practical-2

May 9, 2025

```
[1]:
     import pandas as pd
     import numpy as np
     df=pd.read_csv("StudentsPerformanceTest1.csv")
[4]: df
[4]:
        gender math score reading score
                                            writing score
                                                             Placement Score \
        female
                         72
                                                       74.0
                                                                         78.0
                                        72
     1
        female
                         69
                                        90
                                                       88.0
                                                                          NaN
        female
                         90
                                        95
                                                       93.0
                                                                         74.0
     3
          male
                         47
                                        57
                                                       NaN
                                                                         78.0
     4
          male
                                        78
                                                       75.0
                                                                         81.0
                         na
     5
        female
                         71
                                                       78.0
                                                                         70.0
                                        Na
     6
          male
                         12
                                                       52.0
                                                                         12.0
                                        44
     7
                                                       67.0
          male
                        NaN
                                        65
                                                                         49.0
     8
                          5
                                        77
                                                       89.0
                                                                         55.0
          male
        placement offer count
                                  Region
     0
                              1
                                    Pune
                              2
     1
                                      na
                              2
     2
                                 Nashik
     3
                               1
                                      Na
                              3
     4
                                    Pune
     5
                              4
                                      na
                              2
     6
                                  Nashik
     7
                              1
                                    Pune
     8
                              0
                                     NaN
[5]: df.isnull()
[5]:
                                                               Placement Score
        gender
                 math score
                              reading score
                                               writing score
         False
                                                        False
                       False
                                       False
                                                                           False
     1
         False
                       False
                                       False
                                                        False
                                                                            True
         False
                       False
                                       False
                                                        False
     2
                                                                          False
         False
                       False
                                       False
                                                         True
                                                                           False
```

```
False
                      False
                                      False
     5
                                                      False
                                                                        False
         False
     6
                      False
                                      False
                                                      False
                                                                        False
     7
         False
                       True
                                      False
                                                      False
                                                                        False
         False
                      False
                                      False
                                                      False
                                                                        False
        placement offer count Region
     0
                         False
                                 False
     1
                         False
                                 False
     2
                         False
                                 False
                         False
     3
                                 False
     4
                         False
                                 False
     5
                         False
                                 False
     6
                         False
                                 False
     7
                         False
                                 False
     8
                         False
                                   True
[6]: series=pd.isnull(df["math score"])
     df[series]
       gender math score reading score writing score Placement Score \
[6]:
         male
                      NaN
                                      65
                                                    67.0
                                                                      49.0
        placement offer count Region
     7
                                 Pune
[7]: import pandas as pd
     import numpy as np
[8]: df.notnull()
[8]:
               math score reading score writing score Placement Score \
        gender
     0
          True
                       True
                                       True
                                                       True
                                                                         True
     1
          True
                       True
                                       True
                                                       True
                                                                        False
     2
          True
                       True
                                       True
                                                       True
                                                                         True
     3
          True
                       True
                                       True
                                                      False
                                                                         True
     4
          True
                       True
                                       True
                                                       True
                                                                         True
     5
          True
                       True
                                       True
                                                       True
                                                                         True
     6
          True
                       True
                                       True
                                                       True
                                                                         True
                                                                         True
     7
          True
                      False
                                       True
                                                       True
          True
                       True
                                       True
                                                       True
                                                                         True
        placement offer count
                                Region
     0
                          True
                                   True
                          True
                                   True
     1
     2
                          True
                                   True
     3
                          True
                                   True
```

4

False

False

False

False

False

```
4
                            True
                                     True
      5
                            True
                                     True
      6
                            True
                                     True
      7
                            True
                                     True
      8
                            True
                                   False
 [9]: series=pd.notnull(df["math score"])
      df[series]
 [9]:
         gender math score reading score
                                             writing score Placement Score \
                                                       74.0
                                                                          78.0
      0 female
                          72
                                         72
         female
                          69
                                         90
                                                       88.0
                                                                           NaN
      1
        female
                                                       93.0
                                                                          74.0
                          90
                                         95
           male
                          47
                                                                          78.0
      3
                                         57
                                                        NaN
      4
           male
                                         78
                                                       75.0
                                                                          81.0
                          na
                                                       78.0
                                                                          70.0
      5
        female
                          71
                                         Na
      6
           male
                          12
                                         44
                                                       52.0
                                                                          12.0
           male
                           5
                                                       89.0
                                                                          55.0
      8
                                         77
         placement offer count
                                  Region
      0
                                     Pune
                               2
      1
                                       na
      2
                               2
                                  Nashik
      3
                               1
                                       Na
      4
                               3
                                     Pune
      5
                               4
                                       na
                               2
      6
                                  Nashik
      8
                               0
                                      NaN
[14]: from sklearn.preprocessing import LabelEncoder
[15]: le=LabelEncoder()
[16]: df['gender']=le.fit_transform(df['gender'])
[17]: newdf=df
      df
         gender math score reading score writing score Placement Score \
[17]:
      0
               0
                          72
                                         72
                                                       74.0
                                                                          78.0
               0
      1
                          69
                                         90
                                                       88.0
                                                                           {\tt NaN}
      2
               0
                          90
                                         95
                                                       93.0
                                                                          74.0
                          47
                                                                          78.0
      3
               1
                                         57
                                                        NaN
      4
               1
                          na
                                         78
                                                       75.0
                                                                          81.0
      5
               0
                                                       78.0
                                                                          70.0
                          71
                                         {\tt Na}
      6
               1
                          12
                                         44
                                                       52.0
                                                                          12.0
      7
               1
                                                       67.0
                         NaN
                                         65
                                                                          49.0
```

```
8
               1
                           5
                                        77
                                                       89.0
                                                                         55.0
         placement offer count
                                  Region
      0
                                    Pune
                               1
      1
                               2
                                      na
      2
                               2
                                  Nashik
      3
                               1
                                      Na
      4
                               3
                                    Pune
      5
                               4
                                      na
      6
                               2
                                  Nashik
      7
                               1
                                    Pune
                                     NaN
[18]: missing_values=["Na", "na"]
[20]: df = pd.read_csv("StudentsPerformanceTest1.csv", na_values=["missing", "N/A", __

¬"NaN"])
[21]: df
[21]:
         gender math score reading score writing score Placement Score \
      0 female
                         72
                                         72
                                                       74.0
                                                                         78.0
      1 female
                         69
                                         90
                                                       88.0
                                                                          {\tt NaN}
         female
                         90
                                         95
                                                       93.0
                                                                         74.0
      3
           male
                         47
                                         57
                                                        NaN
                                                                         78.0
           male
      4
                         na
                                         78
                                                       75.0
                                                                         81.0
                                                       78.0
      5
        female
                         71
                                         Na
                                                                         70.0
      6
           male
                         12
                                         44
                                                       52.0
                                                                         12.0
                        NaN
                                                       67.0
                                                                         49.0
      7
           male
                                         65
                           5
           male
                                         77
                                                       89.0
                                                                         55.0
         placement offer count
                                  Region
                                    Pune
      0
                               2
      1
                                      na
                               2
      2
                                  Nashik
      3
                               1
                                      Na
                               3
      4
                                    Pune
      5
                               4
                                      na
      6
                               2
                                  Nashik
      7
                               1
                                    Pune
                                     NaN
[22]: ndf=df
[23]: ndf.fillna(0)
```

```
[23]:
         gender math score reading score writing score Placement Score \
         female
                                                     74.0
                                                                        78.0
                         72
                                                     88.0
                                                                         0.0
      1
         female
                         69
                                        90
      2
         female
                         90
                                        95
                                                     93.0
                                                                        74.0
      3
           male
                         47
                                        57
                                                      0.0
                                                                        78.0
      4
           male
                                        78
                                                     75.0
                                                                        81.0
                         na
         female
                                                     78.0
                                                                        70.0
      5
                         71
                                        Na
           male
                                                     52.0
                                                                        12.0
      6
                         12
                                        44
      7
           male
                          0
                                        65
                                                     67.0
                                                                        49.0
                          5
                                                                        55.0
           male
                                        77
                                                     89.0
         placement offer count
                                 Region
      0
                                   Pune
                              2
      1
                                      na
      2
                                 Nashik
      3
                              1
                                      Na
      4
                              3
                                   Pune
      5
                              4
                                      na
      6
                              2
                                 Nashik
      7
                              1
                                   Pune
                                       0
      8
                              0
[26]: df['math score'] = pd.to_numeric(df['math score'], errors='coerce')
[27]: m_v = df['math score'].mean()
      print(m_v)
     52.285714285714285
[29]: df['math score'] = df['math score'].fillna(value=m_v)
[30]: df.fillna({'math score': m_v}, inplace=True)
[31]: df
[31]:
         gender
                 math score reading score writing score Placement Score \
         female
                   72.000000
                                         72
                                                       74.0
                                                                         78.0
      1 female
                   69.000000
                                         90
                                                       88.0
                                                                         NaN
      2
         female
                   90.000000
                                         95
                                                       93.0
                                                                         74.0
      3
           male
                  47.000000
                                         57
                                                                         78.0
                                                       NaN
      4
           male
                   52.285714
                                         78
                                                       75.0
                                                                         81.0
        female
                  71.000000
                                                       78.0
                                                                         70.0
      5
                                         Na
                  12.000000
           male
                                                       52.0
                                                                         12.0
      6
                                         44
      7
           male
                  52.285714
                                         65
                                                       67.0
                                                                         49.0
           male
                   5.000000
                                         77
                                                       89.0
                                                                         55.0
```

placement offer count Region

```
2
      2
                                   Nashik
      3
                               1
                                       Na
      4
                               3
                                     Pune
      5
                               4
                                       na
                               2
      6
                                  Nashik
      7
                                1
                                     Pune
      8
                               0
                                      NaN
     ndf.replace(to_replace=np.nan,value=-99)
[32]:
                  math score reading score
                                               writing score
         gender
                                                               Placement Score
         female
                   72.000000
                                          72
                                                         74.0
                                                                           78.0
      1
         female
                   69.000000
                                          90
                                                         88.0
                                                                          -99.0
      2
         female
                   90.000000
                                          95
                                                         93.0
                                                                           74.0
      3
           male
                                                        -99.0
                   47.000000
                                          57
                                                                           78.0
      4
           male
                   52.285714
                                          78
                                                         75.0
                                                                           81.0
         female
      5
                   71.000000
                                          Na
                                                        78.0
                                                                           70.0
      6
           male
                   12.000000
                                                         52.0
                                                                           12.0
                                          44
      7
           male
                   52.285714
                                          65
                                                         67.0
                                                                           49.0
      8
           male
                    5.000000
                                          77
                                                         89.0
                                                                           55.0
         placement offer count
                                   Region
      0
                                1
                                     Pune
                               2
      1
                                       na
                               2
      2
                                   Nashik
      3
                               1
                                       Na
                               3
      4
                                     Pune
      5
                               4
                                       na
                               2
      6
                                  Nashik
      7
                                1
                                     Pune
      8
                               0
                                      -99
[33]: ndf.dropna()
[33]:
         gender
                  math score reading score
                                              writing score
                                                               Placement Score \
         female
                   72.000000
                                          72
                                                         74.0
                                                                           78.0
      2
         female
                   90.000000
                                          95
                                                         93.0
                                                                           74.0
      4
           male
                   52.285714
                                          78
                                                        75.0
                                                                           81.0
                   71.000000
                                                         78.0
                                                                           70.0
      5
         female
                                          Na
      6
            male
                                                         52.0
                   12.000000
                                          44
                                                                           12.0
      7
           male
                   52.285714
                                          65
                                                         67.0
                                                                           49.0
         placement offer count
                                   Region
      0
                                     Pune
      2
                                  Nashik
```

0

1

1

2

Pune

na

```
4
                               3
                                     Pune
      5
                               4
                                       na
                               2
      6
                                   Nashik
      7
                                     Pune
[34]: ndf.dropna(how='all')
[34]:
         gender
                  math score reading score
                                              writing score
                                                               Placement Score
         female
                   72.000000
                                                         74.0
                                                                           78.0
                                          72
      1
         female
                   69.000000
                                          90
                                                         88.0
                                                                            NaN
      2
         female
                                          95
                                                         93.0
                                                                           74.0
                   90.000000
      3
           male
                   47.000000
                                          57
                                                          NaN
                                                                           78.0
      4
           male
                                          78
                                                         75.0
                   52.285714
                                                                           81.0
      5
         female
                   71.000000
                                          Na
                                                         78.0
                                                                           70.0
      6
           male
                   12.000000
                                          44
                                                         52.0
                                                                           12.0
      7
           male
                   52.285714
                                                         67.0
                                                                           49.0
                                          65
      8
           male
                    5.000000
                                          77
                                                         89.0
                                                                           55.0
         placement offer count
                                  Region
      0
                               1
                                     Pune
                               2
      1
                                       na
      2
                               2
                                   Nashik
      3
                               1
                                       Na
      4
                               3
                                     Pune
      5
                               4
                                       na
      6
                               2
                                  Nashik
      7
                               1
                                     Pune
      8
                               0
                                      NaN
[35]: ndf.dropna(axis=1)
[35]:
         gender
                  math score reading score
                                              placement offer count
      0
         female
                   72.000000
                                          72
                                                                     1
      1
         female
                   69.000000
                                          90
                                                                     2
      2
         female
                                          95
                                                                     2
                   90.000000
      3
           male
                   47.000000
                                          57
                                                                     1
           male
                                          78
                                                                     3
      4
                   52.285714
      5
         female
                   71.000000
                                          Na
                                                                     4
      6
           male
                                          44
                                                                     2
                   12.000000
      7
           male
                   52.285714
                                                                     1
                                          65
      8
                    5.000000
                                                                     0
           male
                                          77
[36]: new_data=ndf.dropna(axis=0, how='any')
[37]: new_data
```

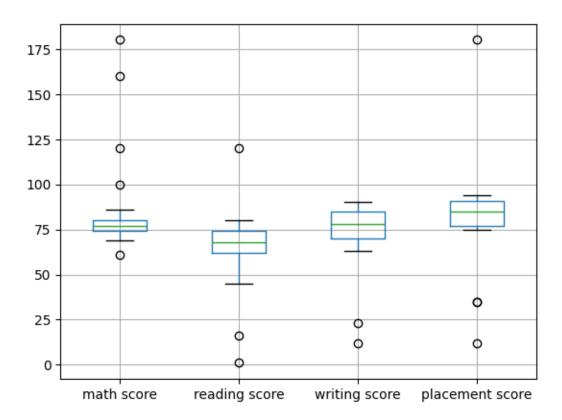
```
[37]:
         gender math score reading score writing score Placement Score \
         female
                   72.000000
                                                        74.0
                                                                          78.0
      2
         female
                   90.000000
                                          95
                                                        93.0
                                                                          74.0
      4
           male
                   52.285714
                                          78
                                                        75.0
                                                                          81.0
      5
         female
                   71.000000
                                          Na
                                                        78.0
                                                                          70.0
      6
           male
                   12.000000
                                          44
                                                        52.0
                                                                          12.0
      7
           male
                   52.285714
                                          65
                                                        67.0
                                                                          49.0
         placement offer count
                                  Region
      0
                                    Pune
      2
                               2
                                  Nashik
      4
                               3
                                    Pune
      5
                               4
                                      na
      6
                               2
                                  Nashik
      7
                               1
                                    Pune
 []:
 []:
     import pandas as pd
[39]: import numpy as np
[40]: df=pd.read_csv("demo1.csv")
[41]: df
[41]:
          math score reading score writing score placement score \
      0
                   80
                                                                      89
                                   68
                                                    70
      1
                                                    85
                                                                      91
                   71
                                   61
      2
                   79
                                   16
                                                   87
                                                                      77
      3
                   61
                                   77
                                                   74
                                                                      76
      4
                   78
                                   71
                                                    67
                                                                      90
      5
                   73
                                   68
                                                   90
                                                                      80
      6
                   77
                                   62
                                                    70
                                                                      35
      7
                   74
                                   45
                                                    80
                                                                      12
                   76
                                   60
                                                    79
                                                                      77
      8
                   75
                                                                      87
      9
                                   65
                                                    85
      10
                  160
                                   67
                                                    12
                                                                      83
                   79
                                   72
                                                    88
      11
                                                                     180
      12
                                   80
                                                    78
                                                                      94
                   80
      13
                   78
                                   69
                                                    71
                                                                      90
      14
                   75
                                    1
                                                    71
                                                                      81
      15
                   78
                                   62
                                                    79
                                                                      93
      16
                   86
                                   78
                                                   80
                                                                      88
      17
                   80
                                   74
                                                   23
                                                                      76
```

```
18
                  75
                                   62
                                                    86
                                                                        87
     19
                  82
                                   70
                                                    87
                                                                        94
     20
                   69
                                                                        35
                                    65
                                                    84
     21
                  100
                                   77
                                                     70
                                                                        91
     22
                  72
                                   60
                                                     78
                                                                        94
     23
                  74
                                   65
                                                    71
                                                                        84
     24
                  75
                                   77
                                                    83
                                                                        77
     25
                  180
                                   67
                                                     63
                                                                        75
                  72
     26
                                   120
                                                    70
                                                                        84
     27
                  71
                                   79
                                                    88
                                                                        85
     28
                  120
                                   73
                                                    71
                                                                        94
          placement offer count
                                   club join year
     0
                                3
                                               2019
                                3
     1
                                               2019
     2
                                2
                                               2018
     3
                                2
                                               2020
                                3
                                               2019
     4
     5
                                2
                                               2019
     6
                                2
                                               2020
     7
                                1
                                               2019
     8
                                2
                                               2020
     9
                                3
                                               2018
     10
                                2
                                               2020
     11
                                2
                                               2019
     12
                                3
                                               2021
     13
                                3
                                               2019
                                2
     14
                                               2019
     15
                                3
                                               2021
     16
                                3
                                               2019
     17
                                2
                                               2021
     18
                                3
                                               2019
     19
                                3
                                               2019
     20
                                1
                                               2018
                                3
                                               2018
     21
     22
                                3
                                               2019
     23
                                2
                                               2019
     24
                                2
                                               2020
     25
                                3
                                               2021
     26
                                2
                                               2021
     27
                                3
                                               2021
     28
                                3
                                               2019
[]: import matplotlib.pyplot as plt
```

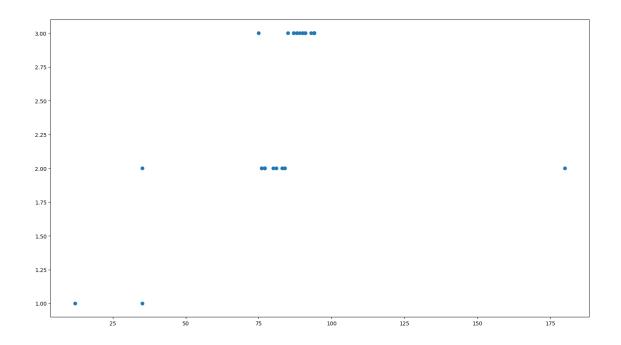
[46]: col=['math score','reading score','writing score','placement score']

```
[47]: df.boxplot(col)
```

[47]: <Axes: >



```
[48]: print(np.where(df['math score']>90))
          (array([10, 21, 25, 28]),)
[49]: print(np.where(df['reading score']<25))
          (array([ 2, 14]),)
[50]: print(np.where(df['writing score']<30))
          (array([10, 17]),)
[98]: fig,ax=plt.subplots(figsize=(18,10))
          ax.scatter(df['placement score'],df['placement offer count'])
          plt.show()</pre>
```



```
[99]: ax.set_ylabel('(Proportion non-retail busines acres)/(town)')
 [99]: Text(4.44444444444452, 0.5, '(Proportion non-retail busines acres)/(town)')
[100]:
      ax.set_ylabel('(Full-value property-tax rate)/($10,000)')
[100]: Text(4.44444444444452, 0.5, '(Full-value property-tax rate)/($10,000)')
[101]: print(np.where((df['placement score']<50) & (df['placement offer count']>1)))
      (array([6]),)
[102]: print(np.where((df['placement score']>85) & (df['placement offer count']<3)))
      (array([11]),)
[103]: from scipy import stats
[104]: z=np.abs(stats.zscore(df['math score']))
[105]: print(z)
      [0.17564553 0.5282877 0.21482799 0.92011234 0.25401045 0.44992277
       0.29319292 0.41074031 0.33237538 0.37155785 2.95895157 0.21482799
       0.17564553 0.25401045 0.37155785 0.25401045 0.05944926 0.17564553
       0.37155785 0.0972806 0.60665263 0.60800375 0.48910524 0.41074031
       0.37155785 3.74260085 0.48910524 0.5282877 1.39165302
```

```
[106]: threshold=0.18
[107]: sample_outliers=np.where(z<threshold)
[108]: sample_outliers
[108]: (array([ 0, 12, 16, 17, 19]),)
[109]: sorted_rscore=sorted(df['reading score'])
[110]: sorted_rscore
[110]: [1,
        16,
        45,
        60,
        60,
        61,
        62,
        62,
        62,
        65,
        65,
        65,
        67,
        67,
        68,
        68,
        69,
        70,
        71,
        72,
        73,
        74,
        77,
        77,
        77,
        78,
        79,
        80,
        120]
[111]: q1 = np.percentile(sorted_rscore, 25)
       q3 = np.percentile(sorted_rscore, 75)
       print(q1,q3)
      62.0 74.0
```

```
[112]: | IQR = q3-q1|
[113]: lwr_bound = q1-(1.5*IQR)
       upr_bound = q3+(1.5*IQR)
       print(lwr_bound, upr_bound)
       44.0 92.0
[114]: r_outliers = []
       for i in sorted rscore:
            if (i<lwr_bound or i>upr_bound):
                r_outliers.append(i)
       print(r_outliers)
       [1, 16, 120]
[115]: new_df=df
[116]: for i in sample_outliers:
            new_df.drop(i,inplace=True)
       new_df
[116]:
            math score reading score writing score placement score \
                    71
                                     61
                                                                        91
                                                     85
       1
       2
                    79
                                                                        77
                                     16
                                                     87
                    61
                                     77
                                                     74
                                                                        76
       3
       4
                    78
                                     71
                                                     67
                                                                        90
       5
                    73
                                     68
                                                     90
                                                                        80
       6
                    77
                                     62
                                                     70
                                                                        35
       7
                    74
                                     45
                                                     80
                                                                        12
       8
                    76
                                     60
                                                     79
                                                                        77
       9
                    75
                                     65
                                                     85
                                                                       87
                   160
                                     67
                                                     12
                                                                        83
       10
                    79
                                     72
       11
                                                     88
                                                                       180
       13
                    78
                                     69
                                                     71
                                                                        90
       14
                    75
                                      1
                                                     71
                                                                        81
       15
                    78
                                     62
                                                     79
                                                                        93
       18
                    75
                                     62
                                                     86
                                                                        87
       20
                    69
                                     65
                                                     84
                                                                        35
       21
                   100
                                     77
                                                     70
                                                                       91
                                                                       94
       22
                    72
                                                     78
                                     60
       23
                    74
                                     65
                                                     71
                                                                        84
       24
                                     77
                    75
                                                     83
                                                                        77
       25
                   180
                                     67
                                                     63
                                                                        75
       26
                    72
                                    120
                                                     70
                                                                        84
       27
                    71
                                    79
                                                     88
                                                                        85
       28
                   120
                                    73
                                                     71
                                                                        94
```

```
2
       2
                                             2018
                                                  1.897627
       3
                               2
                                             2020 1.785330
                               3
                                             2019 1.892095
       4
       5
                               2
                                            2019 1.863323
       6
                               2
                                            2020 1.886491
       7
                               1
                                             2019 1.869232
       8
                               2
                                             2020 1.880814
                               3
       9
                                             2018 1.875061
       10
                               2
                                            2020
                                                  2.204120
       11
                               2
                                             2019 1.897627
                               3
       13
                                             2019 1.892095
       14
                               2
                                             2019 1.875061
                               3
       15
                                             2021 1.892095
                               3
       18
                                             2019 1.875061
       20
                               1
                                             2018 1.838849
       21
                               3
                                             2018 2.000000
       22
                               3
                                             2019 1.857332
       23
                               2
                                             2019 1.869232
                               2
       24
                                            2020 1.875061
       25
                               3
                                             2021 2.255273
                               2
       26
                                             2021 1.857332
       27
                               3
                                             2021
                                                   1.851258
       28
                               3
                                             2019
                                                  2.079181
[118]: df=pd.read_csv("demo1.csv")
       df stud=df
       ninetieth_percentile = np.percentile(df_stud['math score'], 90)
       b = np.where(df_stud['math score']>ninetieth_percentile,
       ninetieth_percentile, df_stud['math score'])
       print("New array:",b)
                                      78. 73.
                                                 77. 74. 76. 75. 104.
                                                                           79.
                                                                                 80.
      New array: [ 80. 71. 79. 61.
      78.
             78.
                       80.
                            75.
        75.
                  86.
                                 82.
                                       69. 100. 72. 74. 75. 104.
                                                                     72.
       104.]
[119]: df_stud.insert(1,"m score",b,True)
       df_stud
                       m score
[119]:
                               reading score writing score placement score \
           math score
       0
                   80
                          80.0
                                                           70
                                            68
                                                                            89
                          71.0
       1
                   71
                                            61
                                                           85
                                                                            91
       2
                   79
                          79.0
                                            16
                                                           87
                                                                            77
       3
                   61
                          61.0
                                            77
                                                           74
                                                                            76
```

club join year

2019

3

log_math

1.851258

placement offer count

1

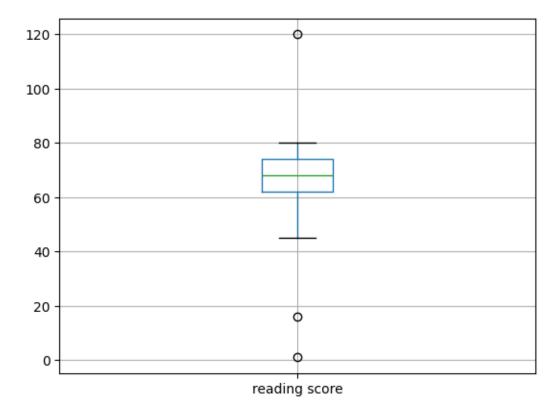
4	78	78.0	71	67	90
5	73	73.0	68	90	80
6	77	77.0	62	70	35
7	74	74.0	45	80	12
8	76	76.0	60	79	77
9	75	75.0	65	85	87
10	160	104.0	67	12	83
11	79	79.0	72	88	180
12	80	80.0	80	78	94
13	78	78.0	69	71	90
14	75	75.0	1	71	81
15	78	78.0	62	79	93
16	86	86.0	78	80	88
17	80	80.0	74	23	76
18	75	75.0	62	86	87
19	82	82.0	70	87	94
20	69	69.0	65	84	35
21	100	100.0	77	70	91
22	72	72.0	60	78	94
23	74	74.0	65	71	84
24	75	75.0	77	83	77
25	180	104.0	67	63	75
26	72	72.0	120	70	84
27	71	71.0	79	88	85
28	120	104.0	73	71	94

	placement	offer	count	club	join	year
0			3			2019
1			3			2019
2			2			2018
3			2			2020
4			3			2019
5			2			2019
6			2			2020
7			1			2019
8			2			2020
9			3			2018
10			2			2020
11			2			2019
12			3			2021
13			3			2019
14			2			2019
15			3			2021
16			3			2019
17			2			2021
18			3			2019
19			3			2019

```
20
                                        2018
                          1
21
                                        2018
                          3
22
                          3
                                        2019
23
                          2
                                        2019
                                        2020
                          2
24
                                        2021
25
                          3
26
                          2
                                        2021
27
                          3
                                        2021
28
                          3
                                        2019
```

```
[124]: import matplotlib.pyplot as plt

col = ['reading score']
  df.boxplot(column=col) # use `column=` to avoid warnings
  plt.show()
```



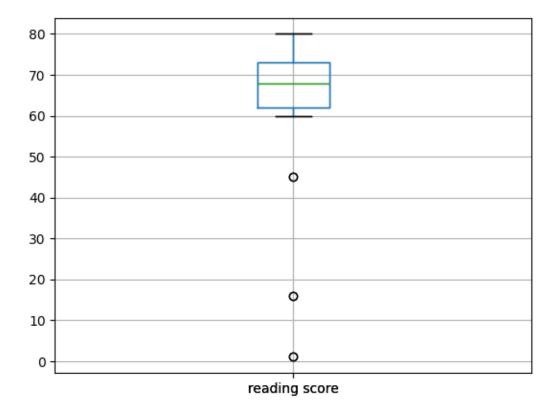
```
[125]: refined_df['reading score'] = np.where(
    refined_df['reading score'] > upr_bound,
    median,
    refined_df['reading score']
)
```

[126]: df [126]: m score reading score writing score placement score \ math score 80.0 71.0 79.0 61.0 78.0 73.0 77.0 74.0 76.0 75.0 104.0 79.0 80.0 78.0 75.0 78.0 86.0 80.0 75.0 82.0 69.0 100.0 72.0 74.0 75.0 104.0 72.0 71.0 104.0 placement offer count club join year

```
13
                                        2019
                          3
14
                          2
                                        2019
15
                          3
                                        2021
16
                          3
                                        2019
                          2
                                        2021
17
18
                          3
                                        2019
19
                          3
                                        2019
20
                          1
                                        2018
21
                          3
                                        2018
                                        2019
22
                          3
                                        2019
23
                          2
24
                          2
                                        2020
25
                          3
                                        2021
                                        2021
26
                          2
27
                          3
                                        2021
28
                          3
                                        2019
```

```
[128]: import matplotlib.pyplot as plt # Make sure this is imported

col = ['reading score']
    refined_df.boxplot(col)
    plt.show() # This will display the boxplot
```



```
import numpy as np
       df=pd.read_csv("demo1.csv")
[130]: df
[130]:
            math score reading score writing score placement score \
       0
                     80
                                      68
                                                      70
                                                                          89
       1
                     71
                                     61
                                                      85
                                                                         91
       2
                     79
                                      16
                                                      87
                                                                         77
       3
                                     77
                     61
                                                      74
                                                                         76
       4
                     78
                                     71
                                                      67
                                                                         90
       5
                     73
                                                      90
                                                                         80
                                      68
       6
                     77
                                      62
                                                      70
                                                                          35
       7
                     74
                                      45
                                                      80
                                                                          12
       8
                     76
                                      60
                                                      79
                                                                         77
       9
                     75
                                      65
                                                      85
                                                                         87
       10
                    160
                                     67
                                                      12
                                                                         83
       11
                     79
                                     72
                                                      88
                                                                        180
       12
                     80
                                     80
                                                      78
                                                                         94
       13
                                      69
                                                      71
                                                                         90
                     78
       14
                     75
                                       1
                                                      71
                                                                         81
       15
                     78
                                     62
                                                      79
                                                                         93
       16
                     86
                                     78
                                                      80
                                                                         88
       17
                     80
                                     74
                                                      23
                                                                         76
       18
                     75
                                     62
                                                      86
                                                                         87
       19
                     82
                                     70
                                                      87
                                                                         94
       20
                     69
                                      65
                                                      84
                                                                          35
                                     77
                                                      70
       21
                    100
                                                                         91
       22
                     72
                                      60
                                                      78
                                                                         94
       23
                     74
                                     65
                                                      71
                                                                         84
       24
                     75
                                     77
                                                      83
                                                                         77
       25
                    180
                                                      63
                                                                         75
                                     67
                     72
       26
                                     120
                                                      70
                                                                         84
       27
                     71
                                     79
                                                      88
                                                                         85
       28
                                                      71
                                                                         94
                    120
                                     73
            placement offer count
                                     club join year
       0
                                  3
                                                 2019
       1
                                  3
                                                 2019
       2
                                  2
                                                 2018
                                  2
       3
                                                 2020
       4
                                  3
                                                 2019
       5
                                  2
                                                 2019
                                  2
                                                 2020
       6
       7
                                  1
                                                 2019
       8
                                  2
                                                 2020
```

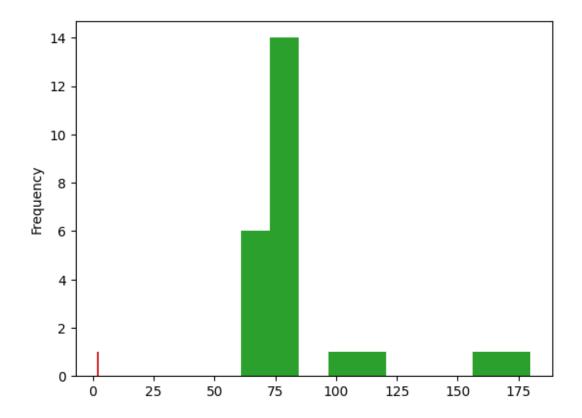
[129]: import pandas as pd

```
9
                                        2018
                         3
10
                         2
                                        2020
11
                         2
                                        2019
12
                         3
                                        2021
                         3
                                        2019
13
                         2
                                        2019
14
15
                         3
                                       2021
16
                         3
                                       2019
17
                         2
                                       2021
18
                         3
                                       2019
19
                         3
                                       2019
20
                         1
                                       2018
21
                                        2018
                         3
                         3
                                        2019
22
23
                         2
                                        2019
24
                         2
                                        2020
25
                         3
                                        2021
26
                         2
                                        2021
27
                         3
                                        2021
28
                         3
                                        2019
```

```
[132]: import matplotlib.pyplot as plt
import numpy as np # Make sure numpy is also imported

new_df['math score'].plot(kind='hist')
df['log_math'] = np.log10(df['math score'])
df['log_math'].plot(kind='hist')

plt.show() # This line displays the plots
```



[]: