

vxhmjbj9x

February 17, 2025

```
[1]: import pandas as pd
import numpy as np
```

```
[2]: df=pd.read_csv("student.csv")
```

```
[3]: df
```

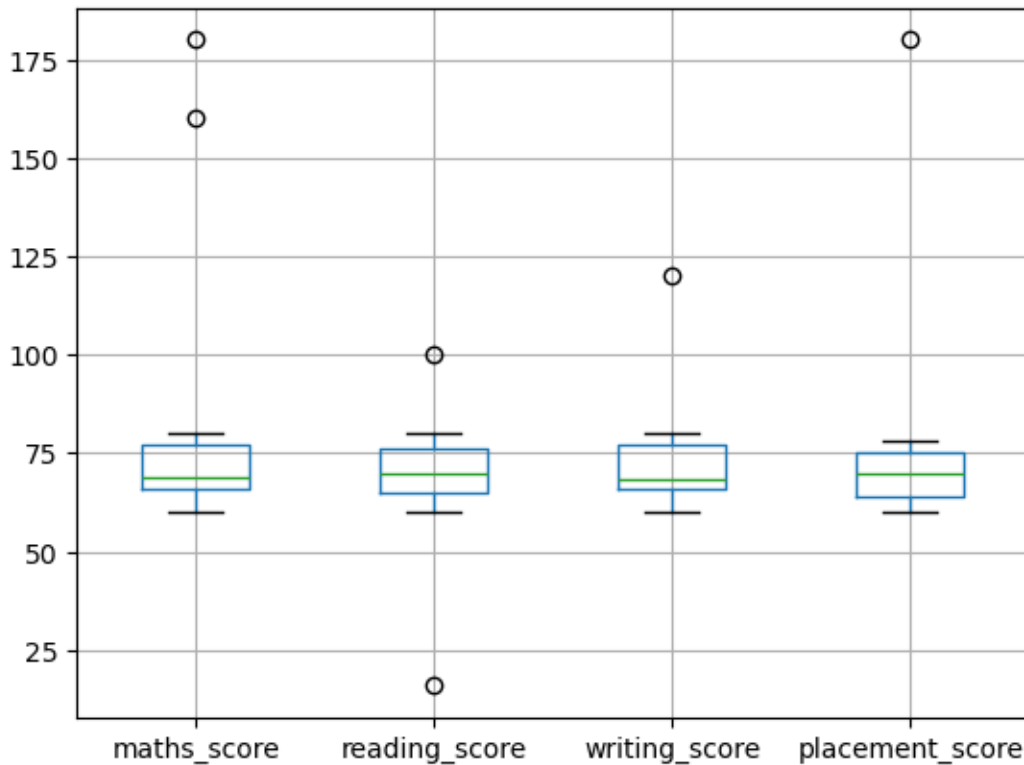
```
[3]:
```

| | maths_score | reading_score | writing_score | placement_score | \ |
|----|-------------|---------------|---------------|-----------------|---|
| 0 | 78 | 63 | 67 | 68 | |
| 1 | 67 | 16 | 73 | 63 | |
| 2 | 62 | 78 | 60 | 63 | |
| 3 | 73 | 69 | 69 | 66 | |
| 4 | 67 | 100 | 77 | 71 | |
| 5 | 68 | 70 | 64 | 69 | |
| 6 | 79 | 64 | 72 | 73 | |
| 7 | 78 | 70 | 78 | 64 | |
| 8 | 160 | 67 | 75 | 62 | |
| 9 | 72 | 80 | 79 | 69 | |
| 10 | 180 | 70 | 68 | 72 | |
| 11 | 65 | 79 | 69 | 75 | |
| 12 | 62 | 74 | 80 | 75 | |
| 13 | 66 | 63 | 120 | 74 | |
| 14 | 76 | 60 | 60 | 64 | |
| 15 | 70 | 61 | 61 | 63 | |
| 16 | 72 | 69 | 65 | 74 | |
| 17 | 80 | 76 | 74 | 78 | |
| 18 | 65 | 70 | 63 | 76 | |
| 19 | 71 | 65 | 68 | 180 | |
| 20 | 68 | 65 | 68 | 75 | |
| 21 | 77 | 79 | 66 | 62 | |
| 22 | 68 | 76 | 80 | 74 | |
| 23 | 80 | 71 | 79 | 78 | |
| 24 | 62 | 71 | 62 | 60 | |
| 25 | 62 | 78 | 66 | 78 | |
| 26 | 66 | 77 | 68 | 69 | |
| 27 | 60 | 75 | 77 | 65 | |

| | club_join_year | placement_offer_count | Gender |
|----|----------------|-----------------------|--------|
| 0 | 2021 | 1 | Female |
| 1 | 2019 | 1 | Male |
| 2 | 2021 | 1 | Male |
| 3 | 2021 | 1 | Male |
| 4 | 2020 | 1 | Male |
| 5 | 2018 | 1 | Male |
| 6 | 2020 | 1 | Male |
| 7 | 2021 | 1 | Male |
| 8 | 2020 | 1 | Male |
| 9 | 2020 | 1 | Male |
| 10 | 2019 | 1 | Male |
| 11 | 2019 | 2 | Female |
| 12 | 2018 | 2 | Female |
| 13 | 2020 | 1 | Female |
| 14 | 2018 | 1 | Female |
| 15 | 2021 | 1 | Female |
| 16 | 2020 | 1 | Female |
| 17 | 2020 | 2 | Female |
| 18 | 2019 | 2 | Female |
| 19 | 2019 | 3 | Female |
| 20 | 2019 | 2 | Female |
| 21 | 2020 | 1 | Female |
| 22 | 2020 | 1 | Male |
| 23 | 2020 | 2 | Male |
| 24 | 2018 | 1 | Male |
| 25 | 2018 | 2 | Male |
| 26 | 2020 | 1 | Female |
| 27 | 2021 | 1 | Male |

```
[4]: col = ['maths_score', 'reading_score', 'writing_score', 'placement_score']
df.boxplot(col)
```

```
[4]: <Axes: >
```

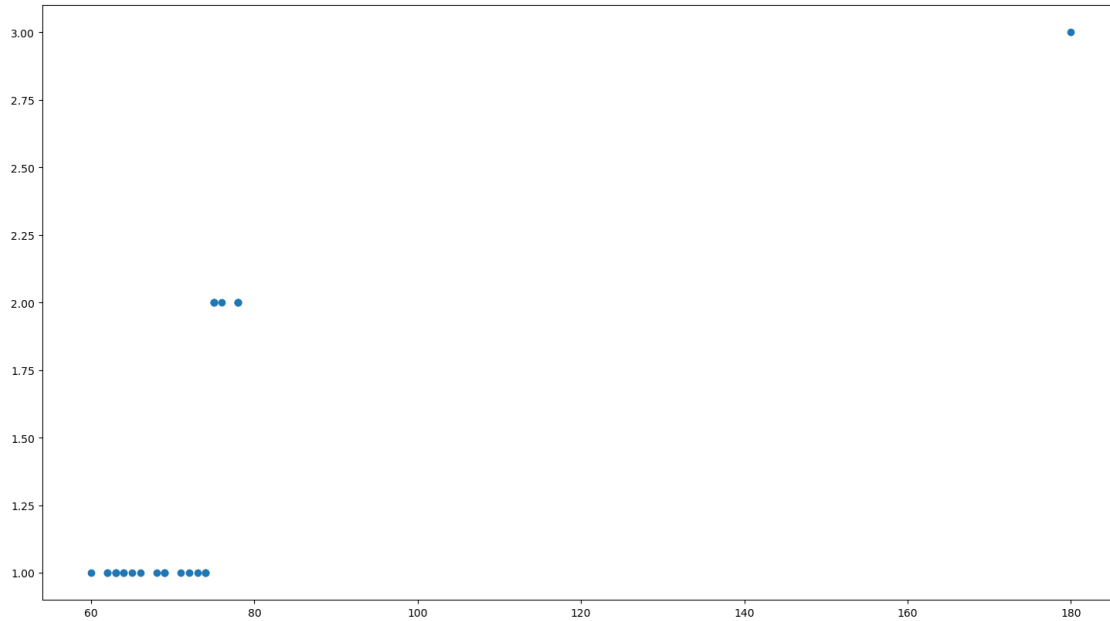


```
[6]: print(np.where(df['maths_score']>90))
      print(np.where(df['reading_score']<25))
      print(np.where(df['writing_score']<30))
```

```
(array([ 8, 10], dtype=int64),)
(array([1], dtype=int64),)
(array([], dtype=int64),)
```

```
[7]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
```

```
[8]: fig, ax = plt.subplots(figsize = (18,10))
      ax.scatter(df['placement_score'], df['placement_offer_count'])
      plt.show()
```



```
[10]: print(np.where((df['placement_score']<50) & (df['placement_offer_count']>1)))
      print(np.where((df['placement_score']>85) & (df['placement_offer_count']<3)))
```

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

```
[11]: import numpy as np
      from scipy import stats
```

```
[12]: z = np.abs(stats.zscore(df['maths_score']))
```

```
[13]: print(z)
```

```
0    0.040255
1    0.373027
2    0.560882
3    0.147601
4    0.373027
5    0.335456
6    0.077826
7    0.040255
8    3.121082
9    0.185172
10   3.872504
11   0.448169
12   0.560882
13   0.410598
```

```
14    0.034887
15    0.260314
16    0.185172
17    0.115397
18    0.448169
19    0.222743
20    0.335456
21    0.002684
22    0.335456
23    0.115397
24    0.560882
25    0.560882
26    0.410598
27    0.636024
Name: maths_score, dtype: float64
```

```
[14]: threshold = 0.18
```

```
[15]: sample_outliers = np.where(z < threshold)
      sample_outliers
```

```
[15]: (array([ 0,  3,  6,  7, 14, 17, 21, 23], dtype=int64),)
```

```
[16]: import numpy as np
```

```
[17]: sorted_rscore= sorted(df['reading_score'])
```

```
[18]: sorted_rscore
```

```
[18]: [16,
      60,
      61,
      63,
      63,
      64,
      65,
      65,
      67,
      69,
      69,
      70,
      70,
      70,
      70,
      71,
      71,
      74,
```

```
75,  
76,  
76,  
77,  
78,  
78,  
79,  
79,  
80,  
100]
```

```
[19]: q1 = np.percentile(sorted_rscore, 25)
```

```
[20]: q3 = np.percentile(sorted_rscore, 75)  
print(q1,q3)
```

```
65.0 76.25
```

```
[21]: IQR = q3-q1
```

```
[22]: lwr_bound = q1-(1.5*IQR)  
upr_bound = q3+(1.5*IQR)  
print(lwr_bound, upr_bound)
```

```
48.125 93.125
```

```
[24]: r_outliers = []  
for i in sorted_rscore:  
    if (i<lwr_bound or i>upr_bound):  
        r_outliers.append(i)  
print(r_outliers)
```

```
[16, 100]
```

```
[ ]:
```