

Data wrangling, II

- create an "Academic performance" dataset of students & perform following operations.
 - 1) scan all variables for missing values & inconsistencies if there are missing values and inconsistencies.
 - 2) scan all numeric values for outliers if there are outliers, use any of suitable techniques.
 - 3) Apply data transformⁿ on at least one of the variables. The purpose of this transformⁿ should be one of the following reasons:
to change the scale for better understanding of the variables, to convert a non-linear relath into a linear one or to decrease skewness & convert the distributn into a normal distributn.

Reason & document your approach properly.
- Objectives of assignment:-
Students should be able to perform the data wrangling operatⁿ using python on any open source dataset.
- Prerequisite:-
 - Basic of python programming.
 - concept of data processing, data formatting, data normalization & data cleaning.

= contents for theory:-

- 1) creatⁿ of dataset using microsoft excel.
- 2) Identificatⁿ & handling of null values.
- 3) Identificatⁿ & handling of outliers.

4) Data transformtn for purpose of:

- (a) To change the scale for better understanding.
- (b) To decrease the skewness & convert distribution into normal distribution.

= Theory :

1) creatn of dataset using microsoft excel.

dataset is created in "csv" format.

- name of dataset is student performance.
- The features of dataset are : math-score, Reading-score, writing-score, placement-score, club-join-date.

- no. of instances: 30

- Response variable is: placement_offer_count.

- Range of values:

math-score [60-80], Reading-score [75-95], writing-score [60, 80], placement-score [75-100], club-join-date [2018-2017].

- The response variable is the no. of placement offers facilitated to particular students.

= conclusion:- In this way we have explored the functions of the python library for data identifying & handling the outliers. Data transformtn techniques are explored with the purpose of creating new variables & reducing skewness from datasets.

+ ⌛ ✎ ⟲ ⟳ ⟴ Run All

Code ▾

Draft Session

```
import numpy as np  
import pandas as pd
```

```
df=pd.read_csv("../input/datascience-lab2/studentPerformance.csv")
```

[3]:
df

```
[3]:
```

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
3	NaN	67	76	96	2020	3	male
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male

Console

34°C



+ - X ☐ ↗ ↘ Run All

Code ▾

Draft

5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male
14	70.0	78	62	91	2020	3	male
15	70.0	86	72	91	2019	3	male
16	77.0	86	76	71	2019	3	female
17	71.0	65	78	92	2019	3	female
18	65.0	74	80	92	2020	3	female
19	78.0	72	69	96	2020	3	female
20	70.0	100	69	70	2019	3	female
21	79.0	62	63	71	2019	3	female
22	76.0	79	73	84	2020	3	female
23	72.0	86	71	84	2020	3	female
24	70.0	83	78	91	2019	3	female

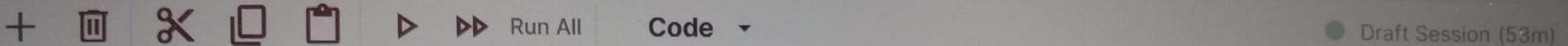


Console



notebook098aae0116 Draft saved

File Edit View Run Add-ons Help



23	72.0	86	71	84	2020	3	female
24	70.0	83	78	91	2019	3	female
25	66.0	64	64	94	2019	3	female
26	60.0	64	75	87	2020	3	female
27	50.0	63	71	71	2019	3	male
28	77.0	76	69	91	2019	3	male
29	Nan	61	60	92	2019	3	female

[4]: df.isnull()

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	True	False	False	False	False	False	False
4	False	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

Console



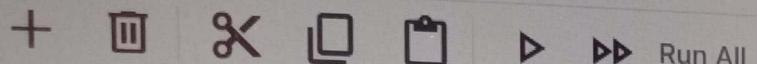
Console

34°C

Code ▾

6	False							
7	False							
8	False							
9	False							
10	False							
11	False							
12	False							
13	False							
14	False							
15	False							
16	False							
17	False							
18	False							
19	False							
20	False							
21	False							
22	False							
23	False							
24	False							
25	False							





Code ▾

Draft Session (53m)

```
29      True      False      False      False      False      False      False      False
```

```
[6]:  
series=pd.isnull(df["math score"])
```

```
[7]:  
df[series]
```

```
[7]:  
math score  reading score  writing score  placement score  club join year  placement offer count  Gender  
3           NaN          67            76              96                2020                  3    male  
29          NaN          61            60              92                2019                  3    female
```

```
[8]:  
df.notnull()
```

```
[8]:  
math score  reading score  writing score  placement score  club join year  placement offer count  Gender  
0           True          True          True          True          True          True          True          True  
1           True          True          True          True          True          True          True          True  
2           True          True          True          True          True          True          True          True
```

Console





Run All

Code ▾

4	True							
5	True							
6	True							
7	True							
8	True							
9	True							
10	True							
11	True							
12	True							
13	True							
14	True							
15	True							
16	True							
17	True							
18	True							
19	True							
20	True							
21	True							
22	True							
23	True							
24	True							



Console



+ ⌛ ✎ ⟲ ⟳ Run All Code

Draft Sess

[9]:

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

[10]:

```
df[series1]
```

[10]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male

Console

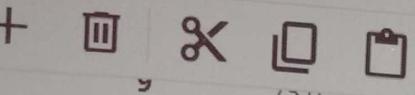
34°C
Haze



notebook098aae0116

Draft saved

File Edit View Run Add-ons Help



Run All

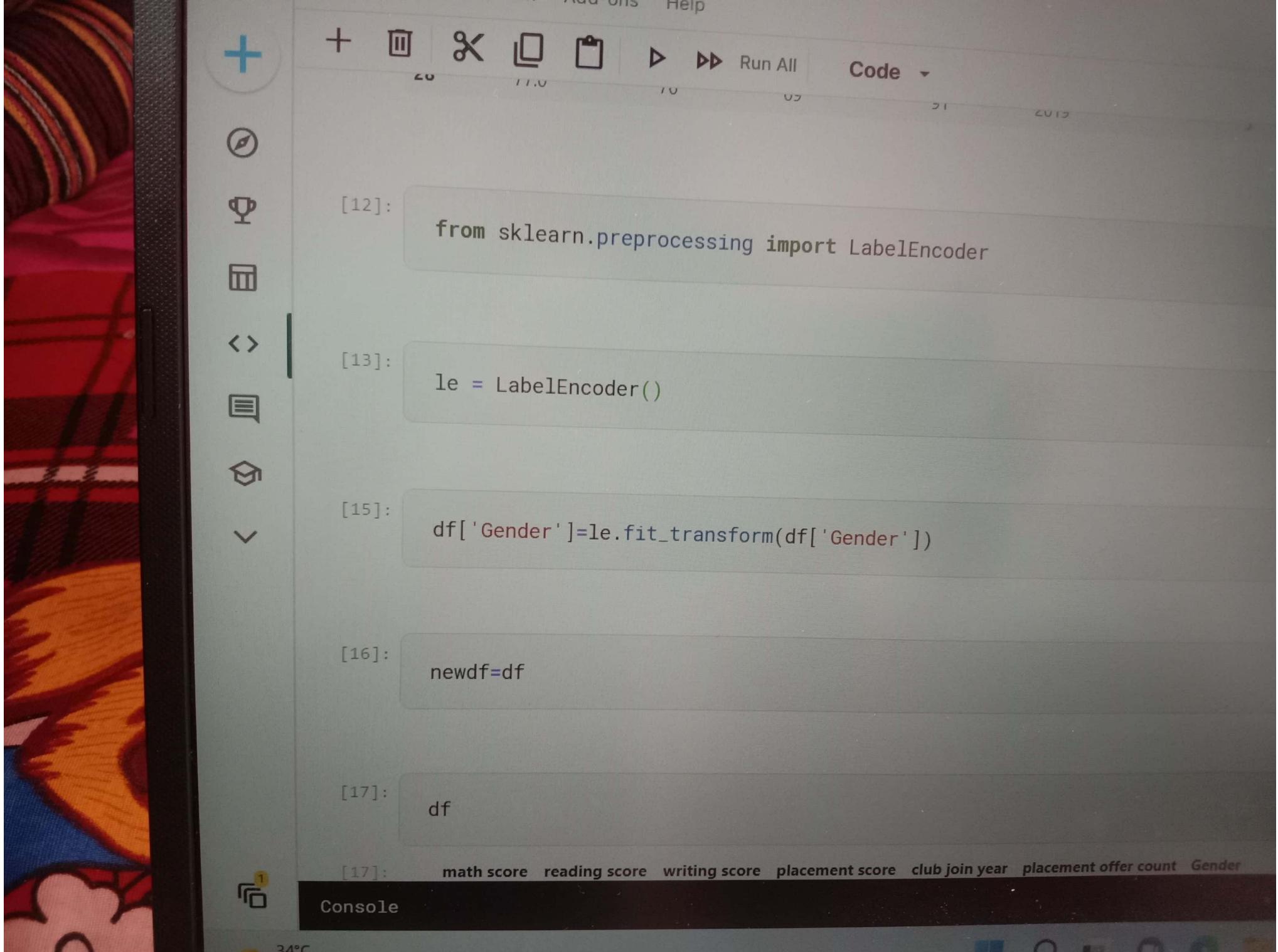
Code



	10	77.0	77	70	91	2019	3	Male
	11	63.0	71	79	100	2020	3	male
	12	100.0	80	60	100	2017	3	male
	13	66.0	64	69	91	2019	1	male
	14	70.0	78	62	76	2020	3	male
	15	70.0	86	72	91	2020	3	male
	16	77.0	86	76	91	2019	3	male
	17	71.0	65	78	71	2019	3	female
	18	65.0	74	80	92	2020	3	female
	19	78.0	72	69	96	2020	3	female
	20	70.0	100	69	70	2019	3	female
	21	79.0	62	63	71	2019	3	female
	22	76.0	79	73	84	2020	3	female
	23	72.0	86	71	84	2020	3	female
	24	70.0	83	78	91	2019	3	female
	25	66.0	64	64	94	2019	3	female
	26	60.0	64	75	87	2020	3	female
	27	50.0	63	71	71	2019	3	male
	28	77.0	76	69	91	2019	3	male



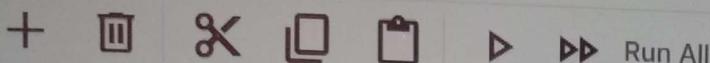
Console



notebook098aae0116

Draft saved

File Edit View Run Add-ons Help



Code

	math score	reading score	writing score	placement score	grad year	placement offer count	selected
0	77.0	82	70	85	2020	3	0
1	78.0	79	77	85	2019	3	1
2	94.0	88	65	97	2019	3	0
3	Nan	67	76	96	2020	3	1
4	99.0	62	78	93	2020	1	0
5	73.0	76	64	70	2020	3	2
6	63.0	84	76	76	2020	3	1
7	80.0	80	67	95	2019	3	1
8	66.0	60	88	96	2020	3	1
9	75.0	89	66	91	2019	3	1
10	77.0	77	70	100	2020	3	1
11	63.0	71	79	100	2017	3	1
12	100.0	80	60	91	2019	1	1
13	66.0	64	69	76	2020	3	1
14	70.0	78	62	91	2020	3	1
15	70.0	86	72	91	2019	3	1
16	77.0	86	76	71	2019	3	0
17	71.0	65	78	92	2019	3	0
18	65.0	74	80	92	2020	3	0
19	78.0	72	69	96	2020	3	0



Run All

Code ▾

	22	76.0	79	73	84	2020	3	0
	23	72.0	86	71	84	2020	3	0
	24	70.0	83	78	91	2019	3	0
	25	66.0	64	64	94	2019	3	0
	26	60.0	64	75	87	2020	3	0
	27	50.0	63	71	71	2019	3	1
	28	77.0	76	69	91	2019	3	1
	29	NaN	61	60	92	2019	3	0

[18]:

```
missing_values = [ "Na", "na" ]
```

[19]:

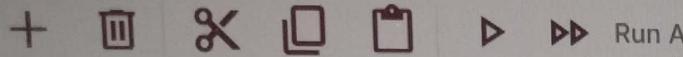
```
df = pd.read_csv("../input/datascience-lab2/studentPerformance.csv", na_values=missing_values)
```

[20]:

```
df
```

Console





Code ▾

Draft Session (55m)

[20]:	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
3	NaN	67	76	96	2020	3	male
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male
14	70.0	78	62	91	2020	3	male
15	70.0	86	72	91	2019	3	female
16	77.0	86	76	71	2019	3	female
17	71.0	65	78	92	2019	3	female
18	65.0	74	80	92	2020	3	female

Console

34°C
Haze

ASUS

+ ┌ └ X ┌ └ Run All

Code ▾

Draft Session (5)

22	76.0	79	73	84	2020	3	female
23	72.0	86	71	84	2020	3	female
24	70.0	83	78	91	2019	3	female
25	66.0	64	64	94	2019	3	female
26	60.0	64	75	87	2020	3	female
27	50.0	63	71	71	2019	3	male
28	77.0	76	69	91	2019	3	male
29	NaN	61	60	92	2019	3	female

[21]:

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

[21]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
3	0.0	67	76	96	2020	1	female
4	99.0	62	78	93	2020	3	male
5	73.0	76	64	70	2020		



Console

notebookOSRbaseOTM

File Edit View Insert Tools Database Help

+   X    >     

--



Run All

Code ▾

Draft

7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male
14	70.0	78	62	91	2020	3	male
15	70.0	86	72	91	2019	3	male
16	77.0	86	76	71	2019	3	female
17	71.0	65	78	92	2019	3	female
18	65.0	74	80	92	2020	3	female
19	78.0	72	69	96	2020	3	female
20	70.0	100	69	70	2019	3	female
21	79.0	62	63	71	2019	3	female
22	76.0	79	73	84	2020	3	female
23	72.0	86	71	84	2020	3	female
24	70.0	83	78	91	2019	3	female
25	66.0	64	64	94	2019	3	female
26	60.0	64	75	87	2020	3	female

```
29      0.0      61      60      92      2019  
3       male  
3       female
```



```
df['math score']=df['math score'].fillna(df['math score'].mean())
```

[24]:

```
df['math score']=df['math score'].fillna(df['math score'].median())
```

[25]:

```
df['math score']=df['math score'].fillna(df['math score'].std())
```

[26]:

```
df['math score']=df['math score'].fillna(df['math score'].min())
```



Console

34°C
Haze



ASUS

[27]:

```
df['math score']=df['math score'].fillna(df['math score'].max())
```

[28]:

```
m_v=df['math score'].mean()
```

[29]:

```
df['math score'].fillna(value=m_v,inplace=True)
```

[30]:

```
df
```

[30]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.000000	82	70	85	2020	3	female
1	78.000000	79	77	85	2019	3	male
2	94.000000	88	65	97	2019	3	female
3	73.642857	67	76	96	2020	3	male

Console



Run All

Code ▾

4	99.000000	62	78	93	2020	1	female
5	73.000000	76	64	70	2020	3	male
6	63.000000	84	76	76	2020	3	male
7	80.000000	80	67	95	2019	3	male
8	66.000000	60	88	96	2020	3	male
9	75.000000	89	66	91	2019	3	male
10	77.000000	77	70	100	2020	3	male
11	63.000000	71	79	100	2017	3	male
12	100.000000	80	60	91	2019	1	male
13	66.000000	64	69	76	2020	3	male
14	70.000000	78	62	91	2020	3	male
15	70.000000	86	72	91	2019	3	male
16	77.000000	86	76	71	2019	3	female
17	71.000000	65	78	92	2019	3	female
18	65.000000	74	80	92	2020	3	female
19	78.000000	72	69	96	2020	3	female
20	70.000000	100	69	70	2019	3	female
21	79.000000	62	63	71	2019	3	female
22	76.000000	79	73	84	2020	3	female
23	72.000000	86	71	84	2020	3	female

+ Run All Code

29 73.642857

61

60

92

2019

3 female

Draft Set



[33]:

```
import numpy as np  
import pandas as pd
```

[34]:

```
df=pd.read_csv("../input/datascience-lab2/studentPerformance.csv")
```

[36]:

```
df
```

[36]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020		3 female
1	78.0	79	77	85	2019		3 male
2	94.0	88	65	97	2019		3 female
3	NaN	67	76	96	2020		3 male
4	99.0	62	78	93	2020		1 female
5	73.0	76	64	70	2020		3 male



Console



34°C
Haze



notebook098aae0116 Draft saved

File Edit View Run Add-ons Help

					Code			
	9	75.0	89	66	91	2019	3	male
	10	77.0	77	70	100	2020	3	male
	11	63.0	71	79	100	2017	3	male
	12	100.0	80	60	91	2019	1	male
	13	66.0	64	69	76	2020	3	male
	14	70.0	78	62	91	2020	3	male
	15	70.0	86	72	91	2019	3	male
	16	77.0	86	76	71	2019	3	female
	17	71.0	65	78	92	2019	3	female
	18	65.0	74	80	92	2020	3	female
	19	78.0	72	69	96	2020	3	female
	20	70.0	100	69	70	2019	3	female
	21	79.0	62	63	71	2019	3	female
	22	76.0	79	73	84	2020	3	female
	23	72.0	86	71	84	2020	3	female
	24	70.0	83	78	91	2019	3	female
	25	66.0	64	64	94	2019	3	female
	26	60.0	64	75	87	2020	3	female
	27	50.0	63	71	71	2019	3	male
	28	77.0	76	69	91	2019	3	male



+ - X Copy Run All

Code ▾

Draft

29 NaN 61 60 92 2019 3 female



[38]: df.dropna()

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male
14	70.0	78	62	91	2020	3	male



Console

Run All

Code

28 77.0 76 69 91 2019 3 male

[39]: df.dropna(how='all')

[39]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
3	NaN	67	76	96	2020	3	male
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male
13	66.0	64	69	76	2020	3	male

Console

34°C Haze

Scanned with CamScanner

	28	77.0	76	69	91	2019	3	male
29	NaN		61	60	92	2019		3 female

[40]:

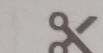
```
df.dropna(axis=1)
```

[40]:

	reading score	writing score	placement score	club join year	placement offer count	Gender
0	82	70	85	2020		3 female
1	79	77	85	2019		3 male
2	88	65	97	2019		3 female
3	67	76	96	2020		3 male
4	62	78	93	2020		1 female
5	76	64	70	2020		3 male
6	84	76	76	2020		3 male
7	80	67	95	2019		3 male
8	60	88	96	2020		3 male
9	89	66	91	2019		3 male
10	77	70	100	2020		3 male
11	71	79	100	2017		3 male
12	80	60	91	2019	1	male

Console

34°C



Run All

Code ▾

25	64	64	94	2019	3	female
26	64	75	87	2020	3	female
27	63	71	71	2019	3	male
28	76	69	91	2019	3	male
29	61	60	92	2019	3	female



[41]:

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

[42]:

```
new_data
```

[42]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male



Console

notebook098aae0116

Draft saved

File Edit View Run Add-ons Help



Run All

Code ▾

18	65.0	74	80	92	2020		
19	78.0	72	69	96	2020	3	female
20	70.0	100	69	70	2019	3	female
21	79.0	62	63	71	2019	3	female
22	76.0	79	73	84	2020	3	female
23	72.0	86	71	84	2020	3	female
24	70.0	83	78	91	2019	3	female
25	66.0	64	64	94	2019	3	female
26	60.0	64	75	87	2020	3	female
27	50.0	63	71	71	2019	3	male
28	77.0	76	69	91	2019	3	male

[43]:

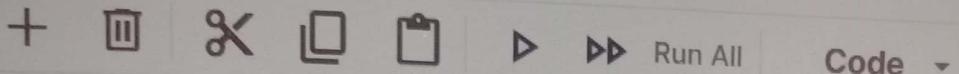
```
import pandas as pd
import numpy as np
```

[45]:

```
df=pd.read_csv("../input/datascience-lab2/studentPerformance.csv")
```

Console

34°C
Haze



[45]:

```
df=pd.read_csv("../input/datascience-lab2/studentPerformance.csv")
```

Draft S

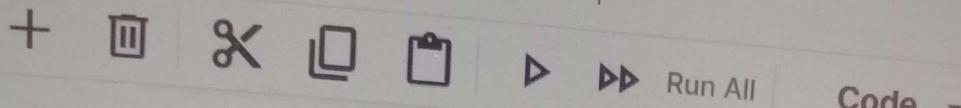
[46]:

```
df
```

[46]:

	math score	reading score	writing score	placement score	club join year	placement offer count	Gender
0	77.0	82	70	85	2020	3	female
1	78.0	79	77	85	2019	3	male
2	94.0	88	65	97	2019	3	female
3	NaN	67	76	96	2020	3	male
4	99.0	62	78	93	2020	1	female
5	73.0	76	64	70	2020	3	male
6	63.0	84	76	76	2020	3	male
7	80.0	80	67	95	2019	3	male
8	66.0	60	88	96	2020	3	male
9	75.0	89	66	91	2019	3	male
10	77.0	77	70	100	2020	3	male
11	63.0	71	79	100	2017	3	male
12	100.0	80	60	91	2019	1	male

Console



Code ▾

20	70.0	100	69	70	2019	3 female
21	79.0	62	63	71	2019	3 female
22	76.0	79	73	84	2020	3 female
23	72.0	86	71	84	2020	3 female
24	70.0	83	78	91	2019	3 female
25	66.0	64	64	94	2019	3 female
26	60.0	64	75	87	2020	3 female
27	50.0	63	71	71	2019	3 female
28	77.0	76	69	91	2019	3 male
29	NaN	61	60	92	2019	3 female

[47]:

```
col=['math score', 'reading score', 'writing score', 'placement score']
```

[48]:

```
df.boxplot(col)
```

[48]: <AxesSubplot:>



Console

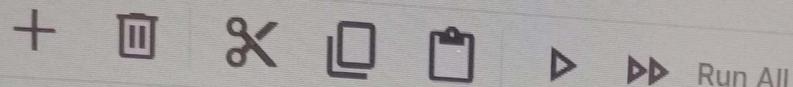
34°C



notebook098aae0116

Draft saved

File Edit View Run Add-ons Help



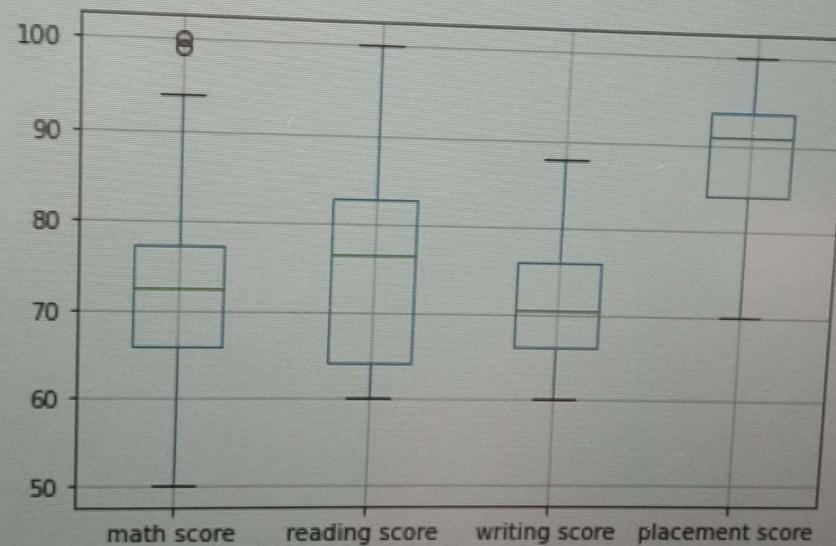
Code ▾



[48]:

```
df.boxplot(col)
```

[48]: <AxesSubplot:>



[49]:

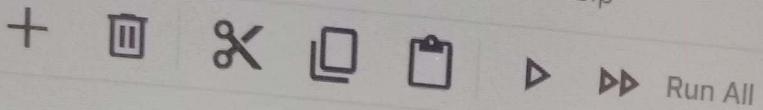
```
print(np.where(df['math score']>90))
print(np.where(df['math score']<25))
print(np.where(df['math score']<30))
```

```
(array([ 2,  4, 12]),)
```

notebook098aae0116

Draft saved

File Edit View Run Add-ons Help



Code ▾

[49]:

```
print(np.where(df['math score']>90))
print(np.where(df['math score']<25))
print(np.where(df['math score']<30))

(array([ 2,  4, 12]),)
(array([], dtype=int64),)
(array([], dtype=int64),)
```



[50]:

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

[52]:

```
df=pd.read_csv('../input/datascience-lab2/studentPerformance.csv')
```

[64]:

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



Console

notebook098aae0116 Draft saved

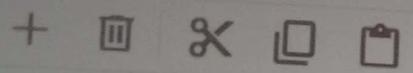
File Edit View Run Add-ons Help

+     ▶ ▶▶ Run All Code

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

notebook098aae0116 Draft saved

File Edit View Run Add-ons Help



Code ▾

Draft Session (27m)

[5]:

```
print(np.where((df['placement score']<50) & (df['placement offer count']>1)))  
print(np.where((df['placement score']>85) & (df['placement offer count']<3)))  
(array([], dtype=int64),)  
(array([ 4, 12]),)
```

[6]:

```
import numpy as np  
from scipy import stats
```

[20]:

```
df=pd.read_csv('../input/datascience-lab2/studentPerformance.csv')
```

[21]:

```
z = np.abs(stats.zscore(df['math score']))  
print(z)
```

```
0    NaN  
1    NaN  
2    NaN  
3    NaN  
4    NaN  
5    NaN
```

Console

File Edit View Run Add-ons Help

+ ⌫ ✎ ⟲ ⟳ ⟴ Run All

Code

Draft Session (27m)

```
8     NaN
9     NaN
10    NaN
11    NaN
12    NaN
13    NaN
14    NaN
15    NaN
16    NaN
17    NaN
18    NaN
19    NaN
20    NaN
21    NaN
22    NaN
23    NaN
24    NaN
25    NaN
26    NaN
27    NaN
28    NaN
29    NaN
Name: math score, dtype: float64
```

[16]:

```
threshold=0
sample_outliers=np.where(z<threshold)
sample_outliers
```

[16]: (array([], dtype=int64),)



Console

[23]:

```
import numpy as np
```

[24]:

```
sorted_rscore=sorted(df['reading score'])  
sorted_rscore
```

```
[24]: [60,  
       61,  
       62,  
       62,  
       63,  
       64,  
       64,  
       64,  
       65,  
       67,  
       71,  
       72,  
       74,  
       76,  
       76,  
       77,  
       78,  
       79,  
       79,  
       80,  
       80,  
       82,
```

Console

```
86,  
88,  
89,  
100]
```

Code ▾

Draft Session

[25]:

```
q1=np.percentile(sorted_rscores,25)  
q3=np.percentile(sorted_rscores,75)  
print(q1,q3)
```

64.25 82.75

[26]:

```
IQR=q3-q1
```

[28]:

```
lwr_bound=q1-(1.5*IQR)  
upr_bound=q3+(1.5*IQR)  
print(lwr_bound,upr_bound)
```

36.5 110.5

[29]:

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

```
[30]:  
new_df=df
```

```
[31]:  
for i in sample_outliers:  
    new_df.drop(i,inplace=True)  
new_df
```

```
[31]:  
   math score  reading score  writing score  placement score  club join year  placement offer count  Gender  
0      77.0          82            70             85           2020                      3     female  
1      78.0          79            77             85           2019                      3     male  
2      94.0          88            65             97           2019                      3     female  
3       NaN          67            76             96           2020                      3     male  
4      99.0          62            78             93           2020                      1     female
```



Run All

Code

29

NaN

61

60

92

2019

3 female

Draft Session

[33]:

```
df=pd.read_csv("../input/datascience-lab2/studentPerformance.csv")
```



[34]:

```
df_stud=df
```



[36]:

```
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)
```

New array: [77. 78. 94. nan 99. 73. 63. 80. 66. 75. 77. 63. 100. 66.
 70. 70. 77. 71. 65. 78. 70. 79. 76. 72. 70. 66. 60. 50.
 77. nan]

[38]:

```
df_stud.insert(1,"m score",b, True)
df_stud
```