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

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

In [3]:
```
df
```

Out[3]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	64.0	84.0	2020.0	2.0
1	71.0	80	76.0	86.0	2018.0	3.0
2	64.0	81	66.0	81.0	2020.0	2.0
3	71.0	85	77.0	91.0	2018.0	3.0
4	68.0	86	76.0	92.0	2021.0	3.0
5	79.0	86	61.0	100.0	2019.0	3.0
6	75.0	79	66.0	76.0	2020.0	2.0
7	71.0	79	66.0	95.0	2019.0	3.0
8	66.0	88	66.0	88.0	2020.0	2.0
9	70.0	79	61.0	87.0	2021.0	2.0
10	78.0	80	65.0	85.0	2021.0	2.0
11	76.0	84	73.0	92.0	2020.0	3.0
12	74.0	79	79.0	98.0	2019.0	3.0
13	NaN	85	NaN	NaN	NaN	NaN
14	NaN	78	NaN	NaN	NaN	NaN

In [4]:
```
1 df.isnull()
```

Out[4]:

	math score	reading score	writing score	placement score	club join year	placement offer count
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	False	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	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	True	False	True	True	True	True
14	True	False	True	True	True	True

In [5]:
```
1 series=pd.isnull(df["math score"])
2 df[series]
```

Out[5]:

	math score	reading score	writing score	placement score	club join year	placement offer count
13	NaN	85	NaN	NaN	NaN	NaN
14	NaN	78	NaN	NaN	NaN	NaN

In [6]:
```
1 df.notnull()
```

Out[6]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
5	True	True	True	True	True	True
6	True	True	True	True	True	True
7	True	True	True	True	True	True
8	True	True	True	True	True	True
9	True	True	True	True	True	True
10	True	True	True	True	True	True
11	True	True	True	True	True	True
12	True	True	True	True	True	True
13	False	True	False	False	False	False
14	False	True	False	False	False	False

In [7]:
```
1 series1=pd.notnull(df["math score"])
2 df[series]
```

Out[7]:

	math score	reading score	writing score	placement score	club join year	placement offer count
13	NaN	85	NaN	NaN	NaN	NaN
14	NaN	78	NaN	NaN	NaN	NaN

In [9]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['writing score']=le.fit_transform(df['writing score'])
newdf=df
df
```

Out[9]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	1	84.0	2020.0	2.0
1	71.0	80	5	86.0	2018.0	3.0
2	64.0	81	3	81.0	2020.0	2.0
3	71.0	85	6	91.0	2018.0	3.0
4	68.0	86	5	92.0	2021.0	3.0
5	79.0	86	0	100.0	2019.0	3.0
6	75.0	79	3	76.0	2020.0	2.0
7	71.0	79	3	95.0	2019.0	3.0
8	66.0	88	3	88.0	2020.0	2.0
9	70.0	79	0	87.0	2021.0	2.0
10	78.0	80	2	85.0	2021.0	2.0
11	76.0	84	4	92.0	2020.0	3.0
12	74.0	79	7	98.0	2019.0	3.0
13	NaN	85	8	NaN	NaN	NaN
14	NaN	78	8	NaN	NaN	NaN

In [10]:
```
missing_values=["Na","na"]
df=pd.read_csv("TANMAY.csv",)
df
```

Out[10]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	64.0	84.0	2020.0	2.0
1	71.0	80	76.0	86.0	2018.0	3.0
2	64.0	81	66.0	81.0	2020.0	2.0
3	71.0	85	77.0	91.0	2018.0	3.0
4	68.0	86	76.0	92.0	2021.0	3.0
5	79.0	86	61.0	100.0	2019.0	3.0
6	75.0	79	66.0	76.0	2020.0	2.0
7	71.0	79	66.0	95.0	2019.0	3.0
8	66.0	88	66.0	88.0	2020.0	2.0
9	70.0	79	61.0	87.0	2021.0	2.0
10	78.0	80	65.0	85.0	2021.0	2.0
11	76.0	84	73.0	92.0	2020.0	3.0
12	74.0	79	79.0	98.0	2019.0	3.0
13	NaN	85	NaN	NaN	NaN	NaN
14	NaN	78	NaN	NaN	NaN	NaN

In [11]:
```
ndf=df
ndf.fillna(0)
```

Out[11]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	64.0	84.0	2020.0	2.0
1	71.0	80	76.0	86.0	2018.0	3.0
2	64.0	81	66.0	81.0	2020.0	2.0
3	71.0	85	77.0	91.0	2018.0	3.0
4	68.0	86	76.0	92.0	2021.0	3.0
5	79.0	86	61.0	100.0	2019.0	3.0
6	75.0	79	66.0	76.0	2020.0	2.0
7	71.0	79	66.0	95.0	2019.0	3.0
8	66.0	88	66.0	88.0	2020.0	2.0
9	70.0	79	61.0	87.0	2021.0	2.0
10	78.0	80	65.0	85.0	2021.0	2.0
11	76.0	84	73.0	92.0	2020.0	3.0
12	74.0	79	79.0	98.0	2019.0	3.0
13	0.0	85	0.0	0.0	0.0	0.0
14	0.0	78	0.0	0.0	0.0	0.0

In [12]:
```
m_v=df['math score'].mean()
df['math score'].fillna(value=m_v,inplace=True)
df
```

Out[12]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.000000	84	64.0	84.0	2020.0	2.0
1	71.000000	80	76.0	86.0	2018.0	3.0
2	64.000000	81	66.0	81.0	2020.0	2.0
3	71.000000	85	77.0	91.0	2018.0	3.0
4	68.000000	86	76.0	92.0	2021.0	3.0
5	79.000000	86	61.0	100.0	2019.0	3.0
6	75.000000	79	66.0	76.0	2020.0	2.0
7	71.000000	79	66.0	95.0	2019.0	3.0
8	66.000000	88	66.0	88.0	2020.0	2.0
9	70.000000	79	61.0	87.0	2021.0	2.0
10	78.000000	80	65.0	85.0	2021.0	2.0
11	76.000000	84	73.0	92.0	2020.0	3.0
12	74.000000	79	79.0	98.0	2019.0	3.0
13	71.230769	85	NaN	NaN	NaN	NaN
14	71.230769	78	NaN	NaN	NaN	NaN

In [13]:
```
ndf.replace(to_replace=np.nan,value=-99)
```

Out[13]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.000000	84	64.0	84.0	2020.0	2.0
1	71.000000	80	76.0	86.0	2018.0	3.0
2	64.000000	81	66.0	81.0	2020.0	2.0
3	71.000000	85	77.0	91.0	2018.0	3.0
4	68.000000	86	76.0	92.0	2021.0	3.0
5	79.000000	86	61.0	100.0	2019.0	3.0
6	75.000000	79	66.0	76.0	2020.0	2.0
7	71.000000	79	66.0	95.0	2019.0	3.0
8	66.000000	88	66.0	88.0	2020.0	2.0
9	70.000000	79	61.0	87.0	2021.0	2.0
10	78.000000	80	65.0	85.0	2021.0	2.0
11	76.000000	84	73.0	92.0	2020.0	3.0
12	74.000000	79	79.0	98.0	2019.0	3.0
13	71.230769	85	-99.0	-99.0	-99.0	-99.0
14	71.230769	78	-99.0	-99.0	-99.0	-99.0

In [14]:
```
1 ndf.dropna()
```

Out[14]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	64.0	84.0	2020.0	2.0
1	71.0	80	76.0	86.0	2018.0	3.0
2	64.0	81	66.0	81.0	2020.0	2.0
3	71.0	85	77.0	91.0	2018.0	3.0
4	68.0	86	76.0	92.0	2021.0	3.0
5	79.0	86	61.0	100.0	2019.0	3.0
6	75.0	79	66.0	76.0	2020.0	2.0
7	71.0	79	66.0	95.0	2019.0	3.0
8	66.0	88	66.0	88.0	2020.0	2.0
9	70.0	79	61.0	87.0	2021.0	2.0
10	78.0	80	65.0	85.0	2021.0	2.0
11	76.0	84	73.0	92.0	2020.0	3.0
12	74.0	79	79.0	98.0	2019.0	3.0

In [15]:
```
1 ndf.dropna(how='all')
```

Out[15]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.000000	84	64.0	84.0	2020.0	2.0
1	71.000000	80	76.0	86.0	2018.0	3.0
2	64.000000	81	66.0	81.0	2020.0	2.0
3	71.000000	85	77.0	91.0	2018.0	3.0
4	68.000000	86	76.0	92.0	2021.0	3.0
5	79.000000	86	61.0	100.0	2019.0	3.0
6	75.000000	79	66.0	76.0	2020.0	2.0
7	71.000000	79	66.0	95.0	2019.0	3.0
8	66.000000	88	66.0	88.0	2020.0	2.0
9	70.000000	79	61.0	87.0	2021.0	2.0
10	78.000000	80	65.0	85.0	2021.0	2.0
11	76.000000	84	73.0	92.0	2020.0	3.0
12	74.000000	79	79.0	98.0	2019.0	3.0
13	71.230769	85	NaN	NaN	NaN	NaN
14	71.230769	78	NaN	NaN	NaN	NaN

In [16]:
```
1 ndf.dropna(axis=1)
```

Out[16]:

	math score	reading score
0	63.000000	84
1	71.000000	80
2	64.000000	81
3	71.000000	85
4	68.000000	86
5	79.000000	86
6	75.000000	79
7	71.000000	79
8	66.000000	88
9	70.000000	79
10	78.000000	80
11	76.000000	84
12	74.000000	79
13	71.230769	85
14	71.230769	78

In [17]:
```
1 new_data=ndf.dropna(axis=0,how='any')
2 new_data
```

Out[17]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.0	84	64.0	84.0	2020.0	2.0
1	71.0	80	76.0	86.0	2018.0	3.0
2	64.0	81	66.0	81.0	2020.0	2.0
3	71.0	85	77.0	91.0	2018.0	3.0
4	68.0	86	76.0	92.0	2021.0	3.0
5	79.0	86	61.0	100.0	2019.0	3.0
6	75.0	79	66.0	76.0	2020.0	2.0
7	71.0	79	66.0	95.0	2019.0	3.0
8	66.0	88	66.0	88.0	2020.0	2.0
9	70.0	79	61.0	87.0	2021.0	2.0
10	78.0	80	65.0	85.0	2021.0	2.0
11	76.0	84	73.0	92.0	2020.0	3.0
12	74.0	79	79.0	98.0	2019.0	3.0

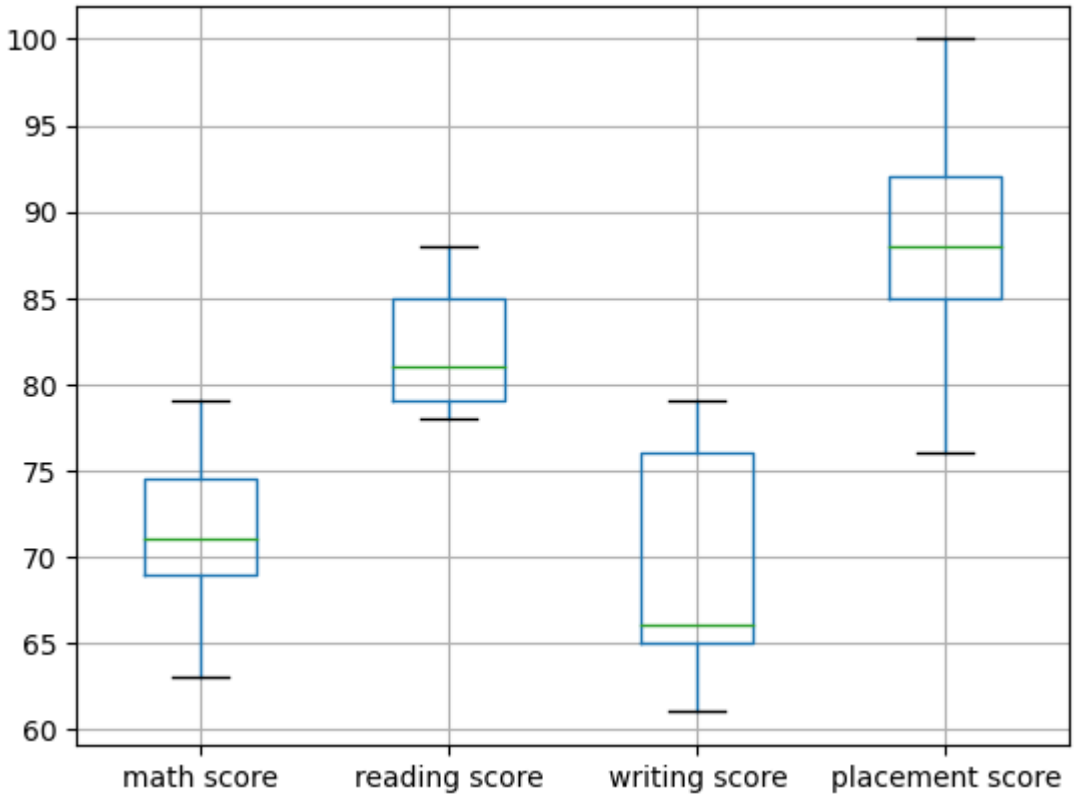
In [18]:
```
1 df
```

Out[18]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.000000	84	64.0	84.0	2020.0	2.0
1	71.000000	80	76.0	86.0	2018.0	3.0
2	64.000000	81	66.0	81.0	2020.0	2.0
3	71.000000	85	77.0	91.0	2018.0	3.0
4	68.000000	86	76.0	92.0	2021.0	3.0
5	79.000000	86	61.0	100.0	2019.0	3.0
6	75.000000	79	66.0	76.0	2020.0	2.0
7	71.000000	79	66.0	95.0	2019.0	3.0
8	66.000000	88	66.0	88.0	2020.0	2.0
9	70.000000	79	61.0	87.0	2021.0	2.0
10	78.000000	80	65.0	85.0	2021.0	2.0
11	76.000000	84	73.0	92.0	2020.0	3.0
12	74.000000	79	79.0	98.0	2019.0	3.0
13	71.230769	85	NaN	NaN	NaN	NaN
14	71.230769	78	NaN	NaN	NaN	NaN

In [19]:
```
1 col=['math score','reading score','writing score','placement score']
2 df.boxplot(col)
```

Out[19]: <Axes: >



In [21]:
```
print(np.where(df['math score']>90))
print(np.where(df['reading score']<25))
print(np.where(df['writing score']<30))
```

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

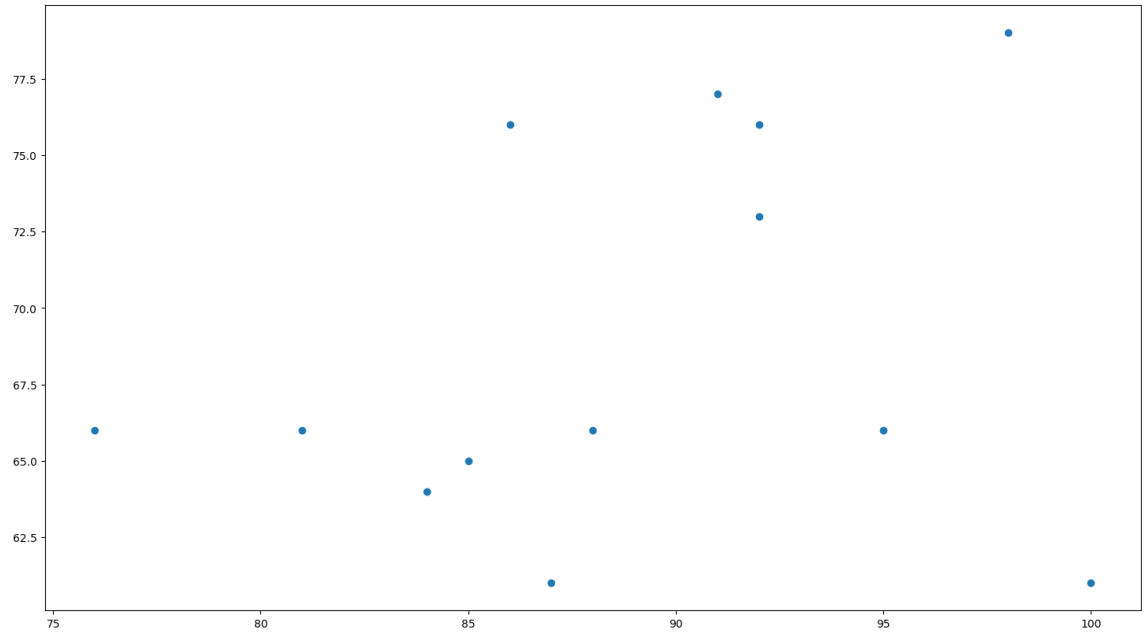
In [22]:
```
df
```

Out[22]:

	math score	reading score	writing score	placement score	club join year	placement offer count
0	63.000000	84	64.0	84.0	2020.0	2.0
1	71.000000	80	76.0	86.0	2018.0	3.0
2	64.000000	81	66.0	81.0	2020.0	2.0
3	71.000000	85	77.0	91.0	2018.0	3.0
4	68.000000	86	76.0	92.0	2021.0	3.0
5	79.000000	86	61.0	100.0	2019.0	3.0
6	75.000000	79	66.0	76.0	2020.0	2.0
7	71.000000	79	66.0	95.0	2019.0	3.0
8	66.000000	88	66.0	88.0	2020.0	2.0
9	70.000000	79	61.0	87.0	2021.0	2.0
10	78.000000	80	65.0	85.0	2021.0	2.0
11	76.000000	84	73.0	92.0	2020.0	3.0
12	74.000000	79	79.0	98.0	2019.0	3.0
13	71.230769	85	NaN	NaN	NaN	NaN
14	71.230769	78	NaN	NaN	NaN	NaN

In [23]:
```
import matplotlib.pyplot as plt
```

In [24]:
```
fig,ax=plt.subplots(figsize=(18,10))
ax.scatter(df['placement score'],df['writing score'])
plt.show()
```



In [25]:
```
print(np.where((df['placement score']<50) & (df['writing score']>1)))
print(np.where((df['placement score']>85) & (df['writing score']<3)))
```

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

In [ ]:
```

```

```
# NAME : TANMAY DIXIT
# ROLL NO : 13143
```