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

In [5]:
```
import numpy as np
```

In [11]:
```
df=pd.read_csv(r"C:\Users\Vaishnavi Jangam\Desktop\StudentsPerformance(B).csv")
```

In [12]:
```
df
```

Out[12]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.0	42.0	42	67	2019	1
1	69.0	24.0	30	74	2019	1
2	90.0	14.0	19	74	2018	1
3	99.0	24.0	72	85	2019	3
4	76.0	77.0	42	74	2020	1
5	71.0	34.0	51	74	2019	1
6	88.0	37.0	64	74	2018	1
7	40.0	NaN	72	74	2018	1
8	64.0	5.0	18	74	2018	1
9	38.0	93.0	11	74	2019	1
10	58.0	82.0	76	74	2020	1
11	40.0	65.0	53	74	2018	1
12	65.0	55.0	3	74	2018	1
13	NaN	34.0	80	74	2019	1
14	50.0	64.0	86	74	2020	1
15	69.0	20.0	59	74	2018	1
16	88.0	NaN	67	74	2020	1
17	18.0	NaN	27	74	2019	1
18	46.0	27.0	87	46	2020	1
19	54.0	77.0	73	61	2018	1

In [13]:
```
df.isnull()
```

Out[13]:

	Maths 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	True	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	False	False	False	False
14	False	False	False	False	False	False
15	False	False	False	False	False	False
16	False	True	False	False	False	False
17	False	True	False	False	False	False
18	False	False	False	False	False	False
19	False	False	False	False	False	False

In [15]:
```
series = pd.isnull(df["Maths Score"])
df[series]
```

Out[15]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
13	NaN	34.0	80	74	2019	1

In [17]:
```
df.notnull()
```

Out[17]:

	Maths 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	False	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	True	True	True	True
14	True	True	True	True	True	True
15	True	True	True	True	True	True
16	True	False	True	True	True	True
17	True	False	True	True	True	True
18	True	True	True	True	True	True
19	True	True	True	True	True	True

In [20]:

```
series1 = pd.notnull(df["Maths Score"])
df[series1]
```

Out[20]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.0	42.0	42	67	2019	1
1	69.0	24.0	30	74	2019	1
2	90.0	14.0	19	74	2018	1
3	99.0	24.0	72	85	2019	3
4	76.0	77.0	42	74	2020	1
5	71.0	34.0	51	74	2019	1
6	88.0	37.0	64	74	2018	1
7	40.0	NaN	72	74	2018	1
8	64.0	5.0	18	74	2018	1
9	38.0	93.0	11	74	2019	1
10	58.0	82.0	76	74	2020	1
11	40.0	65.0	53	74	2018	1
12	65.0	55.0	3	74	2018	1
14	50.0	64.0	86	74	2020	1
15	69.0	20.0	59	74	2018	1
16	88.0	NaN	67	74	2020	1
17	18.0	NaN	27	74	2019	1
18	46.0	27.0	87	46	2020	1
19	54.0	77.0	73	61	2018	1

In [22]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['Writing Score'] = le.fit_transform(df['Writing Score'])
newdf=df
df
```

Out[22]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.0	42.0	6	67	2019	1
1	69.0	24.0	5	74	2019	1
2	90.0	14.0	3	74	2018	1
3	99.0	24.0	12	85	2019	3
4	76.0	77.0	6	74	2020	1
5	71.0	34.0	7	74	2019	1
6	88.0	37.0	10	74	2018	1
7	40.0	NaN	12	74	2018	1
8	64.0	5.0	2	74	2018	1
9	38.0	93.0	1	74	2019	1
10	58.0	82.0	14	74	2020	1
11	40.0	65.0	8	74	2018	1
12	65.0	55.0	0	74	2018	1
13	NaN	34.0	15	74	2019	1
14	50.0	64.0	16	74	2020	1
15	69.0	20.0	9	74	2018	1
16	88.0	NaN	11	74	2020	1
17	18.0	NaN	4	74	2019	1
18	46.0	27.0	17	46	2020	1
19	54.0	77.0	13	61	2018	1

In [23]:

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

Out[23]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.0	42.0	6	67	2019	1
1	69.0	24.0	5	74	2019	1
2	90.0	14.0	3	74	2018	1
3	99.0	24.0	12	85	2019	3
4	76.0	77.0	6	74	2020	1
5	71.0	34.0	7	74	2019	1
6	88.0	37.0	10	74	2018	1
7	40.0	0.0	12	74	2018	1
8	64.0	5.0	2	74	2018	1
9	38.0	93.0	1	74	2019	1
10	58.0	82.0	14	74	2020	1
11	40.0	65.0	8	74	2018	1
12	65.0	55.0	0	74	2018	1
13	0.0	34.0	15	74	2019	1
14	50.0	64.0	16	74	2020	1
15	69.0	20.0	9	74	2018	1
16	88.0	0.0	11	74	2020	1
17	18.0	0.0	4	74	2019	1
18	46.0	27.0	17	46	2020	1
19	54.0	77.0	13	61	2018	1

In [24]:

```
m_v=df['Maths Score'].mean()
df['Maths Score'].fillna(value=m_v, inplace=True)
df
```

Out[24]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
7	40.000000	NaN	12	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
16	88.000000	NaN	11	74	2020	1
17	18.000000	NaN	4	74	2019	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [25]:
```
ndf.replace(to_replace = np.nan, value = -99)
```

Out[25]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
7	40.000000	-99.0	12	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
16	88.000000	-99.0	11	74	2020	1
17	18.000000	-99.0	4	74	2019	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [26]:
```
ndf.dropna()
```

Out[26]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [27]:

```
ndf.dropna(how = 'all')
```

Out[27]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
7	40.000000	NaN	12	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
16	88.000000	NaN	11	74	2020	1
17	18.000000	NaN	4	74	2019	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [28]:
```
ndf.dropna(axis = 1)
```

Out[28]:

	Maths Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	6	67	2019	1
1	69.000000	5	74	2019	1
2	90.000000	3	74	2018	1
3	99.000000	12	85	2019	3
4	76.000000	6	74	2020	1
5	71.000000	7	74	2019	1
6	88.000000	10	74	2018	1
7	40.000000	12	74	2018	1
8	64.000000	2	74	2018	1
9	38.000000	1	74	2019	1
10	58.000000	14	74	2020	1
11	40.000000	8	74	2018	1
12	65.000000	0	74	2018	1
13	62.894737	15	74	2019	1
14	50.000000	16	74	2020	1
15	69.000000	9	74	2018	1
16	88.000000	11	74	2020	1
17	18.000000	4	74	2019	1
18	46.000000	17	46	2020	1
19	54.000000	13	61	2018	1

In [29]:

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

Out[29]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

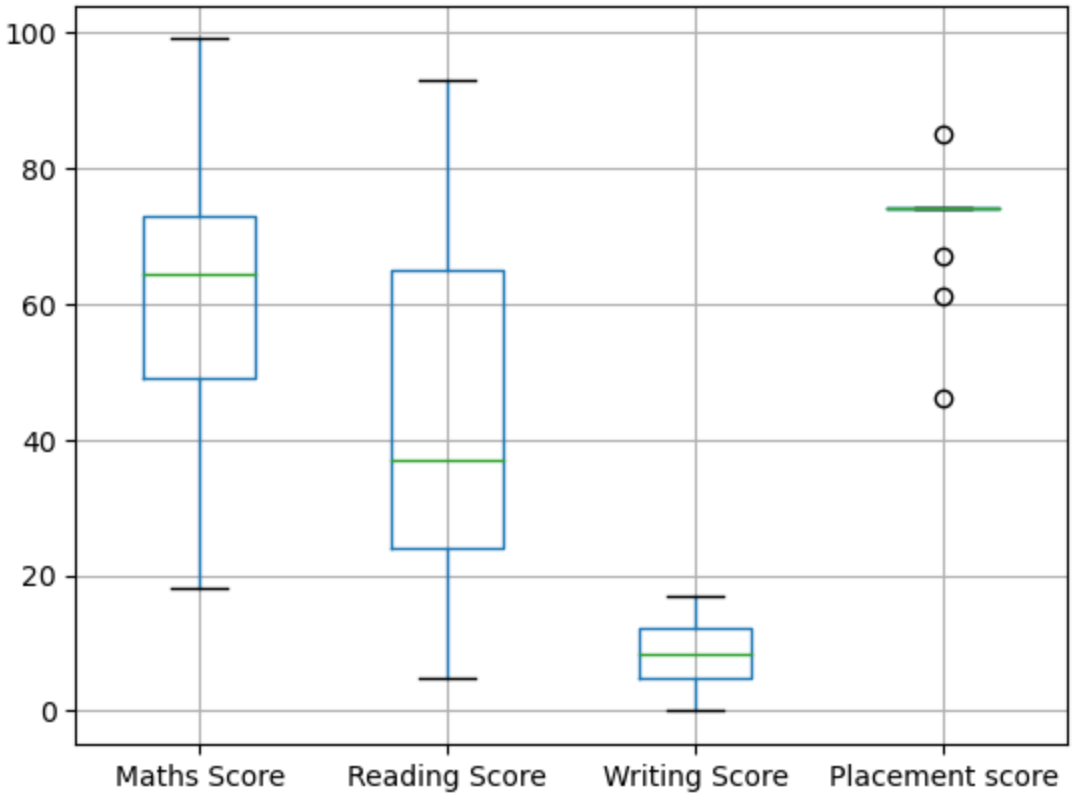
In [30]:
```
df
```

Out[30]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
7	40.000000	NaN	12	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
16	88.000000	NaN	11	74	2020	1
17	18.000000	NaN	4	74	2019	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [31]:
```
col = ['Maths Score' , 'Reading Score' , 'Writing Score' , 'Placement score']
df.boxplot(col)
```

Out[31]:
```
<Axes: >
```



In [34]:
```
print(np.where(df['Maths Score']>90))
print(np.where(df['Reading Score']<25))
print(np.where(df['Writing Score']<30))
```

```
(array([3], dtype=int64),)
(array([ 1,  2,  3,  8, 15], dtype=int64),)
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
       17, 18, 19], dtype=int64),)
```

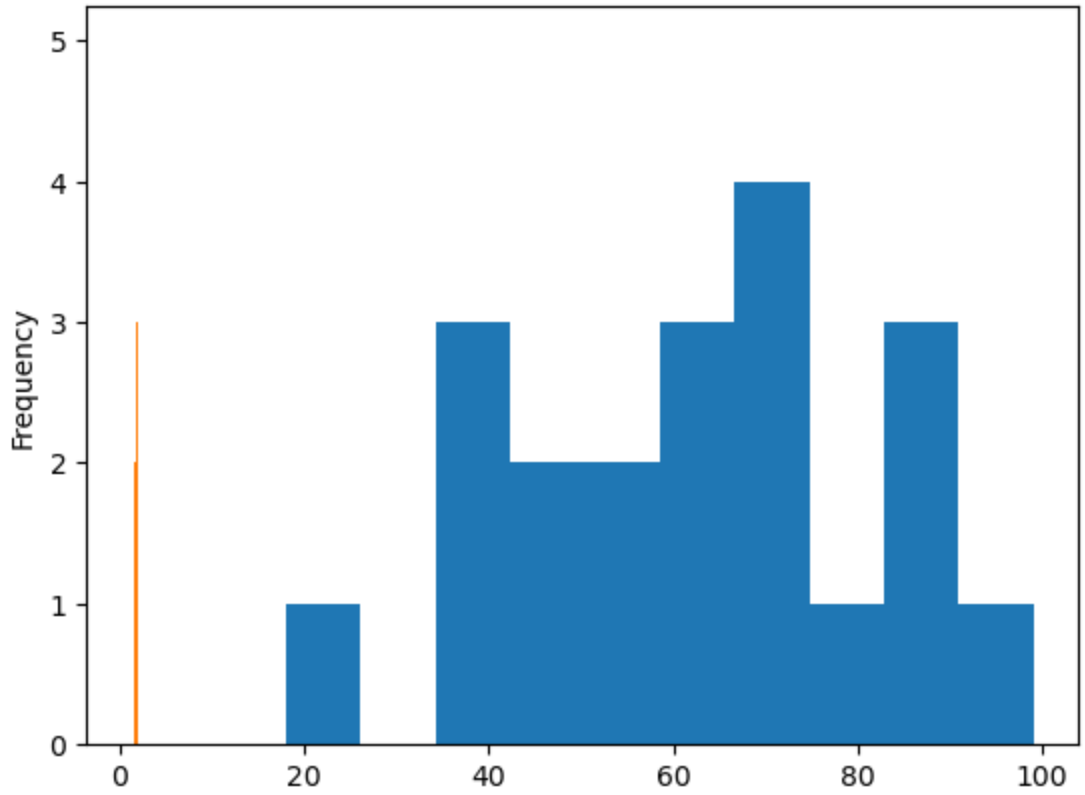
In [35]:
```
df
```

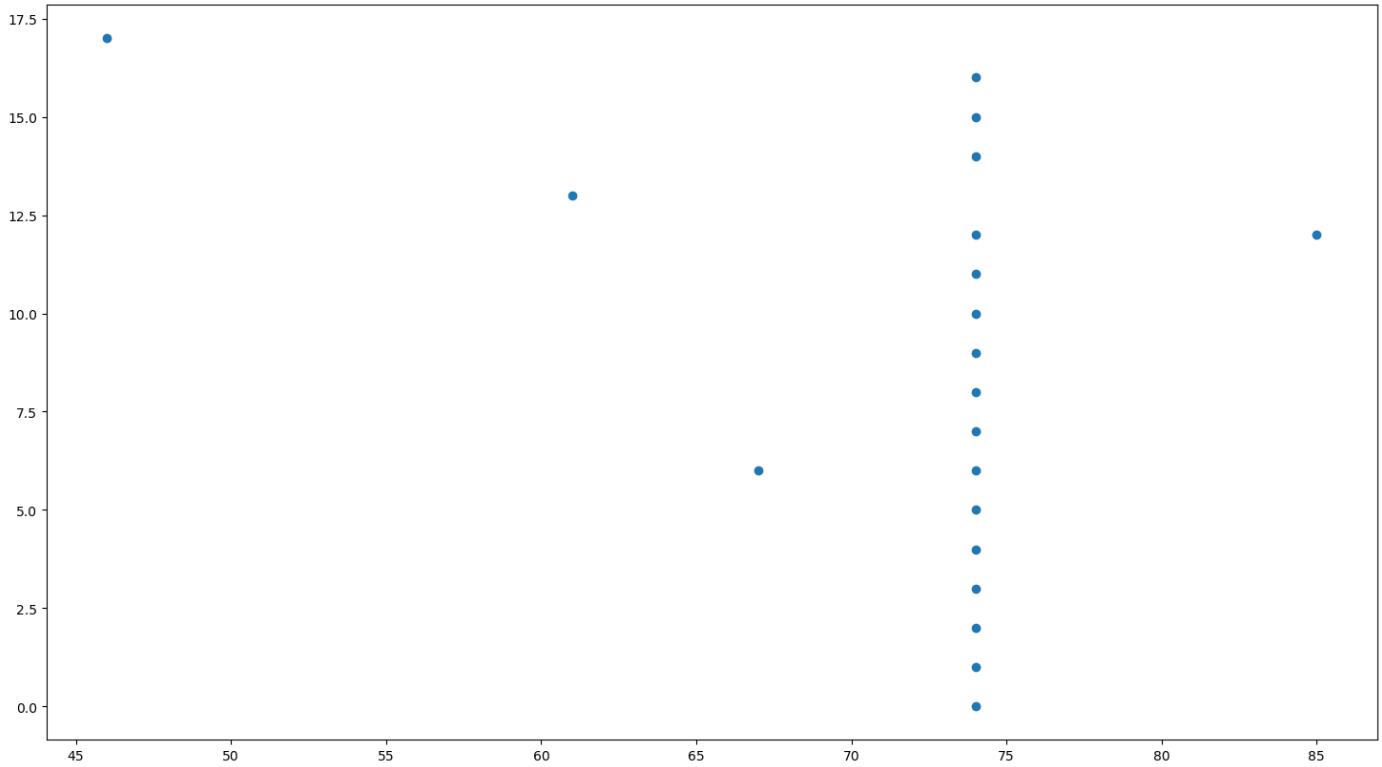
Out[35]:

	Maths Score	Reading Score	Writing Score	Placement score	Club Join Year	Placement Offer Count
0	72.000000	42.0	6	67	2019	1
1	69.000000	24.0	5	74	2019	1
2	90.000000	14.0	3	74	2018	1
3	99.000000	24.0	12	85	2019	3
4	76.000000	77.0	6	74	2020	1
5	71.000000	34.0	7	74	2019	1
6	88.000000	37.0	10	74	2018	1
7	40.000000	NaN	12	74	2018	1
8	64.000000	5.0	2	74	2018	1
9	38.000000	93.0	1	74	2019	1
10	58.000000	82.0	14	74	2020	1
11	40.000000	65.0	8	74	2018	1
12	65.000000	55.0	0	74	2018	1
13	62.894737	34.0	15	74	2019	1
14	50.000000	64.0	16	74	2020	1
15	69.000000	20.0	9	74	2018	1
16	88.000000	NaN	11	74	2020	1
17	18.000000	NaN	4	74	2019	1
18	46.000000	27.0	17	46	2020	1
19	54.000000	77.0	13	61	2018	1

In [37]:
```
import matplotlib.pyplot as plt
```

In [47]:
```
fig, ax = plt.subplots(figsize = (18,10))
ax.scatter(df['Placement score'], df['Writing Score'])
plt.show()
```





In [ ]: