In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

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

In [3]:
```
df = pd.read_csv(r"C:\Users\tejas\Downloads\StudentsPerformance.csv")
```

In [4]:
```
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        72.0            72           74.0
1                 completed         NaN            90           88.0
2                      none        90.0            na           93.0
3                      none        47.0            57            NaN
4                      none        76.0            78           75.0
..                      ...         ...           ...           ...
995               completed        88.0            99           95.0
996                    none        62.0            55           55.0
997               completed        59.0            71           65.0
998               completed        68.0            78           77.0
999                    none        77.0            86           86.0

[1000 rows x 8 columns]
```

In [6]:
```
print(df['math score'])
```

```
0      72.0
1       NaN
2      90.0
3      47.0
4      76.0
       ...
995    88.0
996    62.0
997    59.0
998    68.0
999    77.0
Name: math score, Length: 1000, dtype: float64
```

In [8]:
```
print(df['math score'].isnull())
```

```
0      False
1       True
2      False
3      False
4      False
       ...
995    False
996    False
997    False
998    False
999    False
Name: math score, Length: 1000, dtype: bool
```

In [9]:
```
print(df['reading score'])
```

```
0      72
1      90
2      na
3      57
4      78
       ..
995    99
996    55
997    71
998    78
999    86
Name: reading score, Length: 1000, dtype: object
```

In [11]:
```
print(df['reading score'].isnull())
```

```
0      False
1      False
2      False
3      False
4      False
       ...
995    False
996    False
997    False
998    False
999    False
Name: reading score, Length: 1000, dtype: bool
```

In [12]:
```
missing_values = ['na','n/a','NA', '--','-','',' ','A']
df = pd.read_csv(r"C:\Users\bhagyashree\Downloads\StudentsPerformance.csv", na_va
```

In [13]:
```
print(df['reading score'].isnull())
```

```
0      False
1      False
2       True
3      False
4      False
       ...
995    False
996    False
997    False
998    False
999    False
Name: reading score, Length: 1000, dtype: bool
```

In [14]:
```
dataset =[11,41,20,3,101,55,68,97,99,6]
```

In [15]:
```
sorted(dataset)
```

Out[15]:
```
[3, 6, 11, 20, 41, 55, 68, 97, 99, 101]
```

In [16]:
```
quantile1, quantile3 = np.percentile(dataset,[25,75])
```

In [17]:
```
print(quantile1, quantile3)
```

```
13.25 89.75
```

In [18]:
```
iqr_value = quantile3 - quantile1
print(iqr_value)
```

```
76.5
```

In [19]:
```
lower_bound_val=quantile1-(1.5*iqr_value)
```

In [22]:
```
upper_bound_val=quantile3+(1.5*iqr_value)
```

In [23]:
```
print(lower_bound_val,upper_bound_val)
```

```
-101.5 204.5
```

In [9]:
```
df = pd.read_csv(r"C:\Users\tejas\Downloads\StudentsPerformance (2).csv")
```

In [10]:
```
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
...      ...            ...                         ...           ...
2235     NaN            NaN                         NaN           NaN
2236     NaN            NaN                         NaN           NaN
2237     NaN            NaN                         NaN           NaN
2238     NaN            NaN                         NaN           NaN
2239     NaN            NaN                         NaN           NaN

     test preparation course  Year_Birth  math score reading score  \
0                       none      1970.0        72.0            72
1                  completed      1961.0         NaN            na
2                       none      1958.0        90.0            95
3                       none      1967.0         NaN           NaN
4                       none      1989.0        76.0            78
...                      ...         ...         ...           ...
2235                     NaN         NaN         NaN           NaN
2236                     NaN         NaN         NaN           NaN
2237                     NaN         NaN         NaN           NaN
2238                     NaN         NaN         NaN           NaN
2239                     NaN         NaN         NaN           NaN

     writing score Dt_Admission College_Fees
0               74      6/16/14  $84,835.00
1                A      6/15/14  $57,091.00
2               93      5/13/14  $67,267.00
3               44   05-11-2014  $32,474.00
4               75   04-08-2014  $21,474.00
...            ...          ...         ...
2235           NaN          NaN         NaN
2236           NaN          NaN         NaN
2237           NaN          NaN         NaN
2238           NaN          NaN         NaN
2239           NaN          NaN         NaN

[2240 rows x 11 columns]
```

In [11]:
```
from datetime import date
df['age'] = date.today().year - df['Year_Birth']
```

In [12]:
```
df.head(5)
```

Out[12]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	Year_Birth	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	1970.0	72.0	72	74
1	female	group C	some college	standard	completed	1961.0	NaN	na	A
2	female	group B	master's degree	standard	none	1958.0	90.0	95	93
3	male	group A	associate's degree	free/reduced	none	1967.0	NaN	NaN	44
4	male	group C	some college	standard	none	1989.0	76.0	78	75

In [13]:
```
df['Year'] = pd.DatetimeIndex(df['Dt_Admission']).year
df['E_L'] = date.today().year - df['Year']
```

In [14]:
```
df.head(5)
```

Out[14]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	Year_Birth	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	1970.0	72.0	72	74
1	female	group C	some college	standard	completed	1961.0	NaN	na	A
2	female	group B	master's degree	standard	none	1958.0	90.0	95	93
3	male	group A	associate's degree	free/reduced	none	1967.0	NaN	NaN	44
4	male	group C	some college	standard	none	1989.0	76.0	78	75

In [15]:
```
df['Fees$'] = df['College_Fees'].str.replace(',','').str.replace('$','').str.repl
df['Fees_M$'] = df['Fees$'].apply(lambda X:round(X/1000000))
```

In [16]:
```
df.head(5)
```

Out[16]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	Year_Birth	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	1970.0	72.0	72	74
1	female	group C	some college	standard	completed	1961.0	NaN	na	A
2	female	group B	master's degree	standard	none	1958.0	90.0	95	93
3	male	group A	associate's degree	free/reduced	none	1967.0	NaN	NaN	44
4	male	group C	some college	standard	none	1989.0	76.0	78	75

In [ ]:
```

```