In [10]:

```
import pandas as pd
import numpy as np

df=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")

df.head()
df
```

Out[10]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.0	90.0	77.0	88.0	
1	Female	5.0	82.0	78.0	93.0	
2	Female	170.0	100.0	74.0	32.0	
3	Male	200.0	89.0	32.0	NaN	
4	NaN	1.0	79.0	67.0	100.0	
5	Female	56.0	79.0	33.0	80.0	
6	Male	16.0	94.0	NaN	80.0	
7	Female	NaN	90.0	68.0	83.0	
8	Male	36.0	50.0	72.0	90.0	
9	Male	69.0	90.0	79.0	78.0	
10	Female	0.0	83.0	74.0	89.0	
11	Female	6.0	87.0	79.0	98.0	
12	Male	4.0	87.0	75.0	93.0	
13	Male	150.0	43.0	65.0	21.0	
14	Male	76.0	75.0	62.0	78.0	
15	Female	3.0	46.0	74.0	93.0	
16	Male	103.0	NaN	62.0	94.0	
17	Male	26.0	78.0	100.0	87.0	
18	Male	78.0	95.0	75.0	11.0	
19	Female	82.0	82.0	NaN	76.0	
20	Female	17.0	82.0	69.0	85.0	

In [19]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['Gender'] = le.fit_transform(df['Gender'])
newdf=df
df
```

Out[19]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	0	15.0	90.0	77.0	88.0	2	Nashik
1	0	5.0	82.0	78.0	93.0	2	Pune
2	0	93.0	93.0	74.0	85.0	2	Nashik
3	1	34.0	89.0	67.0	NaN	1	Nashik
4	2	1.0	79.0	67.0	100.0	3	Nashik
5	0	56.0	79.0	70.0	80.0	2	Nashik
6	1	16.0	94.0	NaN	80.0	2	Pune
7	0	NaN	90.0	68.0	83.0	2	Pune
8	1	36.0	82.0	72.0	90.0	2	Nashik
9	1	69.0	90.0	79.0	78.0	1	Nashik
10	0	0.0	83.0	74.0	89.0	2	Pune
11	0	6.0	87.0	79.0	98.0	3	Nashik
12	1	4.0	87.0	75.0	93.0	2	Pune
13	1	58.0	83.0	65.0	78.0	1	Pune
14	1	76.0	75.0	62.0	78.0	1	NaN
15	0	3.0	84.0	74.0	93.0	2	Nashik
16	1	39.0	NaN	62.0	94.0	2	Nashik
17	1	26.0	78.0	70.0	87.0	2	Nashik
18	1	78.0	95.0	75.0	81.0	2	Nashik
19	0	82.0	82.0	NaN	76.0	1	Pune
20	0	17.0	82.0	69.0	85.0	2	Nashik

In [20]:

```
missing_values = ["Na", "na"]
df = pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")
df
```

Out[20]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	Female	15.0	90.0	77.0	88.0	2	Nashik
1	Female	5.0	82.0	78.0	93.0	2	Pune
2	Female	93.0	93.0	74.0	85.0	2	Nashik
3	Male	34.0	89.0	67.0	NaN	1	Nashik
4	NaN	1.0	79.0	67.0	100.0	3	Nashik
5	Female	56.0	79.0	70.0	80.0	2	Nashik
6	Male	16.0	94.0	NaN	80.0	2	Pune
7	Female	NaN	90.0	68.0	83.0	2	Pune
8	Male	36.0	82.0	72.0	90.0	2	Nashik
9	Male	69.0	90.0	79.0	78.0	1	Nashik
10	Female	0.0	83.0	74.0	89.0	2	Pune
11	Female	6.0	87.0	79.0	98.0	3	Nashik
12	Male	4.0	87.0	75.0	93.0	2	Pune
13	Male	58.0	83.0	65.0	78.0	1	Pune
14	Male	76.0	75.0	62.0	78.0	1	NaN
15	Female	3.0	84.0	74.0	93.0	2	Nashik
16	Male	39.0	NaN	62.0	94.0	2	Nashik
17	Male	26.0	78.0	70.0	87.0	2	Nashik
18	Male	78.0	95.0	75.0	81.0	2	Nashik
19	Female	82.0	82.0	NaN	76.0	1	Pune
20	Female	17.0	82.0	69.0	85.0	2	Nashik

In [12]:

```
ndf = df
ndf.fillna(0)
```

Out[12]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.0	90.0	77.0	88.0	8
1	Female	5.0	82.0	78.0	93.0	6
2	Female	170.0	100.0	74.0	32.0	2
3	Male	200.0	89.0	32.0	0.0	1
4	0	1.0	79.0	67.0	100.0	3
5	Female	56.0	79.0	33.0	80.0	2
6	Male	16.0	94.0	0.0	80.0	2
7	Female	0.0	90.0	68.0	83.0	3
8	Male	36.0	50.0	72.0	90.0	2
9	Male	69.0	90.0	79.0	78.0	1
10	Female	0.0	83.0	74.0	89.0	2
11	Female	6.0	87.0	79.0	98.0	3
12	Male	4.0	87.0	75.0	93.0	2
13	Male	150.0	43.0	65.0	21.0	1
14	Male	76.0	75.0	62.0	78.0	1
15	Female	3.0	46.0	74.0	93.0	2
16	Male	103.0	0.0	62.0	94.0	1
17	Male	26.0	78.0	100.0	87.0	2
18	Male	78.0	95.0	75.0	11.0	2
19	Female	82.0	82.0	0.0	76.0	0
20	Female	17.0	82.0	69.0	85.0	2

In [20]:

```
m_v = df['Math_score'].mean()
df['Math_score'].fillna(value=m_v, inplace=True)
df
```

Out[20]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.00	90.0	77.0	88.0	8
1	Female	5.00	82.0	78.0	93.0	6
2	Female	170.00	100.0	74.0	32.0	2
3	Male	200.00	89.0	32.0	NaN	1
4	NaN	1.00	79.0	67.0	100.0	3
5	Female	56.00	79.0	33.0	80.0	2
6	Male	16.00	94.0	NaN	80.0	2
7	Female	55.65	90.0	68.0	83.0	3
8	Male	36.00	50.0	72.0	90.0	2
9	Male	69.00	90.0	79.0	78.0	1
10	Female	0.00	83.0	74.0	89.0	2
11	Female	6.00	87.0	79.0	98.0	3
12	Male	4.00	87.0	75.0	93.0	2
13	Male	150.00	43.0	65.0	21.0	1
14	Male	76.00	75.0	62.0	78.0	1
15	Female	3.00	46.0	74.0	93.0	2
16	Male	103.00	NaN	62.0	94.0	1
17	Male	26.00	78.0	100.0	87.0	2
18	Male	78.00	95.0	75.0	11.0	2
19	Female	82.00	82.0	NaN	76.0	0
20	Female	17.00	82.0	69.0	85.0	2

In [23]:

```
ndf.replace(to_replace = np.nan, value = -99)
```

Out[23]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	Female	15.0	90.0	77.0	88.0	2	Nashik
1	Female	5.0	82.0	78.0	93.0	2	Pune
2	Female	93.0	93.0	74.0	85.0	2	Nashik
3	Male	34.0	89.0	67.0	-99.0	1	Nashik
4	-99	1.0	79.0	67.0	100.0	3	Nashik
5	Female	56.0	79.0	70.0	80.0	2	Nashik
6	Male	16.0	94.0	-99.0	80.0	2	Pune
7	Female	35.7	90.0	68.0	83.0	2	Pune
8	Male	36.0	82.0	72.0	90.0	2	Nashik
9	Male	69.0	90.0	79.0	78.0	1	Nashik
10	Female	0.0	83.0	74.0	89.0	2	Pune
11	Female	6.0	87.0	79.0	98.0	3	Nashik
12	Male	4.0	87.0	75.0	93.0	2	Pune
13	Male	58.0	83.0	65.0	78.0	1	Pune
14	Male	76.0	75.0	62.0	78.0	1	-99
15	Female	3.0	84.0	74.0	93.0	2	Nashik
16	Male	39.0	-99.0	62.0	94.0	2	Nashik
17	Male	26.0	78.0	70.0	87.0	2	Nashik
18	Male	78.0	95.0	75.0	81.0	2	Nashik
19	Female	82.0	82.0	-99.0	76.0	1	Pune
20	Female	17.0	82.0	69.0	85.0	2	Nashik

In [24]:

```
ndf.dropna()
```

Out[24]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	Female	15.0	90.0	77.0	88.0	2	Nashik
1	Female	5.0	82.0	78.0	93.0	2	Pune
2	Female	93.0	93.0	74.0	85.0	2	Nashik
5	Female	56.0	79.0	70.0	80.0	2	Nashik
7	Female	35.7	90.0	68.0	83.0	2	Pune
8	Male	36.0	82.0	72.0	90.0	2	Nashik
9	Male	69.0	90.0	79.0	78.0	1	Nashik
10	Female	0.0	83.0	74.0	89.0	2	Pune
11	Female	6.0	87.0	79.0	98.0	3	Nashik
12	Male	4.0	87.0	75.0	93.0	2	Pune
13	Male	58.0	83.0	65.0	78.0	1	Pune
15	Female	3.0	84.0	74.0	93.0	2	Nashik
17	Male	26.0	78.0	70.0	87.0	2	Nashik
18	Male	78.0	95.0	75.0	81.0	2	Nashik
20	Female	17.0	82.0	69.0	85.0	2	Nashik

In [25]:

```
ndf.dropna(how = 'all')
```

Out[25]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	Female	15.0	90.0	77.0	88.0	2	Nashik
1	Female	5.0	82.0	78.0	93.0	2	Pune
2	Female	93.0	93.0	74.0	85.0	2	Nashik
3	Male	34.0	89.0	67.0	NaN	1	Nashik
4	NaN	1.0	79.0	67.0	100.0	3	Nashik
5	Female	56.0	79.0	70.0	80.0	2	Nashik
6	Male	16.0	94.0	NaN	80.0	2	Pune
7	Female	35.7	90.0	68.0	83.0	2	Pune
8	Male	36.0	82.0	72.0	90.0	2	Nashik
9	Male	69.0	90.0	79.0	78.0	1	Nashik
10	Female	0.0	83.0	74.0	89.0	2	Pune
11	Female	6.0	87.0	79.0	98.0	3	Nashik
12	Male	4.0	87.0	75.0	93.0	2	Pune
13	Male	58.0	83.0	65.0	78.0	1	Pune
14	Male	76.0	75.0	62.0	78.0	1	NaN
15	Female	3.0	84.0	74.0	93.0	2	Nashik
16	Male	39.0	NaN	62.0	94.0	2	Nashik
17	Male	26.0	78.0	70.0	87.0	2	Nashik
18	Male	78.0	95.0	75.0	81.0	2	Nashik
19	Female	82.0	82.0	NaN	76.0	1	Pune
20	Female	17.0	82.0	69.0	85.0	2	Nashik

In [26]:

```
ndf.dropna(axis = 1)
```

Out[26]:

	Math score	Placement offer count
0	15.0	2
1	5.0	2
2	93.0	2
3	34.0	1
4	1.0	3
5	56.0	2
6	16.0	2
7	35.7	2
8	36.0	2
9	69.0	1
10	0.0	2
11	6.0	3
12	4.0	2
13	58.0	1
14	76.0	1
15	3.0	2
16	39.0	2
17	26.0	2
18	78.0	2
19	82.0	1
20	17.0	2

In [27]:

```
new_data = ndf.dropna(axis = 0, how = 'any')
new_data
```

Out[27]:

	Gender	Math score	Reading score	Writing score	Placement score	Placement offer count	Region
0	Female	15.0	90.0	77.0	88.0	2	Nashik
1	Female	5.0	82.0	78.0	93.0	2	Pune
2	Female	93.0	93.0	74.0	85.0	2	Nashik
5	Female	56.0	79.0	70.0	80.0	2	Nashik
7	Female	35.7	90.0	68.0	83.0	2	Pune
8	Male	36.0	82.0	72.0	90.0	2	Nashik
9	Male	69.0	90.0	79.0	78.0	1	Nashik
10	Female	0.0	83.0	74.0	89.0	2	Pune
11	Female	6.0	87.0	79.0	98.0	3	Nashik
12	Male	4.0	87.0	75.0	93.0	2	Pune
13	Male	58.0	83.0	65.0	78.0	1	Pune
15	Female	3.0	84.0	74.0	93.0	2	Nashik
17	Male	26.0	78.0	70.0	87.0	2	Nashik
18	Male	78.0	95.0	75.0	81.0	2	Nashik
20	Female	17.0	82.0	69.0	85.0	2	Nashik

In [30]:

```
import pandas as pd
import numpy as np

df1 = pd.read_csv("C:\\Users\\Nancy\\Desktop\\heights.csv")
df1
```

Out[30]:

	name	height
0	mohan	5.9
1	maria	5.2
2	sakib	5.1
3	tao	5.5
4	virat	4.9
5	vaishnav	6.2
6	yash	5.6
7	vanshika	7.1
8	siddharth	14.2
9	taabish	6.5
10	john	6.1
11	joseph	1.2
12	celina	5.5

In [31]:

```
df1['height'].quantile(0.95)
```

Out[31]:

```
9.939999999999989
```

In [32]:

```
max_thresold = df1['height'].quantile(0.95)
max_thresold
```

Out[32]:

```
9.939999999999989
```

In [33]:

```
df1[df1['height']>max_thresold]
```

Out[33]:

	name	height
8	siddharth	14.2

In [34]:

```
min_thresold = df1['height'].quantile(0.05)
min_thresold
```

Out[34]:

```
3.420000000000001
```

In [35]:

```
df1[df1['height']<min_thresold]
```

Out[35]:

	name	height
11	joseph	1.2

In [36]:

```
df1[(df1['height']<max_thresold) & (df1['height']>min_thresold)]
```

Out[36]:

	name	height
0	mohan	5.9
1	maria	5.2
2	sakib	5.1
3	tao	5.5
4	virat	4.9
5	vaishnav	6.2
6	yash	5.6
7	vanshika	7.1
9	taabish	6.5
10	john	6.1
12	celina	5.5

In [37]:

```
df2 = df1[(df1['height']<max_thresold) & (df1['height']>min_thresold)]
df2.shape
```

Out[37]:

```
(11, 2)
```

In [38]:

```
df2.describe()
```

Out[38]:

	height
count	11.000000
mean	5.781818
std	0.656990
min	4.900000
25%	5.350000
50%	5.600000
75%	6.150000
max	7.100000

In [23]:

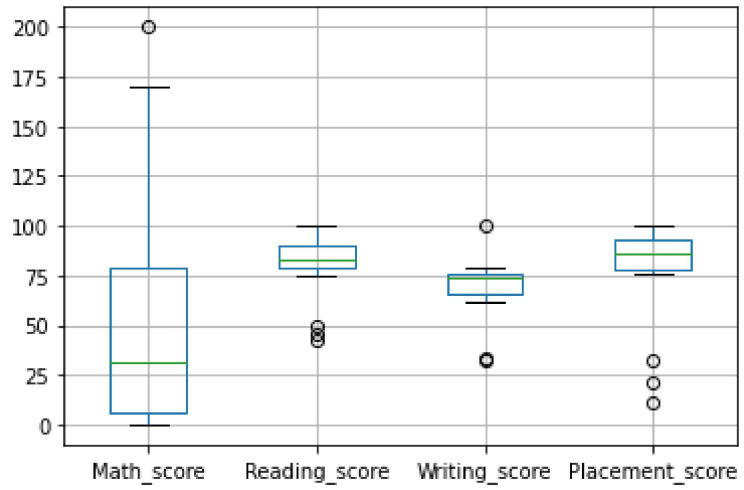
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")

col = ['Math_score', 'Reading_score', 'Writing_score', 'Placement_score']
df.boxplot(col)
```

Out[23]:

```
<AxesSubplot:>
```



In [27]:

```
print(np.where(df['Math_score']>90))
print(np.where(df['Reading_score']<75))
print(np.where(df['Writing_score']<50))
```

```
(array([ 2,  3, 13, 16], dtype=int64),)
(array([ 8, 13, 15], dtype=int64),)
(array([3, 5], dtype=int64),)
```

In [43]:

```
import pandas as pd
df = pd.read_csv("C:\\Users\\Nancy\\Desktop\\heights.csv")
df
```

Out[43]:

	name	height
0	mohan	5.9
1	maria	5.2
2	sakib	5.1
3	tao	5.5
4	virat	4.9
5	vaishnav	6.2
6	yash	5.6
7	vanshika	7.1
8	siddharth	14.2
9	taabish	6.5
10	john	6.1
11	joseph	1.2
12	celina	5.5

In [44]:

```
df.describe()
```

Out[44]:

	height
count	13.000000
mean	6.076923
std	2.814295
min	1.200000
25%	5.200000
50%	5.600000
75%	6.200000
max	14.200000

In [45]:

```
Q1 = df.height.quantile(0.25)
Q3 = df.height.quantile(0.75)
Q1, Q3
```

Out[45]:

```
(5.2, 6.2)
```

In [46]:

```
IQR = Q3 - Q1
IQR
```

Out[46]:

```
1.0
```

In [47]:

```
lower_limit = Q1 - 1.5*IQR
upper_limit = Q3 + 1.5*IQR
lower_limit, upper_limit
```

Out[47]:

```
(3.7, 7.7)
```

In [48]:

```
df[(df.height<lower_limit) | (df.height>upper_limit)]
```

Out[48]:

	name	height
8	siddharth	14.2
11	joseph	1.2

In [49]:

```
df_no_outlier = df[(df.height>lower_limit) & (df.height<upper_limit)]
df_no_outlier
```

Out[49]:

	name	height
0	mohan	5.9
1	maria	5.2
2	sakib	5.1
3	tao	5.5
4	virat	4.9
5	vaishnav	6.2
6	yash	5.6
7	vanshika	7.1
9	taabish	6.5
10	john	6.1
12	celina	5.5

In [53]:

```
df.shape
```

Out[53]:

```
(13, 2)
```

In [52]:

```
df_no_outlier.shape
```

Out[52]:

```
(11, 2)
```

In [7]:

```
import pandas as pd
import numpy as np

df=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")
df.head()
print(df)
```

```
    Gender  Math_score  Reading_score  Writing_score  Placement_score  \
0   Female        15.0           90.0           77.0             88.0
1   Female         5.0           82.0           78.0             93.0
2   Female       170.0          100.0           74.0             32.0
3     Male       200.0           89.0           32.0              NaN
4      NaN         1.0           79.0           67.0            100.0
5   Female        56.0           79.0           33.0             80.0
6     Male        16.0           94.0            NaN             80.0
7   Female         NaN           90.0           68.0             83.0
8     Male        36.0           50.0           72.0             90.0
9     Male        69.0           90.0           79.0             78.0
10  Female         0.0           83.0           74.0             89.0
11  Female         6.0           87.0           79.0             98.0
12    Male         4.0           87.0           75.0             93.0
13    Male       150.0           43.0           65.0             21.0
14    Male        76.0           75.0           62.0             78.0
15  Female         3.0           46.0           74.0             93.0
16    Male       103.0            NaN           62.0             94.0
17    Male        26.0           78.0          100.0             87.0
18    Male        78.0           95.0           75.0             11.0
19  Female        82.0           82.0            NaN             76.0
20  Female        17.0           82.0           69.0             85.0

    Placement_offer_count  Region
0                       8  Nashik
1                       6    Pune
2                       2  Nashik
3                       1  Nashik
4                       3  Nashik
5                       2  Nashik
6                       2    Pune
7                       3    Pune
8                       2  Nashik
9                       1  Nashik
10                      2    Pune
11                      3  Nashik
12                      2    Pune
13                      1    Pune
14                      1     NaN
15                      2  Nashik
16                      1  Nashik
17                      2  Nashik
18                      2  Nashik
19                      0    Pune
20                      2  Nashik
```

In [8]:

```
Q1 = df.Math_score.quantile(0.25)
Q3 = df.Math_score.quantile(0.75)
Q1, Q3
```

Out[8]:

```
(5.75, 79.0)
```

In [9]:

```
IQR = Q3 - Q1
IQR
```

Out[9]:

```
73.25
```

In [10]:

```
lower_limit = Q1 - 1.5*IQR
upper_limit = Q3 + 1.5*IQR
lower_limit, upper_limit
```

Out[10]:

```
(-104.125, 188.875)
```

In [12]:

```
df[(df.Math_score<lower_limit) | (df.Math_score>upper_limit)]
```

Out[12]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_count
3	Male	200.0	89.0	32.0	NaN	1

In [13]:

```
df_no_outlier = df[(df.Math_score>lower_limit) & (df.Math_score<upper_limit)]
df_no_outlier
```

Out[13]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.0	90.0	77.0	88.0	
1	Female	5.0	82.0	78.0	93.0	
2	Female	170.0	100.0	74.0	32.0	
4	NaN	1.0	79.0	67.0	100.0	
5	Female	56.0	79.0	33.0	80.0	
6	Male	16.0	94.0	NaN	80.0	
8	Male	36.0	50.0	72.0	90.0	
9	Male	69.0	90.0	79.0	78.0	
10	Female	0.0	83.0	74.0	89.0	
11	Female	6.0	87.0	79.0	98.0	
12	Male	4.0	87.0	75.0	93.0	
13	Male	150.0	43.0	65.0	21.0	
14	Male	76.0	75.0	62.0	78.0	
15	Female	3.0	46.0	74.0	93.0	
16	Male	103.0	NaN	62.0	94.0	
17	Male	26.0	78.0	100.0	87.0	
18	Male	78.0	95.0	75.0	11.0	
19	Female	82.0	82.0	NaN	76.0	
20	Female	17.0	82.0	69.0	85.0	

In [14]:

```
df.shape
```

Out[14]:

```
(21, 7)
```

In [15]:

```
df_no_outlier.shape
```

Out[15]:

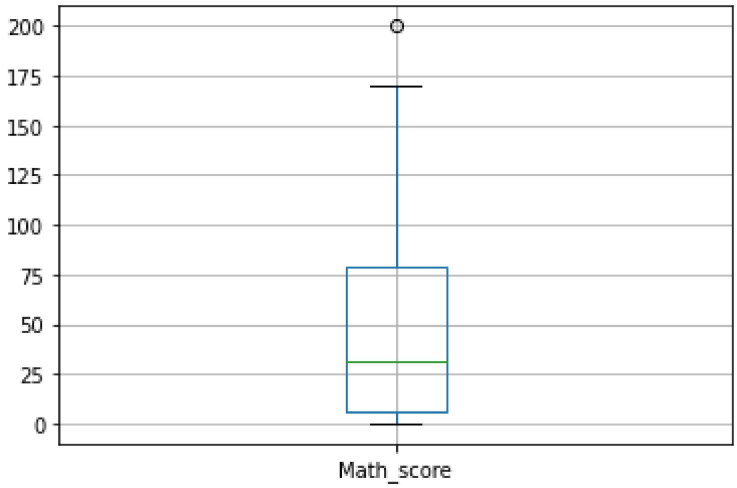
```
(19, 7)
```

In [20]:

```
df.boxplot("Math_score")
```

Out[20]:

<AxesSubplot:>



In [35]:

```
import pandas as pd
import numpy as np

df=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")
df.head()

Q1 = np.nanpercentile(df["Math_score"],25)
Q3 = np.nanpercentile(df["Math_score"],75)
print(Q1, Q3)
```

	Gender	Math_score	Reading_score	Writing_score	Placement_score	\
0	Female	15.0	90.0	77.0	88.0	
1	Female	5.0	82.0	78.0	93.0	
2	Female	170.0	100.0	74.0	32.0	
3	Male	200.0	89.0	32.0	NaN	
4	NaN	1.0	79.0	67.0	100.0	
5	Female	56.0	79.0	33.0	80.0	
6	Male	16.0	94.0	NaN	80.0	
7	Female	NaN	90.0	68.0	83.0	
8	Male	36.0	50.0	72.0	90.0	
9	Male	69.0	90.0	79.0	78.0	
10	Female	0.0	83.0	74.0	89.0	
11	Female	6.0	87.0	79.0	98.0	
12	Male	4.0	87.0	75.0	93.0	
13	Male	150.0	43.0	65.0	21.0	
14	Male	76.0	75.0	62.0	78.0	
15	Female	3.0	46.0	74.0	93.0	
16	Male	103.0	NaN	62.0	94.0	
17	Male	26.0	78.0	100.0	87.0	
18	Male	78.0	95.0	75.0	11.0	
19	Female	82.0	82.0	NaN	76.0	
20	Female	17.0	82.0	69.0	85.0	

	Placement_offer_count	Region
0	8	Nashik
1	6	Pune
2	2	Nashik
3	1	Nashik
4	3	Nashik
5	2	Nashik
6	2	Pune
7	3	Pune
8	2	Nashik
9	1	Nashik
10	2	Pune
11	3	Nashik
12	2	Pune
13	1	Pune
14	1	NaN
15	2	Nashik
16	1	Nashik
17	2	Nashik
18	2	Nashik
19	0	Pune
20	2	Nashik

```
5.75 79.0
```

In [36]:

```
IQR = Q3 - Q1
lower_limit = Q1 - 1.5*IQR
upper_limit = Q3 + 1.5*IQR
lower_limit, upper_limit
```

Out[36]:

```
(-104.125, 188.875)
```

In [40]:

```
col = ["Math_score"]

index_outliers = np.where((df[col] < lower_limit) | (df[col] > upper_limit))
index_outliers
```

Out[40]:

```
(array([3], dtype=int64), array([0], dtype=int64))
```

In [46]:

```
sample_outliers = df[col][(df[col] < lower_limit) | (df[col] > upper_limit)]
sample_outliers
```

Out[46]:

	Math_score
0	NaN
1	NaN
2	NaN
3	200.0
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN
10	NaN
11	NaN
12	NaN
13	NaN
14	NaN
15	NaN
16	NaN
17	NaN
18	NaN
19	NaN
20	NaN

In [47]:

```
df1 = df
df[col] = np.where(df[col] < lower_limit, lower_limit, df[col])
df[col] = np.where(df[col] > upper_limit, upper_limit, df[col])
df
```

Out[47]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.000	90.0	77.0	88.0	8
1	Female	5.000	82.0	78.0	93.0	6
2	Female	170.000	100.0	74.0	32.0	2
3	Male	188.875	89.0	32.0	NaN	1
4	NaN	1.000	79.0	67.0	100.0	3
5	Female	56.000	79.0	33.0	80.0	2
6	Male	16.000	94.0	NaN	80.0	2
7	Female	NaN	90.0	68.0	83.0	3
8	Male	36.000	50.0	72.0	90.0	2
9	Male	69.000	90.0	79.0	78.0	1
10	Female	0.000	83.0	74.0	89.0	2
11	Female	6.000	87.0	79.0	98.0	3
12	Male	4.000	87.0	75.0	93.0	2
13	Male	150.000	43.0	65.0	21.0	1
14	Male	76.000	75.0	62.0	78.0	1
15	Female	3.000	46.0	74.0	93.0	2
16	Male	103.000	NaN	62.0	94.0	1
17	Male	26.000	78.0	100.0	87.0	2
18	Male	78.000	95.0	75.0	11.0	2
19	Female	82.000	82.0	NaN	76.0	0
20	Female	17.000	82.0	69.0	85.0	2

In [49]:

```
ninetieth_percentile = np.nanpercentile(df1["Math_score"], 90)
ninetieth_percentile
```

Out[49]:

```
152.00000000000003
```

In [50]:

```
df1[col] = np.where(df1[col] > upper_limit, ninetieth_percentile, df1[col])
df1
```

Out[50]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.000	90.0	77.0	88.0	8
1	Female	5.000	82.0	78.0	93.0	6
2	Female	170.000	100.0	74.0	32.0	2
3	Male	188.875	89.0	32.0	NaN	1
4	NaN	1.000	79.0	67.0	100.0	3
5	Female	56.000	79.0	33.0	80.0	2
6	Male	16.000	94.0	NaN	80.0	2
7	Female	NaN	90.0	68.0	83.0	3
8	Male	36.000	50.0	72.0	90.0	2
9	Male	69.000	90.0	79.0	78.0	1
10	Female	0.000	83.0	74.0	89.0	2
11	Female	6.000	87.0	79.0	98.0	3
12	Male	4.000	87.0	75.0	93.0	2
13	Male	150.000	43.0	65.0	21.0	1
14	Male	76.000	75.0	62.0	78.0	1
15	Female	3.000	46.0	74.0	93.0	2
16	Male	103.000	NaN	62.0	94.0	1
17	Male	26.000	78.0	100.0	87.0	2
18	Male	78.000	95.0	75.0	11.0	2
19	Female	82.000	82.0	NaN	76.0	0
20	Female	17.000	82.0	69.0	85.0	2

In [56]:

```
median = np.nanmedian(df[col])
median
```

Out[56]:

```
31.0
```

In [57]:

```
for i in index_outliers:
    df.at[i, col] = median

df
```

Out[57]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	31.0	90.0	77.0	88.0	8
1	Female	5.0	82.0	78.0	93.0	6
2	Female	170.0	100.0	74.0	32.0	2
3	Male	31.0	89.0	32.0	NaN	1
4	NaN	1.0	79.0	67.0	100.0	3
5	Female	56.0	79.0	33.0	80.0	2
6	Male	16.0	94.0	NaN	80.0	2
7	Female	NaN	90.0	68.0	83.0	3
8	Male	36.0	50.0	72.0	90.0	2
9	Male	69.0	90.0	79.0	78.0	1
10	Female	0.0	83.0	74.0	89.0	2
11	Female	6.0	87.0	79.0	98.0	3
12	Male	4.0	87.0	75.0	93.0	2
13	Male	150.0	43.0	65.0	21.0	1
14	Male	76.0	75.0	62.0	78.0	1
15	Female	3.0	46.0	74.0	93.0	2
16	Male	103.0	NaN	62.0	94.0	1
17	Male	26.0	78.0	100.0	87.0	2
18	Male	78.0	95.0	75.0	11.0	2
19	Female	82.0	82.0	NaN	76.0	0
20	Female	17.0	82.0	69.0	85.0	2

In [59]:

```
import numpy as np
from scipy import stats

df3=pd.read_csv("C:\\Users\\Nancy\\Desktop\\Nancy_Assign2.csv")
df3.head()
df3
```

Out[59]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.0	90.0	77.0	88.0	8
1	Female	5.0	82.0	78.0	93.0	6
2	Female	170.0	100.0	74.0	32.0	2
3	Male	200.0	89.0	32.0	NaN	1
4	NaN	1.0	79.0	67.0	100.0	3
5	Female	56.0	79.0	33.0	80.0	2
6	Male	16.0	94.0	NaN	80.0	2
7	Female	NaN	90.0	68.0	83.0	3
8	Male	36.0	50.0	72.0	90.0	2
9	Male	69.0	90.0	79.0	78.0	1
10	Female	0.0	83.0	74.0	89.0	2
11	Female	6.0	87.0	79.0	98.0	3
12	Male	4.0	87.0	75.0	93.0	2
13	Male	150.0	43.0	65.0	21.0	1
14	Male	76.0	75.0	62.0	78.0	1
15	Female	3.0	46.0	74.0	93.0	2
16	Male	103.0	NaN	62.0	94.0	1
17	Male	26.0	78.0	100.0	87.0	2
18	Male	78.0	95.0	75.0	11.0	2
19	Female	82.0	82.0	NaN	76.0	0
20	Female	17.0	82.0	69.0	85.0	2

In [70]:

```
z = np.abs(stats.zscore(df3["Math_score"], nan_policy='omit'))
z
```

Out[70]:

```
0     0.690171
1     0.859955
2     1.941477
3     2.450828
4     0.927868
5     0.005942
6     0.673193
7          NaN
8     0.333625
9     0.226661
10    0.944847
11    0.842976
12    0.876933
13    1.601910
14    0.345510
15    0.893911
16    0.803926
17    0.503409
18    0.379467
19    0.447380
20    0.656214
Name: Math_score, dtype: float64
```

In [71]:

```
threshold = 0.18
sample_outliers = np.where(z < threshold)
sample_outliers
```

Out[71]:

```
(array([5], dtype=int64),)
```

In [74]:

```
upper_threshold = 1.4
lower_threshold = 0.18

sample_outliers = np.where((z < lower_threshold) | (z > upper_threshold))
sample_outliers
```

Out[74]:

```
(array([ 2,  3,  5, 13], dtype=int64),)
```

In [75]:

```
new_df1 = df3
for i in sample_outliers:
    new_df1.drop(i, inplace=True)

new_df1
```

Out[75]:

	Gender	Math_score	Reading_score	Writing_score	Placement_score	Placement_offer_coun
0	Female	15.0	90.0	77.0	88.0	8
1	Female	5.0	82.0	78.0	93.0	6
4	NaN	1.0	79.0	67.0	100.0	3
6	Male	16.0	94.0	NaN	80.0	2
7	Female	NaN	90.0	68.0	83.0	3
8	Male	36.0	50.0	72.0	90.0	2
9	Male	69.0	90.0	79.0	78.0	1
10	Female	0.0	83.0	74.0	89.0	2
11	Female	6.0	87.0	79.0	98.0	3
12	Male	4.0	87.0	75.0	93.0	2
14	Male	76.0	75.0	62.0	78.0	1
15	Female	3.0	46.0	74.0	93.0	2
16	Male	103.0	NaN	62.0	94.0	1
17	Male	26.0	78.0	100.0	87.0	2
18	Male	78.0	95.0	75.0	11.0	2
19	Female	82.0	82.0	NaN	76.0	0
20	Female	17.0	82.0	69.0	85.0	2

In [ ]:

```

```