```
#Priya More(305C002)
```

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

```
from google.colab import files
uploads=files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving student performance.csv to student performance.csv

```
df=pd.read_csv("student performance.csv")
```

```
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	NaN	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	NaN	NaN	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	NaN	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	NaN	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	NaN	NaN	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	NaN	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df.shape
```

```
(13, 8)
```

```
df.isnull()
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
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
4	False	False	False	False	False	True	False	False
5	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False
7	False	False	True	True	False	False	False	False
8	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False
11	False	False	False	False	False	True	False	False
12	False	False	False	False	False	False	False	False

```
df.isnull().sum()
```

```
Unnamed: 0         0
gender             0
math score         1
reading score      1
writing score      0
placement score    2
club join year     0
placement offer    0
dtype: int64
```

```
df["math score"]=df["math score"].interpolate()
```

```
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	NaN	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	70.5	NaN	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	NaN	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df["placement score"]=df["placement score"].fillna(method="bfill")
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	100.0	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	70.5	NaN	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	98.0	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df["placement score"]=df["placement score"].fillna(method="bfill")
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	100.0	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	70.5	NaN	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	98.0	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df.isnull().sum()
```

```
Unnamed: 0         0
gender             0
math score         0
reading score      1
writing score      0
placement score    0
club join year     0
placement offer    0
dtype: int64
```

```
df["reading score"]=df["reading score"].replace(to_replace=np.nan,value='80')
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	100.0	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	70.5	80	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	98.0	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
df.isnull().sum()
```

```
Unnamed: 0         0
gender             0
math score         0
reading score      0
writing score      0
placement score    0
club join year     0
placement offer    0
dtype: int64
```

```
df["math score"]=df["math score"].fillna('1')
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63.0	84.0	64	84.0	2020	2
1	1	female	71.0	80.0	76	86.0	2018	3
2	2	female	64.0	81.0	66	81.0	2020	2
3	3	male	71.0	85.0	77	96.0	2018	1
4	4	male	68.0	86.0	76	100.0	2021	3
5	5	female	94.0	86.0	61	100.0	2019	1
6	6	male	75.0	79.0	66	-99.0	2020	1
7	7	female	70.5	80	66	95.0	2019	3
8	8	male	66.0	88.0	66	88.0	2020	3
9	9	male	70.0	79.0	61	87.0	2021	2
10	10	female	-99.0	80.0	65	85.0	2021	1
11	11	male	76.0	84.0	-99	98.0	2020	2
12	12	female	74.0	79.0	79	98.0	2019	2

```
# converting into the type
df.dtypes
```

```
Unnamed: 0           int64
gender              object
math score         float64
reading score       object
writing score        int64
placement score    float64
club join year       int64
placement offer      int64
dtype: object
```

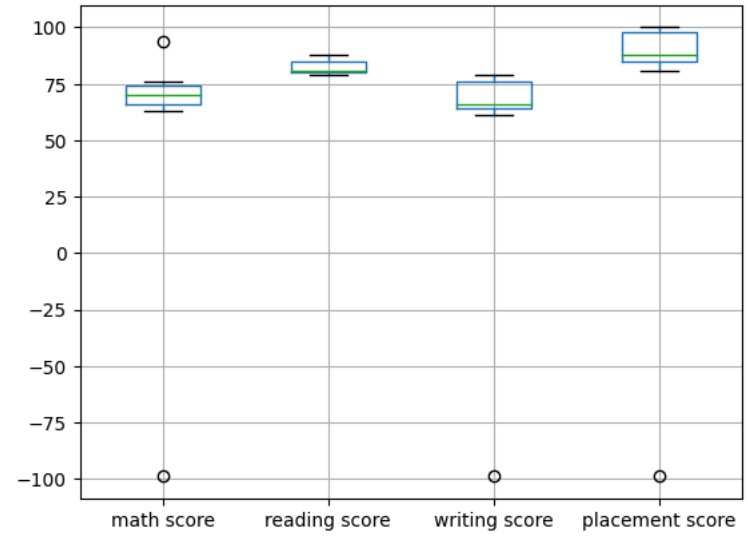
```
df['math score']=df['math score'].astype('int64')
df.dtypes
```

```
Unnamed: 0           int64
gender              object
```

```
math score           int64
reading score       object
writing score        int64
placement score    float64
club join year       int64
placement offer      int64
dtype: object
```

```
columns=['math score','reading score','writing score','placement score']
df.boxplot(columns)
```

```
<Axes: >
```




```
df["reading score"]=df["reading score"].astype('int64')
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63	84	64	84.0	2020	2
1	1	female	71	80	76	86.0	2018	3
2	2	female	64	81	66	81.0	2020	2
3	3	male	71	85	77	96.0	2018	1
4	4	male	68	86	76	100.0	2021	3
5	5	female	94	86	61	100.0	2019	1
6	6	male	75	79	66	-99.0	2020	1
7	7	female	70	80	66	95.0	2019	3
8	8	male	66	88	66	88.0	2020	3
9	9	male	70	79	61	87.0	2021	2
10	10	female	-99	80	65	85.0	2021	1
11	11	male	76	84	-99	98.0	2020	2
12	12	female	74	79	79	98.0	2019	2

```
np.where(df['math score']>65)
```

```
(array([ 1,  3,  4,  5,  6,  7,  8,  9, 11, 12]),)
```

```
np.where(df['math score']<65)
```

```
(array([ 0,  2, 10]),)
```

```
df[(df['placement score']<75)& (df['placement offer']>1)]
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer

```
df[(df['math score']>75)&(df['placement offer']>1)]
```

```
new_df1=df[((df['math score']>=60)&(df['math score']<=80)) &
           ((df['reading score']>=75)&(df['reading score']<=95))&
            ((df['writing score']>=60)&(df['writing score']<=80))&
            ((df['placement score']>=75)&(df['placement score']<=100))&
            ((df['club join year']>=2018)&(df['club join year']<=2021))]
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63	84	64	84.0	2020	2
1	1	female	71	80	76	86.0	2018	3
2	2	female	64	81	66	81.0	2020	2
3	3	male	71	85	77	96.0	2018	1
4	4	male	68	86	76	100.0	2021	3
7	7	female	70	80	66	95.0	2019	3
8	8	male	66	88	66	88.0	2020	3
9	9	male	70	79	61	87.0	2021	2
12	12	female	74	79	79	98.0	2019	2

```
new_df1.shape
```

```
(9, 8)
```

```
df
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63	84	64	84.0	2020	2
1	1	female	71	80	76	86.0	2018	3
2	2	female	64	81	66	81.0	2020	2
3	3	male	71	85	77	96.0	2018	1
4	4	male	68	86	76	100.0	2021	3
5	5	female	94	86	61	100.0	2019	1
6	6	male	75	79	66	-99.0	2020	1
7	7	female	70	80	66	95.0	2019	3
8	8	male	66	88	66	88.0	2020	3
9	9	male	70	79	61	87.0	2021	2
10	10	female	-99	80	65	85.0	2021	1
11	11	male	76	84	-99	98.0	2020	2
12	12	female	74	79	79	98.0	2019	2

```
new_df1[(new_df1['math score']>60)&(new_df1['math score']<90)]
```

	Unnamed: 0	gender	math score	reading score	writing score	placement score	club join year	placement offer
0	0	female	63	84	64	84.0	2020	2
1	1	female	71	80	76	86.0	2018	3