity','parental level of education'],axis = 1)
```

```
In [18]: df.head()
```

```
Out[18]:
```

	gender	lunch	test preparation course	math score	reading score	writing score
0	female	standard	none	72	72	74
1	female	standard	completed	69	90	88
2	female	standard	none	90	95	93
3	male	free/reduced	none	47	57	44
4	male	standard	none	76	78	75

## RELATIONSHIP ANALYSIS

```
In [20]: corr = df.corr()
```

```
In [21]: sns.heatmap(corr,xticklabels=corr.columns,yticklabels=corr.columns,
annot = True)
```

```
Out[21]: <AxesSubplot:>
```

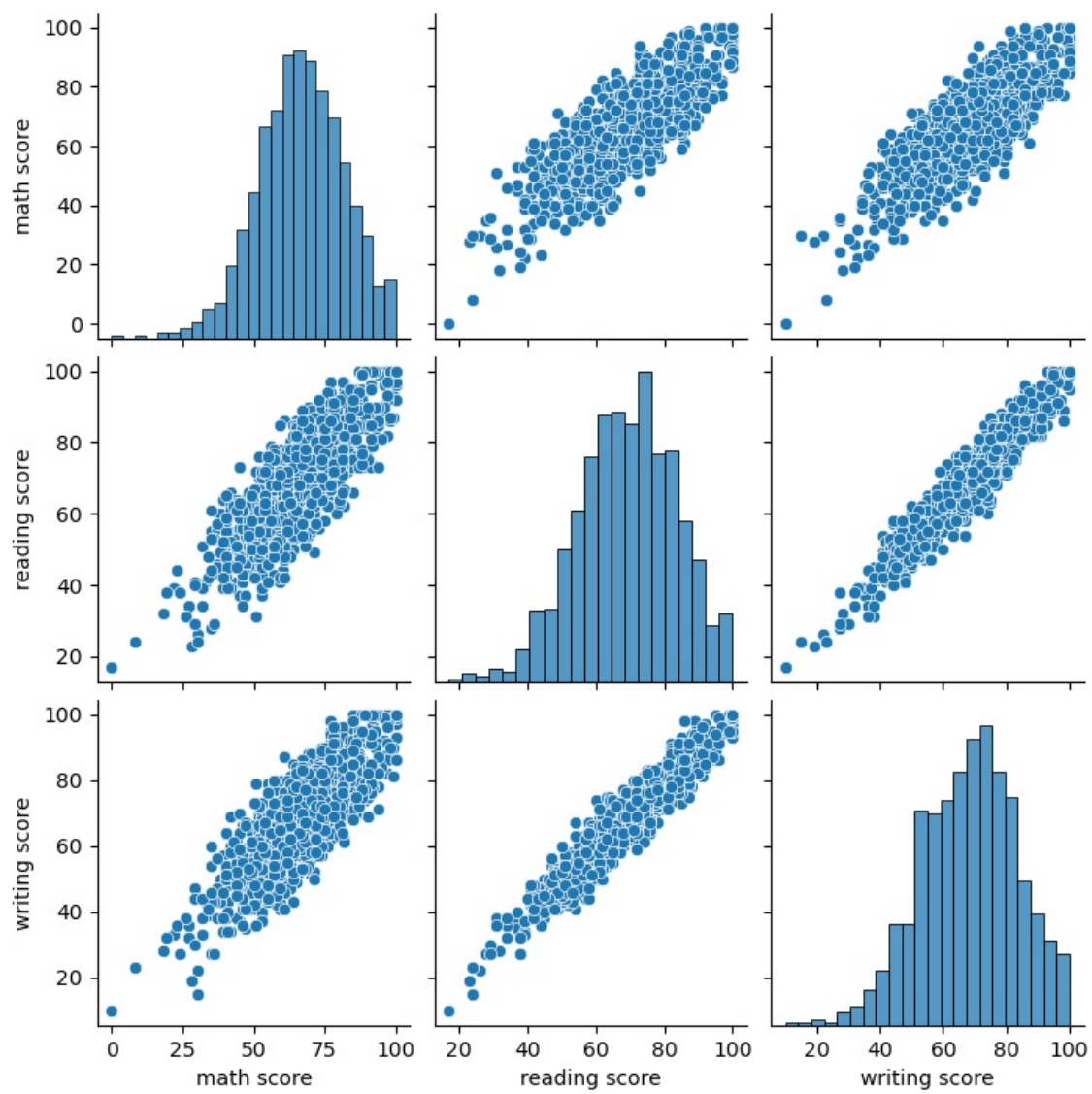


## TO UNDERSTAND THE RELATION BETWEEN DIFFERENT VARIABLES

WHICH IS MATH SCORE, READING SCORE & WRITING SCORE

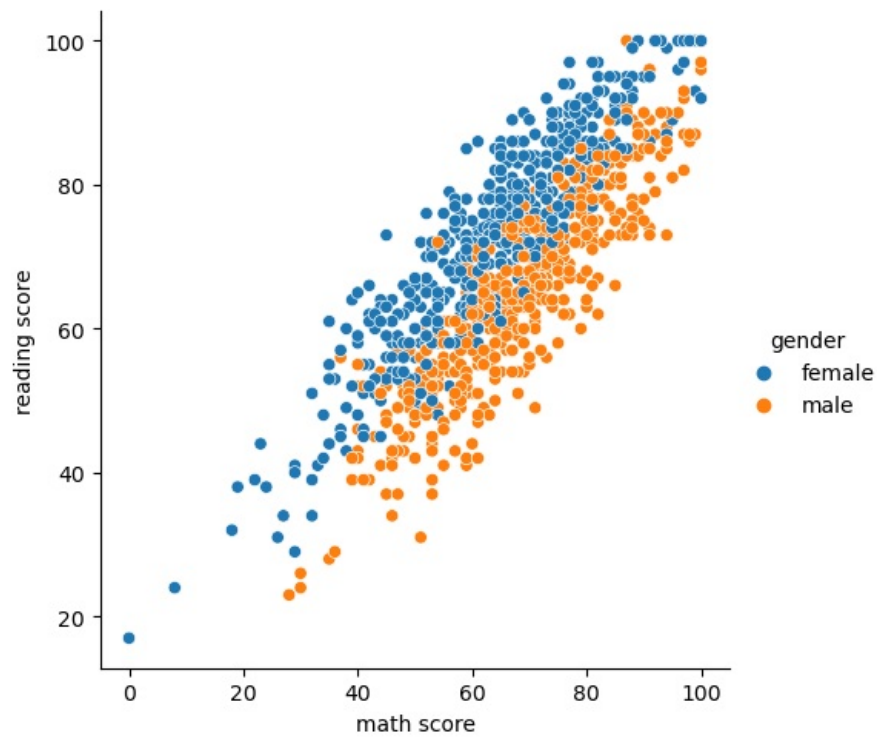
```
In [22]: sns.pairplot(df)
```

```
Out[22]: <seaborn.axisgrid.PairGrid at 0x21997bafb20>
```



```
In [26]: sns.relplot(x='math score',y='reading score',hue='gender',data = df)
```

```
Out[26]: <seaborn.axisgrid.FacetGrid at 0x2199863e4c0>
```

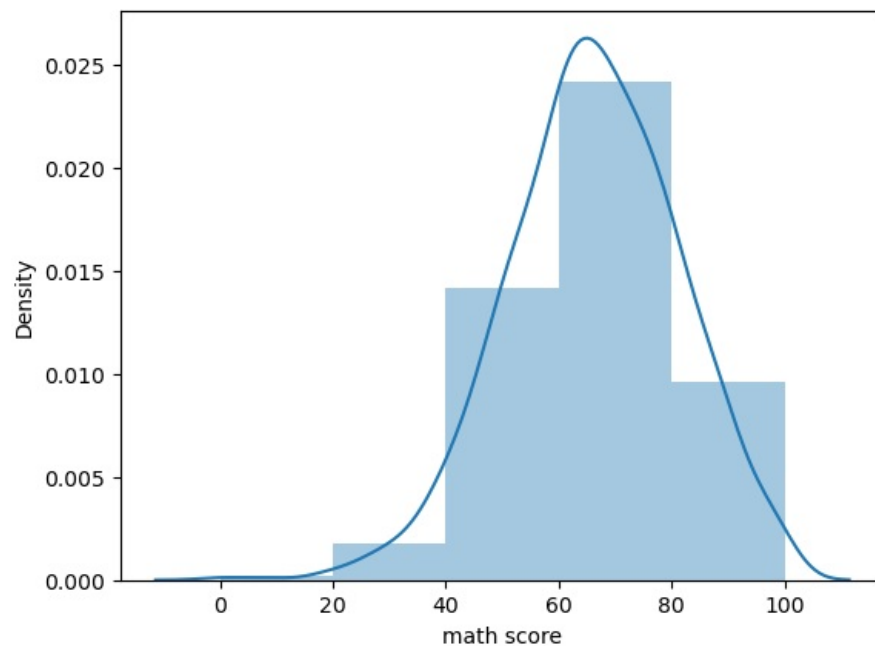


```
In [31]: sns.distplot(df['math score'],bins=5)
```

C:\Users\SAI COMPUTERS\Contacts\anaconda3 new\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

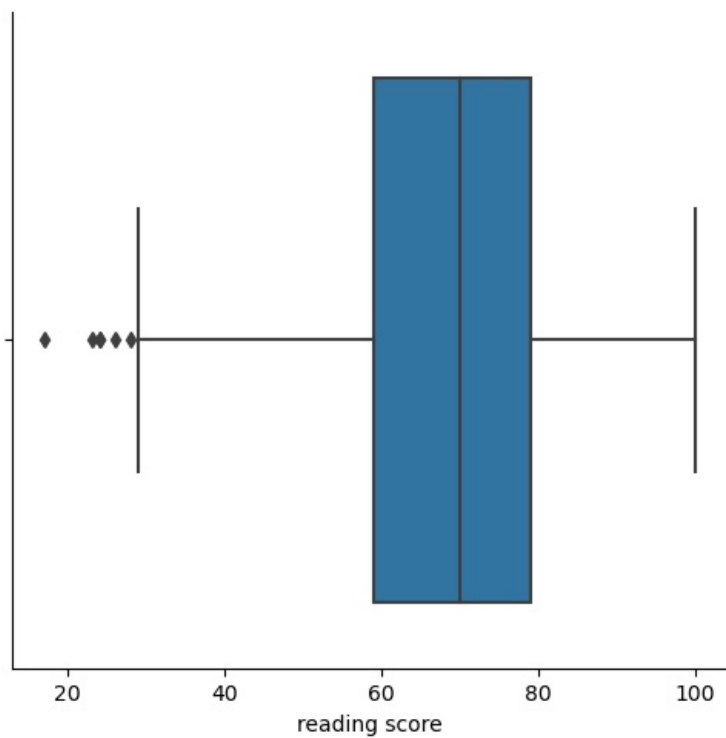
warnings.warn(msg, FutureWarning)

```
Out[31]: <AxesSubplot:xlabel='math score', ylabel='Density'>
```



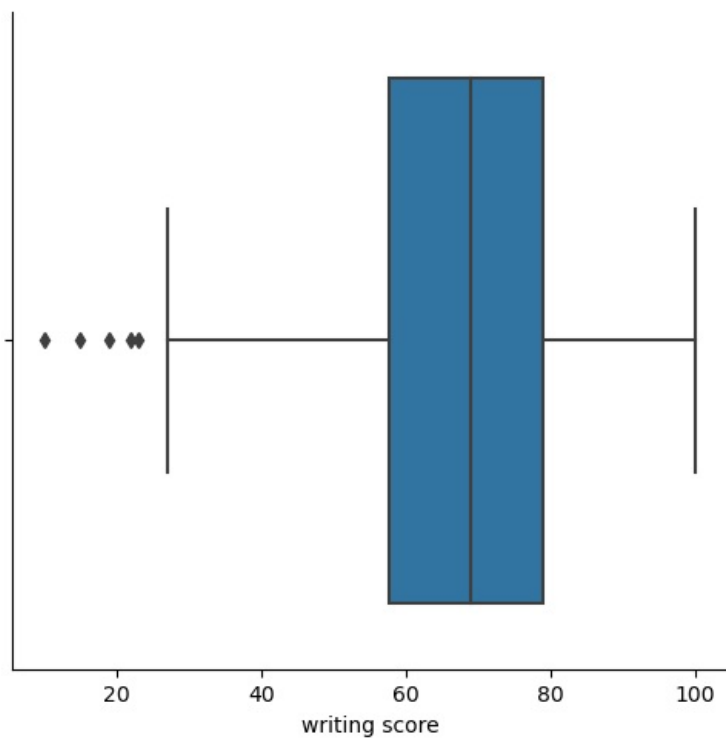
```
In [33]: sns.catplot(x='reading score',kind='box',data=df)
```

```
Out[33]: <seaborn.axisgrid.FacetGrid at 0x21998999e80>
```



```
In [34]: sns.catplot(x='writing score',kind='box',data=df)
```

```
Out[34]: <seaborn.axisgrid.FacetGrid at 0x21999dde8b0>
```



```
In [ ]:
```