ASSIGNMENT NO:2

Aim:

- 1. Creation of Dataset using Microsoft Excel.
- 3. Identification and Handling of Outliers
- 4. Data Transformation for the purpose of : a. To change the scale for better understanding
- b. To decrease the skewness and convert distribution into normal distribution

import pandas as pd

import seaborn as sns

import numpy as py

df= pd.read csv("C:/Users/Welcome/Music/Book1.csv")

df

ioi	math score n year \	reading score	writing score	placememt score	club
0	60.0	63.0	76.0	95.0	
202					
1	75.0	70.0	64.0	85.0	
2020					
2	74.0	50.0	55.0	91.0	
2020 3	68.0	76.0	78.0	97.0	
2020		70.0	70.0	97.0	
4	NaN	67.0	71.0	93.0	
2020					
5	70.0	64.0	80.0	98.0	
2018					
6	61.0	78.0	92.0	94.0	
202		74.0	70.0	MaN	
7 202	61.0	74.0	78.0	NaN	
8	64.0	76.0	79.0	76.0	
2019		7010	7310	7010	
9	65.0	95.0	75.0	90.0	
2020	9				
10	66.0	76.0	NaN	100.0	
2019		67.0	71.0	00.0	
11	84.0	67.0	71.0	92.0	
2020 12	69.0	NaN	70.0	86.0	
202		IVAIN	70.0	00.0	
13	74.0	65.0	65.0	80.0	
202		22.10	3210	22.0	
14	74.0	63.0	72.0	96.0	

2018 15 76.0 2020 16 60.0 2021 17 77.0 2020 18 67.0 2018	64.0 64.0 70.0 95.0 53.0 65.0	80.0 54.0 72.0 NaN 78.0 56.0	96.0 91.0 99.0 87.0 75.0
16 60.0 2021 17 77.0 2020 18 67.0 2018	70.0 95.0 53.0	72.0 NaN 78.0	99.0 87.0
17 77.0 2020 18 67.0 2018	95.0 53.0	NaN 78.0	87.0
18 67.0 2018	53.0	78.0	
10 71 0			75.0
19 71.0 2018	65.0	56.0	
20 58.0 2019		30.0	NaN
21 68.0 2021	63.0	62.0	94.0
22 77.0 2021	63.0	68.0	97.0
23 80.0 2018	NaN	86.0	85.0
24 84.0 2018 25 68.0	63.0 67.0	67.0 73.0	83.0
2019 26 76.0	64.0	68.0	96.0
2021 27 92.0	96.0	61.0	83.0
2018 28 60.0	68.0	59.0	93.0
2020			
placement offer cour 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	female male male male male male male male		

df.isnull() math score reading score writing score placememt score club join year \ 0 False False False False False False False False 1 False False False False False False False False 2 False False False False False False False False 3 False False False False False False False False False 4 True False Fa	18 19 20 21 22 23 24 25 26			2 3 3 3 1 3 3	fema fema fema fema	ale ale ale ale ale ale ale				
math score reading score writing score placememt score club join year \ 0 False False False False False False 1 False False False False False False 2 False False False False False False 3 False False False False False False 4 True False False False False False 5 False False False False False False 6 False False False False False False 7 False False False False False False 8 False False False False False 9 False False False False False False 10 False False False False False False 11 False False False False False False 12 False False False False False False 13 False False False False False False 14 False False False False False False 15 False	28									
join year \ 0 False False False False False False False 1 False False False False False False 2 False False False False False False False 3 False False False False False False False 4 True False False False False False False 5 False False False False False False False 6 False False False False False False False 7 False False False False False False False 8 False False False False False False False 9 False False False False False False 10 False False False False False False 11 False False False False False False 12 False False False False False False 13 False False False False False False 14 False False False False False False 15 False	df.isnul	l()								
False			reading	sco	re	writing	score	placememt	score	club
False 1 False False False False False 2 False False False False 3 False False False False False False 4 True False False False False 5 False False False False False 6 False False False False False 7 False False False False False 8 False False False False False False 9 False False False False False False 10 False False False False False False 11 False False False False False False 12 False False False False False False 13 False False False False False 14 False False False False False False 15 False False False False False 16 False				Fal	se		False		False	
False 2 False False False False False 3 False False False False 4 True False False False False 5 False False False False False 6 False False False False False 7 False False False False False 8 False False False False False 9 False False False False False False 10 False False False False False False 11 False False False False False False 12 False False False False False False 13 False False False False False False 14 False False False False False False 15 False False False False False False 16 False False False False False False 17 False	False									
False		False		Fal	se		False		False	
False False False True False	2	False		Fal	se		False		False	
False 4 True False False False False 5 False False False False 6 False False False False 7 False False False False 8 False False False False False 8 False False False False 9 False False False False 10 False False False False False 11 False False False False False False 12 False True False False 13 False False False False False False 14 False False False False False False 15 False		Ealco		E a l	60		Ealso.		Ealco	
4TrueFalse		гасѕе		гас	se		гасѕе		гасѕе	
False		True		Fal	se		False		False	
False 6 False False False False 7 False False False True False 8 False False False False 9 False False False False 10 False False False False 11 False False False False 12 False True False False 13 False False False False False False 14 False False False False False False 15 False False False False False False 16 False		False		Fal	se		False		False	
False 7 False False False True False 8 False False False False 9 False False False False 10 False False True False False 11 False False False False False False 12 False True False False False 13 False False False False False False 14 False False False False False False 15 False False False False False False 16 False False False False False	False									
7 False False False True False 8 False False False False 9 False False False False 10 False False True False False 11 False False False False False False 12 False True False False False 13 False False False False False False 14 False False False False False False 15 False False False False False False 16 False False False False False		False		Fal	se		False		False	
8FalseFalseFalseFalseFalseFalse9FalseFalseFalseFalseFalseTrueFalse10FalseFalseFalseFalseFalseFalseFalse11FalseFalseFalseFalseFalseFalseFalse12FalseFalseFalseFalseFalseFalseFalseFalseFalse14FalseFalseFalseFalse15FalseFalseFalseFalse16FalseFalseFalseFalse		False		Fal	se		False		True	
False 9 False False False False 10 False False True False 11 False False False False 12 False True False False 13 False False False False 14 False False False False 15 False False False False 16 False False False False 17 False False False False 18 False False False False 19 False False False False 19 False False False False		Ealco		E-1	60		Ealso		Falso	
False 10 False False True False False 11 False False False False 12 False True False False False 13 False False False False False 14 False False False False False 15 False False False False False 16 False False False False		гасѕе		гас	se		гасѕе		гасѕе	
10FalseFalseTrueFalseFalseFalseFalseFalse12FalseTrueFalseFalseFalseFalseFalseFalse13FalseFalseFalseFalseFalseFalseFalseFalse14FalseFalseFalseFalse15FalseFalseFalseFalseFalseFalseFalseFalse	9	False		Fal	se		False		False	
False 11 False False False False False 12 False True False False False 13 False False False False False 14 False False False False False 15 False		False		Fal	SE		True		False	
False 12 False True False False False 13 False False False False False 14 False False False False False 15 False False False False False 16 False False False False	False									
12 False True False False False 13 False False False False False 14 False False False False False 15 False False False False False 16 False False False False		False		Fal	se		False		False	
13FalseFalseFalseFalseFalseFalse14FalseFalseFalseFalseFalseFalse15FalseFalseFalseFalseFalseFalse		False		Tr	ue		False		False	
False 14 False False False False 15 False False False False 16 False False False False		F.1					F-1		F.1	
14FalseFalseFalseFalseFalseFalse15FalseFalseFalseFalseFalseFalse		ratse		ral	se		ratse		ratse	
15 False False False False False False False False	14	False		Fal	se		False		False	
False False False False		False		Fal	C A		Falso		False	
16 False False False		10136		ıat	36		Tacse		Tacse	
1 2 2 2		False		Fal	.se		False		False	

17 False	False	False	False	False	
18	False	False	True	False	
False 19	False	False	False	False	
False 20	False	False	False	True	
False					
21 False	False	False	False	False	
22	False	False	False	False	
False 23	False	True	False	False	
False	14136	True	racse	racse	
24 False	False	False	False	False	
25	False	False	False	False	
False 26	False	False	False	False	
False	racse	racse	racse	racse	
27	False	False	False	False	
False 28	False	False	False	False	
False					
plac	cement offer	count gender			
0	cement offer	False False			
0	cement offer	False False False False			
0	cement offer	False False False False False False			
0	cement offer	False False False False			
0	cement offer	False False False False False False			
0	cement offer	False			
0 1 2 3 4 5 6 7	cement offer	False			
0 1 2 3 4 5 6 7	cement offer	False			
0 1 2 3 4 5 6 7 8	cement offer	False			
0 1 2 3 4 5 6 7 8 9	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	cement offer	False			
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	cement offer	False			

```
23
                    False
                             False
24
                    False
                             False
25
                    False
                             False
26
                    False
                             False
                             False
27
                    False
28
                    False
                             False
series = pd.isnull(df["math score"])
df[series]
   math score reading score writing score placement score club
join year \
                        67.0
                                        71.0
                                                         93.0
          NaN
2020
   placement offer count gender
4
                           male
df.notnull()
    math score reading score writing score placement score club
join year ∖
          True
                         True
                                         True
                                                          True
True
          True
                         True
                                         True
                                                          True
True
          True
                         True
                                         True
                                                          True
2
True
                         True
                                         True
3
          True
                                                          True
True
         False
                         True
                                         True
                                                          True
True
                         True
                                         True
                                                          True
5
          True
True
          True
                         True
                                         True
                                                          True
True
          True
                         True
                                         True
                                                          False
7
True
                         True
                                         True
                                                          True
          True
True
                         True
                                         True
9
          True
                                                          True
True
10
          True
                         True
                                        False
                                                          True
True
11
          True
                         True
                                         True
                                                          True
True
12
                        False
          True
                                         True
                                                          True
True
13
                         True
                                         True
                                                          True
          True
True
```

15	14 True	True	True	True	True
16 True True True True 17 True True True True 17 True True True True 18 True True True True 19 True True True True 20 True True True True 20 True True True True 10 True True True True 11 True True True True 12 True True True True		True	True	True	True
True True True True True True True True	16	True	True	True	True
18 True True False True True 19 True True True True True 20 True True True True 21 True True True True True 21 True True True True True 22 True True True True True 23 True False True True True 24 True True True True True 25 True True True True True 26 True True True True True 27 True True 28 True True True True True 29 True True True True True 20 True True True True 21 True True 22 True True True True True 23 True True True True 24 True True 25 True True True True 26 True True True True 27 True True True 28 True True True 29 True True 30 True True 4 True True 5 True True 6 True True 7 True True 7 True True 8 True True 9 True True 10 True True 11 True True	17	True	True	True	True
19 True True True True True True 20 True True True True False True 21 True True True True True True 22 True True True True True True 23 True False True True True True 24 True True True True True True 25 True True True True True True 26 True True True True True True 27 True True True True True True 28 True True True True True True placement offer count gender True	18	True	True	False	True
True True True True True True True True	19	True	True	True	True
True	20	True	True	True	False
True	21	True	True	True	True
True	22	True	True	True	True
24 True True True True True True 25 True True True True True 26 True True True True True 27 True True True True True 28 True True True True True True placement offer count gender frue True True	23	True	False	True	True
25 True True True True True 26 True True True 27 True True True True 27 True True True True 28 True True True True True placement offer count gender 0 True True 1 True True 2 True True 2 True True 3 True True 4 True True 5 True True 6 True True 7 True True 8 True True 9 True True 10 True True 11 True True 11 True True 12 True True 11 True True	24	True	True	True	True
True True True True True True True True	25	True	True	True	True
True True True True True True True	26	True	True	True	True
True True placement offer count gender True	27	True	True	True	True
placement offer count gender True True True True True True True True True	28	True	True	True	True
0 True True 1 True True 2 True True 3 True True 4 True True 5 True True 6 True True 7 True True 8 True True 9 True True 10 True True 11 True True 12 True True 13 True True 14 True True					
16 True True	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	cement offer	True		

17	True	True
18	True	True
19	True	True
20	True	True
21	True	True
22	True	True
23	True	True
24	True	True
25	True	True
26	True	True
27	True	True
28	True	True

series1 = pd.notnull(df["math score"])
df[series1]

		reading score	writing score	placememt score	club
) 0 1 r	n year \ 60.0	63.0	76.0	95.0	
2021 1	1 75.0	70.0	64.0	85.0	
2020		70.0	04.0	83.0	
2	74.0	50.0	55.0	91.0	
2020 3	68.0	76.0	78.0	97.0	
2020 5		64.0	90.0	00.0	
2018	70.0 3	64.0	80.0	98.0	
6 2023	61.0	78.0	92.0	94.0	
7	61.0	74.0	78.0	NaN	
2021 8	1 64.0	76.0	79.0	76.0	
2019		70.0	79.0	70.0	
9 2020	65.0	95.0	75.0	90.0	
10	66.0	76.0	NaN	100.0	
2019 11	84.0	67.0	71.0	92.0	
2020		07.0	71.0	92.0	
12 202	69.0	NaN	70.0	86.0	
13	74.0	65.0	65.0	80.0	
2021 14	1 74.0	63.0	72.0	96.0	
2018	3				
15 2020	76.0	64.0	80.0	96.0	
16	60.0	64.0	54.0	91.0	
202	l				

17	77.0	70.0	72.0	99.0	
2020 18	67.0	95.0	NaN	87.0	
2018 19	71.0	53.0	78.0	75.0	
2018 20	58.0	65.0	56.0	NaN	
2019					
21 2021	68.0	63.0	62.0	94.0	
22 2021	77.0	63.0	68.0	97.0	
23 2018	80.0	NaN	86.0	85.0	
24 2018	84.0	63.0	67.0	83.0	
25 2019	68.0	67.0	73.0	88.0	
26	76.0	64.0	68.0	96.0	
2021 27	92.0	96.0	61.0	83.0	
2018 28	60.0	68.0	59.0	93.0	
2020					
place	ement offer	_			
place	ement offer	3 female			
place	ement offer	3 female			
place	ement offer	3 female			
place	ement offer	3 female			
place	ement offer	3 female 3 male 3 male 3 female 3 female 3 male 2 male 2 male			
place 0 1 2 3 5 6 7 8 9	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 18	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male			
place 0 1 2 3 5 6 7 8 9 10 11 12 13 14 15 16 17 18	ement offer	3 female 3 male 3 male 3 female 3 female 2 male 2 male 3 female 1 male 3 male 3 male 3 male 3 male 3 male 5 female 4 male 5 female 6 male 7 male 8 male 9 male 9 female 9 female 9 female			

```
24
                         1
                               male
                            female
25
                         3
                         3
26
                             female
27
                         2
                               male
                         3
                               male
28
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
newdf = df
df
    math score reading score writing score placement score club
join year
          60.0
                          63.0
                                          76.0
                                                            95.0
2021
          75.0
                          70.0
                                          64.0
                                                            85.0
2020
          74.0
                          50.0
                                          55.0
                                                            91.0
2
2020
3
          68.0
                          76.0
                                          78.0
                                                            97.0
2020
                          67.0
                                                            93.0
           NaN
                                          71.0
2020
          70.0
                          64.0
                                          80.0
                                                            98.0
2018
                                          92.0
                                                            94.0
          61.0
                          78.0
2021
7
          61.0
                          74.0
                                          78.0
                                                              NaN
2021
          64.0
                          76.0
                                          79.0
                                                            76.0
2019
9
          65.0
                          95.0
                                          75.0
                                                            90.0
2020
10
          66.0
                          76.0
                                           NaN
                                                           100.0
2019
11
          84.0
                          67.0
                                          71.0
                                                            92.0
2020
12
          69.0
                           NaN
                                          70.0
                                                            86.0
2021
13
          74.0
                          65.0
                                          65.0
                                                            80.0
2021
14
          74.0
                          63.0
                                          72.0
                                                            96.0
2018
                          64.0
15
          76.0
                                          80.0
                                                            96.0
2020
                          64.0
                                          54.0
                                                            91.0
16
          60.0
2021
17
          77.0
                          70.0
                                          72.0
                                                            99.0
2020
```

18	67.0	95.0	NaN	87.0
2018 19	71.0	53.0	78.0	75.0
2018 20	58.0	65.0	56.0	NaN
2019 21	68.0	63.0	62.0	94.0
2021				
22 2021	77.0	63.0	68.0	97.0
23 2018	80.0	NaN	86.0	85.0
24 2018	84.0	63.0	67.0	83.0
25 2019	68.0	67.0	73.0	88.0
26 2021	76.0	64.0	68.0	96.0
27	92.0	96.0	61.0	83.0
2018 28	60.0	68.0	59.0	93.0
2020				
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	ement offer.	count gender 3 0 3 1 3 1 3 0 3 1 3 0 3 1 2 1 2 1 2 1 3 0 1 1 3 1 3 0 2 1 3 1 3 0 2 1 3 1 3 0 2 1 3 1 3 0 2 1 3 1 3 0 2 1 1 1 1 3 0 2 1 1 3 1 1 3 0 1 1 1 1 1 3 0 1 1 1 1 1 1 1 1 1 1		

```
25
                         3
                                  0
                         3
26
                                  0
                         2
27
                                  1
28
missing values = ["Na", "na"]
df = pd.read_csv("C:/Users/Welcome/Music/Book1.csv", na_values =
missing values)
df
    math score reading score writing score placement score club
join year
          60.0
                          63.0
                                          76.0
                                                            95.0
2021
          75.0
                          70.0
                                          64.0
                                                            85.0
2020
          74.0
                          50.0
                                          55.0
                                                            91.0
2020
          68.0
                          76.0
                                          78.0
                                                            97.0
2020
           NaN
                          67.0
                                          71.0
                                                            93.0
2020
          70.0
                          64.0
                                          80.0
                                                            98.0
2018
          61.0
                          78.0
                                          92.0
                                                            94.0
2021
          61.0
                          74.0
                                          78.0
                                                             NaN
2021
          64.0
                                                            76.0
                          76.0
                                          79.0
2019
          65.0
                          95.0
                                          75.0
                                                            90.0
2020
10
          66.0
                          76.0
                                           NaN
                                                           100.0
2019
11
          84.0
                          67.0
                                          71.0
                                                            92.0
2020
12
          69.0
                           NaN
                                          70.0
                                                            86.0
2021
13
          74.0
                          65.0
                                          65.0
                                                            80.0
2021
          74.0
                          63.0
                                          72.0
                                                            96.0
14
2018
15
          76.0
                          64.0
                                          80.0
                                                            96.0
2020
                          64.0
16
          60.0
                                          54.0
                                                            91.0
2021
          77.0
                          70.0
                                          72.0
                                                            99.0
17
2020
18
          67.0
                          95.0
                                           NaN
                                                            87.0
```

2018

19	71.0	53.0	78.0	75.0
2018 20	58.0	65.0	56.0	NaN
2019 21	68.0	63.0	62.0	94.0
2021 22	77.0	63.0	68.0	97.0
2021				
23 2018	80.0	NaN	86.0	85.0
24 2018	84.0	63.0	67.0	83.0
25 2019	68.0	67.0	73.0	88.0
26 2021	76.0	64.0	68.0	96.0
27 2018	92.0	96.0	61.0	83.0
28	60.0	68.0	59.0	93.0
2020				
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26		female male male male male male male male		

27 28				nale nale				
<pre>ndf = df ndf.filln</pre>	a(<mark>0</mark>)							
math join year	score	reading	score	writing	score	placememt	score	club
0 2021	60.0		63.0		76.0		95.0	
1 2020	75.0		70.0		64.0		85.0	
2020	74.0		50.0		55.0		91.0	
3	68.0		76.0		78.0		97.0	
2020 4	0.0		67.0		71.0		93.0	
2020 5	70.0		64.0		80.0		98.0	
2018 6	61.0		78.0		92.0		94.0	
2021 7	61.0		74.0		78.0		0.0	
2021								
8 2019	64.0		76.0		79.0		76.0	
9 2020	65.0		95.0		75.0		90.0	
10 2019	66.0		76.0		0.0		100.0	
11	84.0		67.0		71.0		92.0	
2020 12	69.0		0.0		70.0		86.0	
2021 13	74.0		65.0		65.0		80.0	
2021 14	74.0		63.0		72.0		96.0	
2018 15	76.0		64.0		80.0		96.0	
2020								
16 2021	60.0		64.0		54.0		91.0	
17 2020	77.0		70.0		72.0		99.0	
18 2018	67.0		95.0		0.0		87.0	
19	71.0		53.0		78.0		75.0	
2018 20 2019	58.0		65.0		56.0		0.0	

21 68. 2021 22 77. 2021 23 80. 2018 24 84. 2018 25 68. 2019 26 76. 2021 27 92. 2018 28 60. 2020 placement 0 1 2 3 4 5 6 7 8 9	0 63	. 0	62.0	04.0
22 77. 2021 23 80. 2018 24 84. 2018 25 68. 2019 26 76. 2021 27 92. 2018 28 60. 2020			02.0	94.0
23 80. 2018 24 84. 2018 25 68. 2019 26 76. 2021 27 92. 2018 28 60. 2020	0 63	.0	68.0	97.0
24 84. 2018 25 68. 2019 26 76. 2021 27 92. 2018 28 60. 2020	0 0	.0	86.0	85.0
25 68. 2019 26 76. 2021 27 92. 2018 28 60. 2020	0 63	.0	67.0	83.0
26 76. 2021 27 92. 2018 28 60. 2020	0 67	.0	73.0	88.0
27 92. 2018 28 60. 2020	0 64	. 0	68.0	96.0
28 60. 2020 placement	0 96	.0	61.0	83.0
placement	0 68	.0	59.0	93.0
9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27	3 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3	gender female male female female male male male male female male female male female male female		

```
m v=df['math score'].mean()
df['math score'].fillna(value = m v, inplace = True)
df
    math score reading score writing score placement score club
join year \
     60.000000
                          63.0
                                          76.0
                                                            95.0
2021
                          70.0
                                          64.0
                                                            85.0
     75.000000
2020
     74.000000
                          50.0
                                          55.0
                                                            91.0
2020
                                                            97.0
     68.000000
                          76.0
                                          78.0
3
2020
4
     70.678571
                          67.0
                                          71.0
                                                            93.0
2020
5
     70.000000
                          64.0
                                          80.0
                                                            98.0
2018
6
     61.000000
                          78.0
                                          92.0
                                                            94.0
2021
     61.000000
7
                          74.0
                                          78.0
                                                             NaN
2021
8
     64.000000
                          76.0
                                          79.0
                                                            76.0
2019
9
     65.000000
                          95.0
                                          75.0
                                                            90.0
2020
     66.000000
                          76.0
                                           NaN
                                                           100.0
10
2019
11
     84.000000
                          67.0
                                          71.0
                                                            92.0
2020
12
     69.000000
                           NaN
                                          70.0
                                                            86.0
2021
13
                          65.0
                                                            80.0
     74.000000
                                          65.0
2021
14
     74.000000
                          63.0
                                          72.0
                                                            96.0
2018
15
     76.000000
                          64.0
                                          80.0
                                                            96.0
2020
16
                          64.0
                                          54.0
                                                            91.0
     60.000000
2021
17
     77.000000
                          70.0
                                          72.0
                                                            99.0
2020
18
                          95.0
     67.000000
                                           NaN
                                                            87.0
2018
19
                                          78.0
     71.000000
                          53.0
                                                            75.0
2018
                                          56.0
20
     58.000000
                          65.0
                                                             NaN
2019
21
     68.000000
                          63.0
                                          62.0
                                                            94.0
2021
```

```
22
     77.000000
                           63.0
                                           68.0
                                                              97.0
2021
23
     80.000000
                            NaN
                                           86.0
                                                              85.0
2018
24
     84.000000
                           63.0
                                           67.0
                                                              83.0
2018
25
     68.000000
                           67.0
                                           73.0
                                                              88.0
2019
     76.000000
                           64.0
26
                                           68.0
                                                              96.0
2021
27
     92.000000
                           96.0
                                           61.0
                                                              83.0
2018
28
     60.000000
                           68.0
                                           59.0
                                                              93.0
2020
    placement offer count
                             gender
                             female
0
1
                          3
                               male
2
                          3
                               male
3
                          3
                             female
4
                          3
                               male
5
                          3
                             female
6
                          3
                               male
7
                          2
                               male
8
                          2
                               male
9
                          3
                            female
10
                          1
                               male
11
                          3
                               male
                          3
                             female
12
                          2
13
                               male
14
                          3
                               male
                               male
15
                          3
                          3
16
                             female
                          3
                               male
17
                          3
18
                             female
                          2
19
                             female
20
                          3
                             female
21
                          3
                               male
                          3
                            female
22
                          3
23
                             female
24
                          1
                               male
25
                             female
                          3
26
                          3
                             female
27
                          2
                               male
28
                          3
                               male
ndf.replace(to_replace = py.nan, value = -99)
    math score reading score writing score placement score club
join year \
```

0 2021	60.000000	63.0	76.0	95.0
1 2020	75.000000	70.0	64.0	85.0
2	74.000000	50.0	55.0	91.0
2020	68.000000	76.0	78.0	97.0
2020 4	70.678571	67.0	71.0	93.0
2020 5	70.000000	64.0	80.0	98.0
2018 6	61.000000	78.0	92.0	94.0
2021 7	61.000000	74.0	78.0	-99.0
2021 8	64.000000	76.0	79.0	76.0
2019 9	65.000000	95.0	75.0	90.0
2020 10	66.000000	76.0	-99.0	100.0
2019 11	84.000000	67.0	71.0	92.0
2020 12	69.000000	-99.0	70.0	86.0
2021 13	74.000000	65.0	65.0	80.0
2021 14	74.000000	63.0	72.0	96.0
2018 15	76.000000	64.0	80.0	96.0
2020 16	60.000000	64.0	54.0	91.0
2021 17	77.000000	70.0	72.0	99.0
2020 18	67.000000		-99.0	87.0
2018 19	71.000000	53.0	78.0	75.0
2018 20	58.000000	65.0		-99.0
2019 21	68.000000	63.0	62.0	94.0
2021 22	77.000000			97.0
2021		63.0	68.0	
23 2018		-99.0	86.0	85.0
24	84.000000	63.0	67.0	83.0

2018 25	68.000000	67.0	73.0	88.0
2019 26	76.000000	64.0	68.0	96.0
2021 27	92.000000	96.0	61.0	83.0
2018	60.000000	68.0	59.0	93.0
2020				
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	dropna()	gender female male male male male male male male		
	•			
	math score reading year \ 60.000000	score writing 63.0	score placememt 76.0	score club 95.0
2021 1	75.000000	70.0	64.0	85.0
2020				
2	74.000000	50.0	55.0	91.0

2222				
2020 3 2020	68.000000	76.0	78.0	97.0
4 2020	70.678571	67.0	71.0	93.0
5 2018	70.000000	64.0	80.0	98.0
6 2021	61.000000	78.0	92.0	94.0
8 2019	64.000000	76.0	79.0	76.0
9 2020	65.000000	95.0	75.0	90.0
11 2020	84.000000	67.0	71.0	92.0
13 2021	74.000000	65.0	65.0	80.0
14 2018	74.000000	63.0	72.0	96.0
15 2020	76.000000	64.0	80.0	96.0
16 2021	60.000000	64.0	54.0	91.0
17 2020	77.000000	70.0	72.0	99.0
19 2018	71.000000	53.0	78.0	75.0
21 2021	68.000000	63.0	62.0	94.0
22 2021	77.000000	63.0	68.0	97.0
24 2018	84.000000	63.0	67.0	83.0
25 2019	68.000000	67.0	73.0	88.0
26 2021	76.000000	64.0	68.0	96.0
27 2018	92.000000	96.0	61.0	83.0
28 2020	60.000000	68.0	59.0	93.0
0 1 2 3 4 5		t gender 3 female 3 male 3 male 3 female 3 male 5 male		

6		3	male
8		2	male
6 8 9		3	female
11		3	male
13			male
14		3	male
15		2 3 3 3 3	male
16		3	female
17		3	male
19		2	female
21		3	male
22		3	female
24		1	male
25		3	female
26		3	female
27		2	male
28		3	male
16 1	/ 1		

ndf.dropna(how = 'all')

		reading score	writing score	placememt score	club
0	year \ 60.00000	63.0	76.0	95.0	
	75.000000	70.0	64.0	85.0	
2020 2	74.000000	50.0	55.0	91.0	
2020 3	68.000000	76.0	78.0	97.0	
2020 4	70.678571	67.0	71.0	93.0	
2020 5	70.000000	64.0	80.0	98.0	
2018 6	61.000000	78.0	92.0	94.0	
2021	61.000000	74.0	78.0	NaN	
2021	64.000000	76.0	79.0	76.0	
2019 9	65.000000	95.0	75.0	90.0	
2020 10	66.000000	76.0	NaN	100.0	
2019 11	84.000000	67.0	71.0	92.0	
2020 12	69.000000	NaN	70.0	86.0	
2021 13 2021	74.000000	65.0	65.0	80.0	

14 2018	74.000000	63.0	72.0	96.0
15	76.000000	64.0	80.0	96.0
2020 16	60.000000	64.0	54.0	91.0
2021 17	77.000000	70.0	72.0	99.0
2020	67.000000	95.0	NaN	87.0
2018 19	71.000000	53.0	78.0	75.0
2018	58.000000	65.0	56.0	NaN
2019	68.000000	63.0	62.0	94.0
2021	77.000000	63.0	68.0	97.0
2021 23 2018	80.000000	NaN	86.0	85.0
24 2018	84.000000	63.0	67.0	83.0
25 2019	68.000000	67.0	73.0	88.0
26 2021	76.000000	64.0	68.0	96.0
27 2018	92.000000	96.0	61.0	83.0
28 2020	60.000000	68.0	59.0	93.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14		female male male male male male male male		
15 16		3 male 3 male 3 female		

```
17
                           3
                                male
                           3
18
                              female
                           2
19
                              female
20
                           3
                              female
                           3
                                male
21
22
                           3
                              female
                           3
23
                              female
24
                           1
                                male
                           3
25
                              female
                           3
26
                              female
                           2
27
                                male
                           3
28
                                male
ndf.dropna(axis = 1)
                                    placement offer count
    math score
                  club join year
                                                              gender
0
     60,000000
                                                              female
                             2021
1
     75.000000
                                                           3
                             2020
                                                                male
2
     74.000000
                             2020
                                                           3
                                                                male
3
                                                           3
     68.000000
                             2020
                                                              female
                                                          3
4
     70.678571
                             2020
                                                                male
                                                          3
5
     70.000000
                             2018
                                                              female
6
                                                          3
     61.000000
                             2021
                                                                male
                                                          2
7
     61.000000
                             2021
                                                                male
8
     64.000000
                                                           2
                                                                male
                             2019
                                                          3
9
     65.000000
                             2020
                                                              female
10
                                                           1
     66.000000
                             2019
                                                                male
                                                           3
11
     84.000000
                             2020
                                                                male
                                                          3
                                                              female
12
     69.000000
                             2021
                                                          2
13
     74.000000
                             2021
                                                                male
14
     74.000000
                             2018
                                                           3
                                                                male
                                                           3
15
                             2020
     76,000000
                                                                male
                                                           3
16
     60.000000
                             2021
                                                              female
17
     77.000000
                             2020
                                                           3
                                                                male
                                                           3
18
     67.000000
                                                              female
                             2018
                                                          2
19
     71.000000
                             2018
                                                              female
                                                          3
                                                              female
20
     58.000000
                             2019
                                                           3
21
     68.000000
                             2021
                                                                male
22
     77.000000
                                                           3
                                                              female
                             2021
23
     80.000000
                                                           3
                                                              female
                             2018
24
                                                           1
     84.000000
                             2018
                                                                male
                                                           3
25
     68.000000
                             2019
                                                              female
                                                          3
26
     76.000000
                             2021
                                                              female
                                                           2
27
     92,000000
                             2018
                                                                male
                                                           3
28
     60.000000
                             2020
                                                                male
new_data = ndf.dropna(axis = 0, how='any')
new data
```

	math score	reading score	writing score	placememt score	club
Ō	year \ 60.000000	63.0	76.0	95.0	
2021	75.000000	70.0	64.0	85.0	
2020	74.000000	50.0	55.0	91.0	
2020	68.000000	76.0	78.0	97.0	
2020	70.678571	67.0	71.0	93.0	
2020 5	70.000000	64.0	80.0	98.0	
2018 6	61.000000	78.0	92.0	94.0	
2021 8	64.000000	76.0	79.0	76.0	
2019 9	65.000000	95.0	75.0	90.0	
2020 11	84.000000	67.0	71.0	92.0	
2020 13	74.000000	65.0	65.0	80.0	
2021 14	74.000000	63.6	72.0	96.0	
2018 15	76.000000	64.0	80.0	96.0	
2020 16	60.000000	64.0	54.0	91.0	
2021 17	77.000000	70.0	72.0	99.0	
2020 19	71.000000	53.0	78.0	75.0	
2018 21	68.000000	63.6	62.0	94.0	
2021 22	77.000000	63.6	68.0	97.0	
2021 24	84.000000	63.6	67.0	83.0	
2018 25	68.000000	67.0	73.0	88.0	
2019 26	76.000000	64.0	68.0	96.0	
2021 27	92.000000	96.0	61.0	83.0	
2018 28	60.000000	68.6	59.0	93.0	
2020					
þ	olacement o	ffer count ge	nder		

0	3	female
1	3	male
2	3	
2		male
3	3	female
4	3	male
5	3 3 3	female
6	3	male
2 3 4 5 6 8 9	2	male
9	3	female
11	3	male
13	2	male
14		male
	2	
15	3	male
16	3 3 3 3	female
17	3	male
19	2	female
21	3	male
22	3	female
24	1	male
25	3	female
26	3	female
27	2	
		male
28	3	male

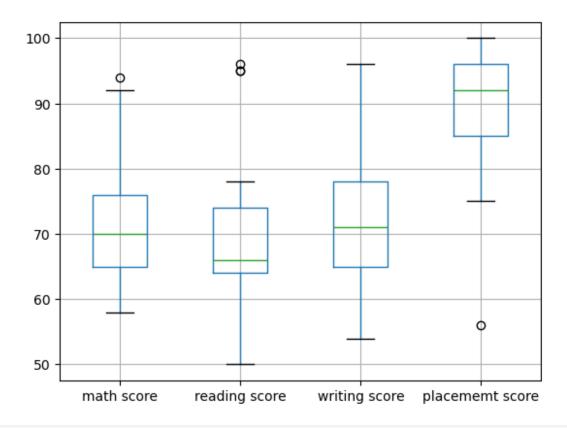
import matplotlib.pyplot as plt

dfl= pd.read_csv("C:/Users/Welcome/Music/Book3.csv")
df1

math score	reading score	writing score	placememt score	club
join year ∖				
0 60	63	76	95	
2021				
1 75	70	64	85	
2020			0.1	
2 74	50	55	91	
2020	76	70	07	
3 68 2020	76	78	97	
4 94	67	71	93	
2020	07	/ 1	93	
5 70	64	80	98	
2018	<u> </u>			
6 61	78	92	94	
2021				
7 61	74	78	80	
2021				
8 64	76	79	76	
2019				
9 65	95	75	90	

2020				
10	66	76	67	100
2019				
11	84	67	71	92
2020 12	60	66	70	E
2021	69	66	70	56
13	74	65	65	80
2021				
14	74	63	72	96
2018				
15	76	64	80	96
2020 16	60	64	54	91
2021	00	04	J 4	91
17	77	70	72	99
2020				
18	67	95	64	87
2018				
19	71	65	78	75
2018 20	58	65	96	92
2019	50	03	30	32
21	68	63	62	94
2021				
22	77	63	68	97
2021	0.0	C 4	0.0	0.5
23 2018	80	64	86	85
24	84	63	67	83
2018	01	03	0,	33
25	68	67	73	88
2019				
26	76	64	68	96
2021 27	92	96	61	83
2018	92	90	01	63
28	60	68	59	93
2020				
_				
	ment offer co			
1		<pre>3 female 3 male</pre>		
2		3 male		
3		3 female		
4		3 male		
5		3 female		
0 1 2 3 4 5 6 7				
1		2 male		

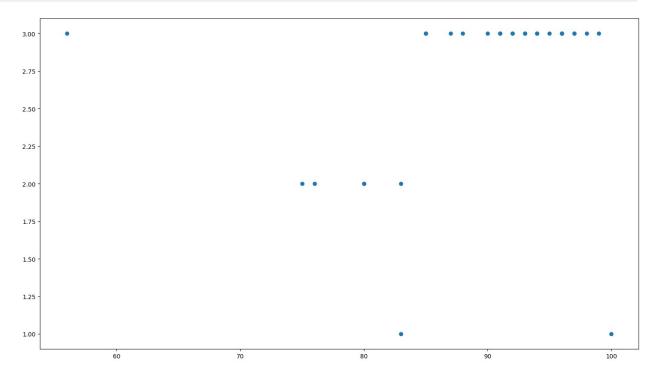
```
2 male
3 female
1 male
8
10
11
                               3
                                     male
                               3 female
12
                               2
13
                                     male
14
                                    male
                               3
                                   male
15
                              3 female
3 male
3 female
2 female
16
17
18
19
                              3 female
3 male
3 female
3 female
1 male
20
21
22
23
24
25
                               3 female
                               3 female
26
                               2
27
                                     male
28
                                     male
col = ['math score', 'reading score', 'writing score', 'placememt
score']
df1.boxplot(col)
<Axes: >
```



```
print(py.where(df1['math score']>90))
print(py.where(df1['reading score']<25))</pre>
print(py.where(df1['writing score']<30))</pre>
(array([ 4, 27], dtype=int64),)
(array([], dtype=int64),)
(array([], dtype=int64),)
pip install matplotlib
Requirement already satisfied: matplotlib in c:\users\welcome\
anaconda3\lib\site-packages (3.8.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy<2,>=1.21 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (23.1)
```

```
Requirement already satisfied: pillow>=6.2.0 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\welcome\
anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\
welcome\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\welcome\anaconda3\
lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.

fig, ax= plt.subplots(figsize = (18, 10))
ax.scatter(df1['placememt score'], df1['placement offer count'])
plt.show()
ax.set_xlabel('(Proportion non-retail business acres)/(town)')
ax.set_ylabel('(Full-value property-tax rate)/($10,000)')
```



```
Text(4.44444444444452, 0.5, '(Full-value property-tax
rate)/($10,000)')

print(py.where((df1['placememt score']<50) & (df1['placement offer
count']>1)))
print(py.where((df1['placememt score']>85) & (df1['placement offer
count']<3)))

(array([], dtype=int64),)
(array([10], dtype=int64),)
from scipy import stats</pre>
```

```
z = py.abs(stats.zscore(df1['math score']))
print(z)
0
      1.252553
1
      0.383665
2
      0.274584
3
      0.379903
4
      2.456207
5
      0.161741
6
      1.143471
7
      1.143471
8
      0.816228
9
      0.707147
10
      0.598066
11
      1.365395
12
      0.270822
13
      0.274584
14
      0.274584
15
      0.492746
16
      1.252553
17
      0.601827
18
      0.488984
19
      0.052660
20
      1.470715
21
      0.379903
22
      0.601827
23
      0.929071
24
      1.365395
25
      0.379903
26
      0.492746
27
      2.238044
28
      1.252553
Name: math score, dtype: float64
threshold = 0.18
sample outliers = py.where(z<threshold)</pre>
sample outliers
(array([ 5, 19], dtype=int64),)
sorted_rscore = sorted(df1['reading score'])
sorted rscore
[50,
 63,
 63,
 63,
 63,
```

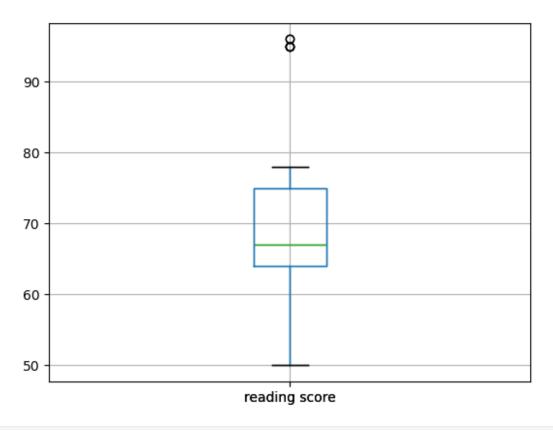
```
63,
 64,
 64,
 64,
 64,
 64,
 65,
 65,
 65,
 66,
 67,
 67,
 67,
 68,
 70,
 70,
 74,
 76,
 76,
 76,
 78,
 95,
 95,
 96]
q1 = py.percentile(sorted_rscore, 25)
q3 = py.percentile(sorted_rscore, 75)
print(q1, q3)
64.0 74.0
IQR = q3-q1
lwr bound = q1-(1.5*IQR)
upr bound = q3+(1.5*IQR)
print(lwr_bound, upr_bound)
49.0 89.0
r outliers = []
for i in sorted rscore:
    if(i<lwr_bound or i>upr_bound):
        r_outliers.append(i)
print(r_outliers)
[95, 95, 96]
new df = df1
for i in sample_outliers:
    new df.drop(i,inplace=True)
new df
```

		core	reading	score	writing	score	placememt	score	club
	year	\							
0		60		63		76		95	
2021									
1		75		70		64		85	
2020									
2		74		50		55		91	
2020									
3		68		76		78		97	
2020									
4		94		67		71		93	
2020									
6		61		78		92		94	
2021									
7		61		74		78		80	
2021									
8		64		76		79		76	
2019									
9		65		95		75		90	
2020									
10		66		76		67		100	
2019	1								
11		84		67		71		92	
2020	1								
12		69		66		70		56	
2021									
13		74		65		65		80	
2021									
14		74		63		72		96	
2018									
15		76		64		80		96	
2020									
16		60		64		54		91	
2021									
17		77		70		72		99	
2020									
18		67		95		64		87	
2018									
20		58		65		96		92	
2019									
21		68		63		62		94	
2021									
22		77		63		68		97	
2021									
23		80		64		86		85	
2018									
24		84		63		67		83	
2018									
25		68		67		73		88	
2019									

```
26
            76
                             64
                                             68
                                                               96
2021
            92
27
                             96
                                             61
                                                               83
2018
                             68
                                             59
                                                               93
28
            60
2020
    placement offer count
                             gender
                             female
0
1
                         3
                               male
2
                         3
                               male
                         3
3
                            female
4
                         3
                               male
                         3
6
                               male
7
                         2
                               male
8
                         2
                               male
9
                         3
                            female
10
                         1
                               male
11
                         3
                               male
                         3
12
                            female
                         2
13
                               male
                         3
14
                               male
                         3
15
                               male
                         3
16
                            female
17
                         3
                               male
                         3
18
                            female
                         3
20
                            female
                         3
21
                               male
                         3
22
                            female
23
                         3
                            female
                         1
24
                               male
25
                         3
                             female
                         3
26
                             female
                         2
27
                               male
28
                         3
                               male
df stud = df1
ninetieth percentile = py.percentile(df stud['math score'], 90)
b = py.where(df stud['math
score']>ninetieth percentile,ninetieth percentile, df stud['math
score'])
print("New array:" ,b)
New array: [60. 75. 74. 68. 84. 61. 61. 64. 65. 66. 84. 69. 74. 74.
76. 60. 77. 67.
58. 68. 77. 80. 84. 68. 76. 84. 60.]
df_stud.insert(1, "m score" , b, True)
df stud
```

	math score	m score	reading score	writing score	placememt score
0	60	60.0	63	76	95
1	75	75.0	70	64	85
2	74	74.0	50	55	91
3	68	68.0	76	78	97
4	94	84.0	67	71	93
6	61	61.0	78	92	94
7	61	61.0	74	78	80
8	64	64.0	76	79	76
9	65	65.0	95	75	90
10	66	66.0	76	67	100
11	84	84.0	67	71	92
12	69	69.0	66	70	56
13	74	74.0	65	65	80
14	74	74.0	63	72	96
15	76	76.0	64	80	96
16	60	60.0	64	54	91
17	77	77.0	70	72	99
18	67	67.0	95	64	87
20	58	58.0	65	96	92
21	68	68.0	63	62	94
22	77	77.0	63	68	97
23	80	80.0	64	86	85
24	84	84.0	63	67	83
25	68	68.0	67	73	88
26	76	76.0	64	68	96

club join year placement offer count gender 0
0 2021 3 female 1 2020 3 male 2 2020 3 male 3 2020 3 female 4 2020 3 male 6 2021 3 male 7 2021 2 male 8 2019 2 male 9 2020 3 female 10 2019 1 male 11 2020 3 male 12 2021 3 male 14 2018 3 male 15 2020 3 male 16 2021 3 female 17 2020 3 male 18 2018 3 male 16 2021 3 female 17 2020 3 male 18 2018 3 female 20 2019 3 female 21 2021 3 female 22 2021 3 female 24 2018 3 female 25 2019 3 female 26 2021 3 female 27 2018 3 female 28 2020 3 female 29 2019 3 female 20 2019 3 female 21 2021 3 female 22 2021 3 female 23 2018 3 female 24 2018 1 male 25 2019 3 female 26 2021 3 female 27 2018 2 male 28 2020 3 male
0 2021 3 female 1 2020 3 male 2 2020 3 male 3 2020 3 female 4 2020 3 male 6 2021 3 male 7 2021 2 male 8 2019 2 male 9 2020 3 female 10 2019 1 male 11 2020 3 male 12 2021 3 male 14 2018 3 male 15 2020 3 male 16 2021 3 female 17 2020 3 male 18 2018 3 male 16 2021 3 female 17 2020 3 male 18 2018 3 female 20 2019 3 female 21 2021 3 female 22 2021 3 female 24 2018 3 female 25 2019 3 female 26 2021 3 female 27 2018 3 female 28 2020 3 female 29 2019 3 female 20 2019 3 female 21 2021 3 female 22 2021 3 female 23 2018 3 female 24 2018 1 male 25 2019 3 female 26 2021 3 female 27 2018 2 male 28 2020 3 male
plt.show()



```
median = py.median(sorted rscore)
median
66.0
refined df = df1
refined_df['reading score'] = py.where(refined_df['reading
score']>upr bound, median, refined df['reading score'])
refined_df
    math score m score reading score writing score placememt score
/
0
            60
                    60.0
                                   63.0
                                                     76
                                                                       95
1
            75
                    75.0
                                   70.0
                                                     64
                                                                       85
2
                    74.0
            74
                                   50.0
                                                     55
                                                                       91
3
            68
                    68.0
                                   76.0
                                                     78
                                                                       97
            94
                    84.0
                                   67.0
                                                     71
                                                                       93
                                    78.0
                                                     92
                                                                       94
            61
                    61.0
                                    74.0
7
            61
                    61.0
                                                     78
                                                                       80
```

8	64	64.0	76.0		79	76
9	65	65.0	66.0		75	90
10	66	66.0	76.0		67	100
11	84	84.0	67.0		71	92
12	69	69.0	66.0		70	56
13	74	74.0	65.0		65	80
14	74	74.0	63.0		72	96
15	76	76.0	64.0		80	96
16	60	60.0	64.0		54	91
17	77	77.0	70.0		72	99
18	67	67.0	66.0		64	87
20	58	58.0	65.0		96	92
21	68	68.0	63.0		62	94
22	77	77.0	63.0		68	97
23	80	80.0	64.0		86	85
24	84	84.0	63.0		67	83
25	68	68.0	67.0		73	88
26	76	76.0	64.0		68	96
27	92	84.0	66.0		61	83
28	60	60.0	68.0		59	93
0 1 2 3 4 6 7 8	club join year 2021 2020 2020 2020 2020 2021 2021 2019 2020	placement	offer count	gender female male female male male male female		

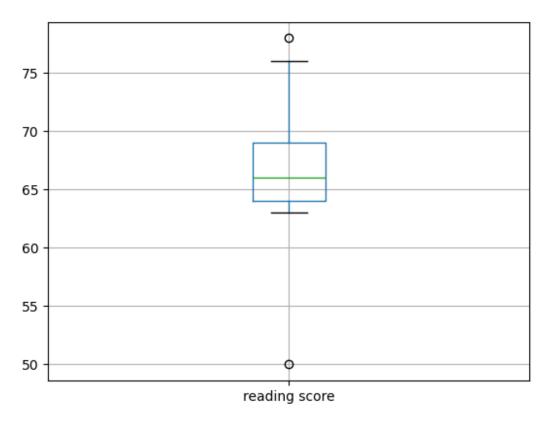
10	2019	1	male
11	2020	3	male
12	2021	3	female
13	2021	2	male
14	2018	3	male
15	2020	3	male
16	2021	3	female
17	2020	3	male
18	2018	3	female
20	2019	3	female
21	2021	3	male
22	2021	3	female
23	2018	3	female
24	2018	1	male
25	2019	3	female
26	2021	3	female
27	2018	2	male
28	2020	3	male

refined_df['reading score'] = py.where(refined_df['reading
score'] < lwr_bound, median, refined_df['reading score'])
refined_df</pre>

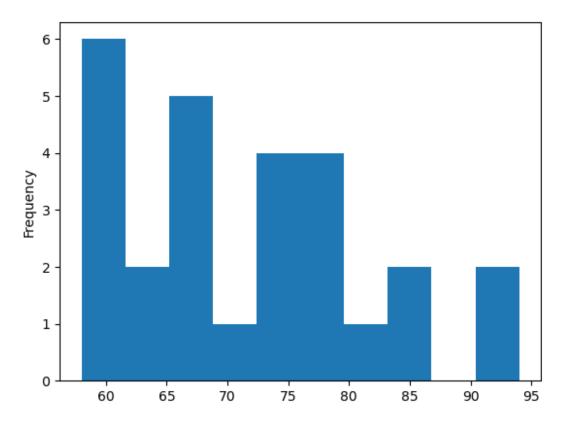
	math score	m score	reading score	writing score	placememt score
0	60	60.0	63.0	76	95
1	75	75.0	70.0	64	85
2	74	74.0	50.0	55	91
3	68	68.0	76.0	78	97
4	94	84.0	67.0	71	93
6	61	61.0	78.0	92	94
7	61	61.0	74.0	78	80
8	64	64.0	76.0	79	76
9	65	65.0	66.0	75	90
10	66	66.0	76.0	67	100
11	84	84.0	67.0	71	92
12	69	69.0	66.0	70	56
13	74	74.0	65.0	65	80

14	74	74.0	63.0		72	96
15	76	76.0	64.0		80	96
16	60	60.0	64.0		54	91
17	77	77.0	70.0		72	99
18	67	67.0	66.0		64	87
20	58	58.0	65.0		96	92
21	68	68.0	63.0		62	94
22	77	77.0	63.0		68	97
23	80	80.0	64.0		86	85
24	84	84.0	63.0		67	83
25	68	68.0	67.0		73	88
26	76	76.0	64.0		68	96
27	92	84.0	66.0		61	83
28	60	60.0	68.0		59	93
0 1 2 3 4 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22	club join year 2021 2020 2020 2020 2021 2021 2021 2019 2020 2021 2021	placement	offer count 3 3 3 3 3 3 2 2 2 3 1 3 3 3 3 3 3 3 3 3	gender female male male male male male male female male female male female female female female female female		

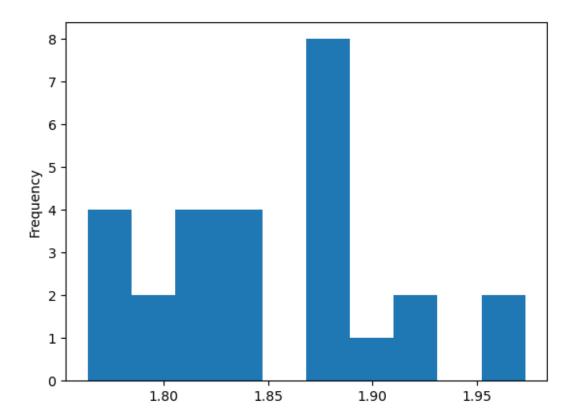
```
23
                 2018
                                                     female
                                                 3
24
                 2018
                                                 1
                                                       male
                 2019
                                                 3
                                                     female
25
                                                 3
26
                 2021
                                                     female
                                                 2
                 2018
                                                       male
27
28
                 2020
                                                       male
col2 = ['reading score']
refined_df.boxplot(col2)
<Axes: >
plt.show()
```



```
new_df['math score'].plot(kind = 'hist')
<Axes: ylabel='Frequency'>
plt.show()
```



```
df1['log_math'] = py.log10(df1['math score'])
df1['log_math'].plot(kind = 'hist')
<Axes: ylabel='Frequency'>
plt.show()
```



Name: Shreyash Tanaji Patil Rollno: 13266 B3