In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("Book1.csv")
df
```

Out[1]:

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
0	70	76.0	68.0	99	2022	3
1	65	65.0	69.0	85	2023	2
2	68	82.0	65.0	99	2018	3
3	90	76.0	68.0	82	2019	2
4	76	NaN	89.0	91	2019	3
5	61	83.0	70.0	90	2019	3
6	76	NaN	NaN	78	2022	1
7	74	84.0	69.0	98	2018	3
8	86	88.0	79.0	75	2019	1
9	90	77.0	NaN	99	2022	3
10	66	69.0	77.0	85	2023	2
11	76	87.0	78.0	83	2023	2
12	78	84.0	77.0	90	2021	3
13	60	72.0	82.0	99	2020	3
14	60	62.0	80.0	99	2019	3
15	76	77.0	74.0	93	2021	3
16	83	61.0	60.0	98	2023	3
17	70	70.0	72.0	78	2022	1
18	78	72.0	78.0	83	2019	2
19	10	85.0	89.0	81	2020	2
20	67	74.0	63.0	80	2018	2
21	61	81.0	62.0	92	2018	3
22	79	85.0	78.0	85	2021	2
23	63	87.0	65.0	96	2019	3
24	80	65.0	60.0	80	2020	2
25	87	82.0	86.0	98	2020	3
26	86	81.0	60.0	95	2019	3
27	72	63.0	70.0	86	2019	2
28	78	85.0	76.0	91	2022	3
29	71	67.0	86.0	95	2019	3
30	64	90.0	69.0	92	2018	3
31	80	68.0	81.0	92	2021	3

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
32	72	70.0	62.0	85	2022	2
33	69	79.0	64.0	89	2020	2
34	77	71.0	69.0	83	2018	2
35	72	72.0	77.0	91	2019	3
36	81	87.0	82.0	91	2022	3
37	60	60.0	76.0	87	2023	2
38	80	61.0	70.0	78	2020	1
39	151	83.0	88.0	95	2018	3
40	62	73.0	87.0	85	2020	2
41	66	83.0	80.0	79	2023	1
42	65	60.0	79.0	77	2022	1
43	71	79.0	71.0	97	2018	3
44	76	79.0	84.0	87	2018	2
45	74	86.0	79.0	86	2022	2
46	90	77.0	87.0	94	2018	3
47	76	67.0	86.0	79	2021	1
48	62	89.0	89.0	87	2018	2

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

Out[4]:
```
Math Score              0
Reading Score           2
Writing Score           2
Placement Score         0
Club join year          0
Placement offer count   0
dtype: int64
```

In [5]:
```
df['Writing Score'].dropna()
```

Out[5]:
```
0     68.0
1     69.0
2     65.0
3     68.0
4     89.0
5     70.0
7     69.0
8     79.0
10    77.0
11    78.0
12    77.0
13    82.0
14    80.0
15    74.0
16    60.0
17    72.0
18    78.0
19    89.0
20    63.0
21    62.0
22    78.0
23    65.0
24    60.0
25    86.0
26    60.0
27    70.0
28    76.0
29    86.0
30    69.0
31    81.0
32    62.0
33    64.0
34    69.0
35    77.0
36    82.0
37    76.0
38    70.0
39    88.0
40    87.0
41    80.0
42    79.0
43    71.0
44    84.0
45    79.0
46    87.0
47    86.0
48    89.0
Name: Writing Score, dtype: float64
```

In [13]:
```
df['Reading Score'].dropna(inplace=True)
```

In [14]:
```
df.isnull().sum()
```

Out[14]:
```
Math Score              0
Reading Score           2
Writing Score           2
Placement Score         0
Club join year          0
Placement offer count   0
dtype: int64
```

In [17]:
```
m1=df['Reading Score'].mean()
m1
```

Out[17]:
```
76.04255319148936
```

In [23]:
```
m1=df['Reading Score'].mean()
df2=df['Reading Score'].fillna(m1,inplace=True)
df
```

Out[23]:

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
0	1.457480	76.000000	68.0	99	2022	3
1	1.444712	65.000000	69.0	85	2023	2
2	1.452499	82.000000	65.0	99	2018	3
3	1.499968	76.000000	68.0	82	2019	2
4	1.471518	76.042553	89.0	91	2019	3
5	1.433680	83.000000	70.0	90	2019	3
6	1.471518	76.042553	NaN	78	2022	1
7	1.466981	84.000000	69.0	98	2018	3
8	1.492372	88.000000	79.0	75	2019	1
9	1.499968	77.000000	NaN	99	2022	3
10	1.447352	69.000000	77.0	85	2023	2
11	1.471518	87.000000	78.0	83	2023	2
12	1.475925	84.000000	77.0	90	2021	3
13	1.430794	72.000000	82.0	99	2020	3
14	1.430794	62.000000	80.0	99	2019	3
15	1.471518	77.000000	74.0	93	2021	3
16	1.486412	61.000000	60.0	98	2023	3
17	1.457480	70.000000	72.0	78	2022	1
18	1.475925	72.000000	78.0	83	2019	2
19	1.072983	85.000000	89.0	81	2020	2
20	1.449947	74.000000	63.0	80	2018	2
21	1.433680	81.000000	62.0	92	2018	3
22	1.478081	85.000000	78.0	85	2021	2
23	1.439294	87.000000	65.0	96	2019	3
24	1.480207	65.000000	60.0	80	2020	2
25	1.494307	82.000000	86.0	98	2020	3
26	1.492372	81.000000	60.0	95	2019	3
27	1.462304	63.000000	70.0	86	2019	2
28	1.475925	85.000000	76.0	91	2022	3
29	1.459911	67.000000	86.0	95	2019	3
30	1.442027	90.000000	69.0	92	2018	3
31	1.480207	68.000000	81.0	92	2021	3

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
32	1.462304	70.000000	62.0	85	2022	2
33	1.455010	79.000000	64.0	89	2020	2
34	1.473738	71.000000	69.0	83	2018	2
35	1.462304	72.000000	77.0	91	2019	3
36	1.482304	87.000000	82.0	91	2022	3
37	1.430794	60.000000	76.0	87	2023	2
38	1.480207	61.000000	70.0	78	2020	1
39	1.583869	83.000000	88.0	95	2018	3
40	1.436512	73.000000	87.0	85	2020	2
41	1.447352	83.000000	80.0	79	2023	1
42	1.444712	60.000000	79.0	77	2022	1
43	1.459911	79.000000	71.0	97	2018	3
44	1.471518	79.000000	84.0	87	2018	2
45	1.466981	86.000000	79.0	86	2022	2
46	1.499968	77.000000	87.0	94	2018	3
47	1.471518	67.000000	86.0	79	2021	1
48	1.436512	89.000000	89.0	87	2018	2

In [20]:
```
df.dropna(inplace=True)
df
```

Out[20]:

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
0	70	76.000000	68.0	99	2022	3
1	65	65.000000	69.0	85	2023	2
2	68	82.000000	65.0	99	2018	3
3	90	76.000000	68.0	82	2019	2
4	76	76.042553	89.0	91	2019	3
5	61	83.000000	70.0	90	2019	3
7	74	84.000000	69.0	98	2018	3
8	86	88.000000	79.0	75	2019	1
10	66	69.000000	77.0	85	2023	2
11	76	87.000000	78.0	83	2023	2
12	78	84.000000	77.0	90	2021	3
13	60	72.000000	82.0	99	2020	3
14	60	62.000000	80.0	99	2019	3
15	76	77.000000	74.0	93	2021	3
16	83	61.000000	60.0	98	2023	3
17	70	70.000000	72.0	78	2022	1
18	78	72.000000	78.0	83	2019	2
19	10	85.000000	89.0	81	2020	2
20	67	74.000000	63.0	80	2018	2
21	61	81.000000	62.0	92	2018	3
22	79	85.000000	78.0	85	2021	2
23	63	87.000000	65.0	96	2019	3
24	80	65.000000	60.0	80	2020	2
25	87	82.000000	86.0	98	2020	3
26	86	81.000000	60.0	95	2019	3
27	72	63.000000	70.0	86	2019	2
28	78	85.000000	76.0	91	2022	3
29	71	67.000000	86.0	95	2019	3
30	64	90.000000	69.0	92	2018	3
31	80	68.000000	81.0	92	2021	3
32	72	70.000000	62.0	85	2022	2
33	69	79.000000	64.0	89	2020	2

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
34	77	71.000000	69.0	83	2018	2
35	72	72.000000	77.0	91	2019	3
36	81	87.000000	82.0	91	2022	3
37	60	60.000000	76.0	87	2023	2
38	80	61.000000	70.0	78	2020	1
39	151	83.000000	88.0	95	2018	3
40	62	73.000000	87.0	85	2020	2
41	66	83.000000	80.0	79	2023	1
42	65	60.000000	79.0	77	2022	1
43	71	79.000000	71.0	97	2018	3
44	76	79.000000	84.0	87	2018	2
45	74	86.000000	79.0	86	2022	2
46	90	77.000000	87.0	94	2018	3
47	76	67.000000	86.0	79	2021	1
48	62	89.000000	89.0	87	2018	2

In [21]:
```
df.isnull().sum()
```

Out[21]:
```
Math Score              0
Reading Score           0
Writing Score           0
Placement Score         0
Club join year          0
Placement offer count   0
dtype: int64
```

In [25]:
```
import numpy as np
m2=df['Math Score'].mean()
df['Math Score']=np.where((df['Math Score']<60) | (df['Math Score']>90),m2,c
df
```

Out[25]:

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
0	70.000000	76.000000	68.0	99	2022	3
1	65.000000	65.000000	69.0	85	2023	2
2	68.000000	82.000000	65.0	99	2018	3
3	90.000000	76.000000	68.0	82	2019	2
4	76.000000	76.042553	89.0	91	2019	3
5	61.000000	83.000000	70.0	90	2019	3
7	74.000000	84.000000	69.0	98	2018	3
8	86.000000	88.000000	79.0	75	2019	1
10	66.000000	69.000000	77.0	85	2023	2
11	76.000000	87.000000	78.0	83	2023	2
12	78.000000	84.000000	77.0	90	2021	3
13	60.000000	72.000000	82.0	99	2020	3
14	60.000000	62.000000	80.0	99	2019	3
15	76.000000	77.000000	74.0	93	2021	3
16	83.000000	61.000000	60.0	98	2023	3
17	70.000000	70.000000	72.0	78	2022	1
18	78.000000	72.000000	78.0	83	2019	2
19	73.170213	85.000000	89.0	81	2020	2
20	67.000000	74.000000	63.0	80	2018	2
21	61.000000	81.000000	62.0	92	2018	3
22	79.000000	85.000000	78.0	85	2021	2
23	63.000000	87.000000	65.0	96	2019	3
24	80.000000	65.000000	60.0	80	2020	2
25	87.000000	82.000000	86.0	98	2020	3
26	86.000000	81.000000	60.0	95	2019	3
27	72.000000	63.000000	70.0	86	2019	2
28	78.000000	85.000000	76.0	91	2022	3
29	71.000000	67.000000	86.0	95	2019	3
30	64.000000	90.000000	69.0	92	2018	3
31	80.000000	68.000000	81.0	92	2021	3
32	72.000000	70.000000	62.0	85	2022	2
33	69.000000	79.000000	64.0	89	2020	2

	Math Score	Reading Score	Writing Score	Placement Score	Club join year	Placement offer count
34	77.000000	71.000000	69.0	83	2018	2
35	72.000000	72.000000	77.0	91	2019	3
36	81.000000	87.000000	82.0	91	2022	3
37	60.000000	60.000000	76.0	87	2023	2
38	80.000000	61.000000	70.0	78	2020	1
39	73.170213	83.000000	88.0	95	2018	3
40	62.000000	73.000000	87.0	85	2020	2
41	66.000000	83.000000	80.0	79	2023	1
42	65.000000	60.000000	79.0	77	2022	1
43	71.000000	79.000000	71.0	97	2018	3
44	76.000000	79.000000	84.0	87	2018	2
45	74.000000	86.000000	79.0	86	2022	2
46	90.000000	77.000000	87.0	94	2018	3
47	76.000000	67.000000	86.0	79	2021	1
48	62.000000	89.000000	89.0	87	2018	2

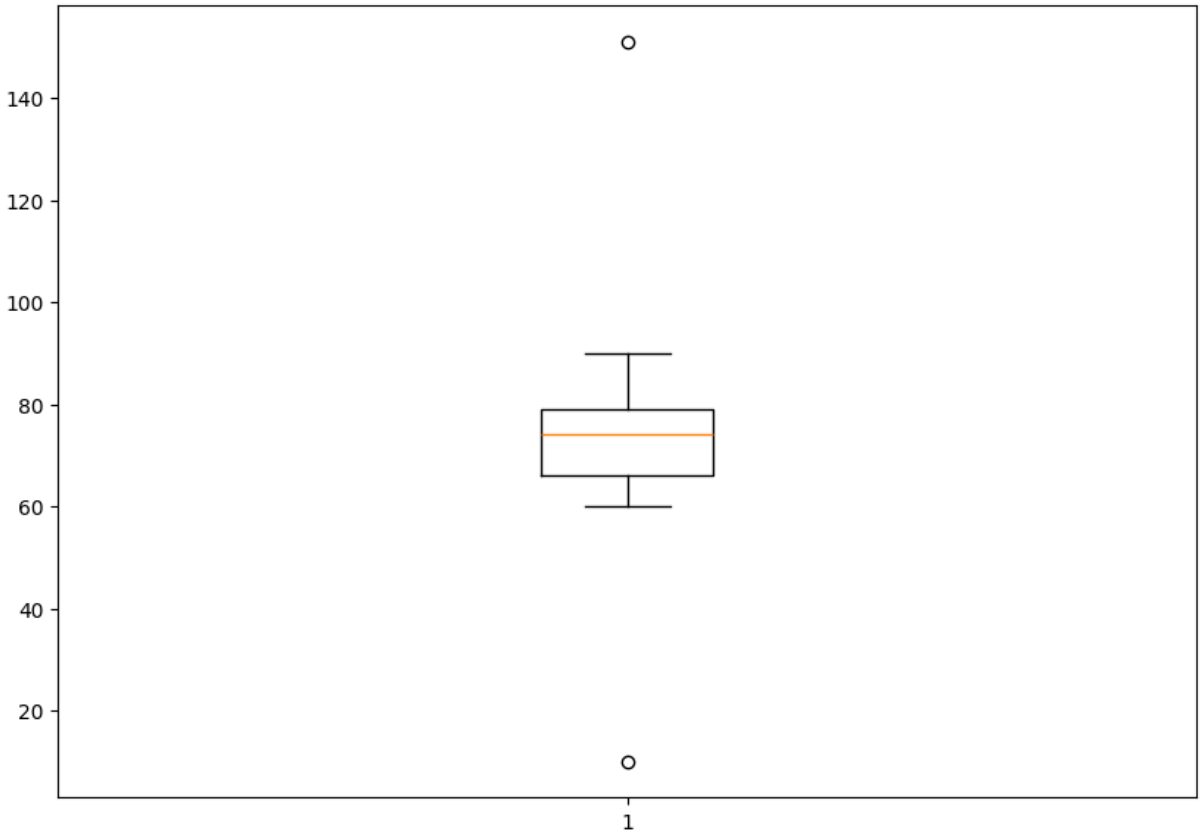
In [26]:
```
df.sum()
```

Out[26]:
```
Math Score                3424.340426
Reading Score             3573.042553
Writing Score             3530.000000
Placement Score           4160.000000
Club join year           94944.000000
Placement offer count      111.000000
dtype: float64
```

In [27]:
```
df.count()
```

Out[27]:
```
Math Score                47
Reading Score             47
Writing Score             47
Placement Score           47
Club join year            47
Placement offer count     47
dtype: int64
```

In [3]:
```
plt.figure(figsize =(10, 7))
plt.boxplot(df['Math Score'])
plt.show()
```



In [36]:
```
from scipy.stats import skew
import seaborn as sns
for col in df:
    print(col)
    print(skew(df[col]))
    # plt.figure()
    # sns.distplot(df[col])
    # plt.show()
```

```
Math Score
0.25810537579528003
Reading Score
-0.30038633049364255
Writing Score
nan
Placement Score
-0.0727034222170501
Club join year
0.25810537579528003
Placement offer count
-0.6218680623897693
Placement Scorer
0.02740156992096087
```

In [37]:
```
plt.figure()
sns.distplot(df["Placement offer count"])
plt.show()
```

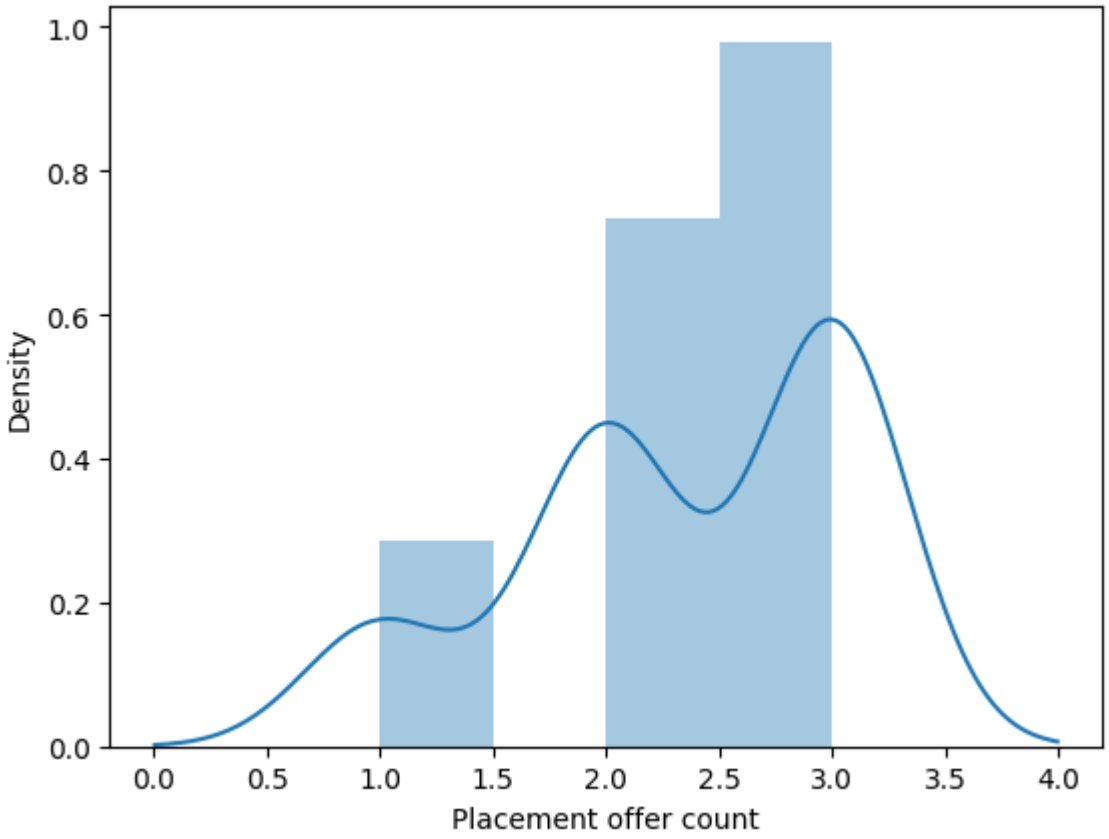
```
C:\Users\sneha\AppData\Local\Temp\ipykernel_11056\2189122855.py:2: UserWarni
ng:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["Placement offer count"])
```



In [44]:
```
df["Placement offer count"]=df['Placement offer count']**2
skew(df["Placement offer count"])
```

Out[44]: -0.2806110198833401

In [38]:
```
df["Placement offer count"]=df['Placement offer count']**2
skew(df["Placement offer count"])
plt.figure()
sns.distplot(df["Placement offer count"])
plt.show()
```

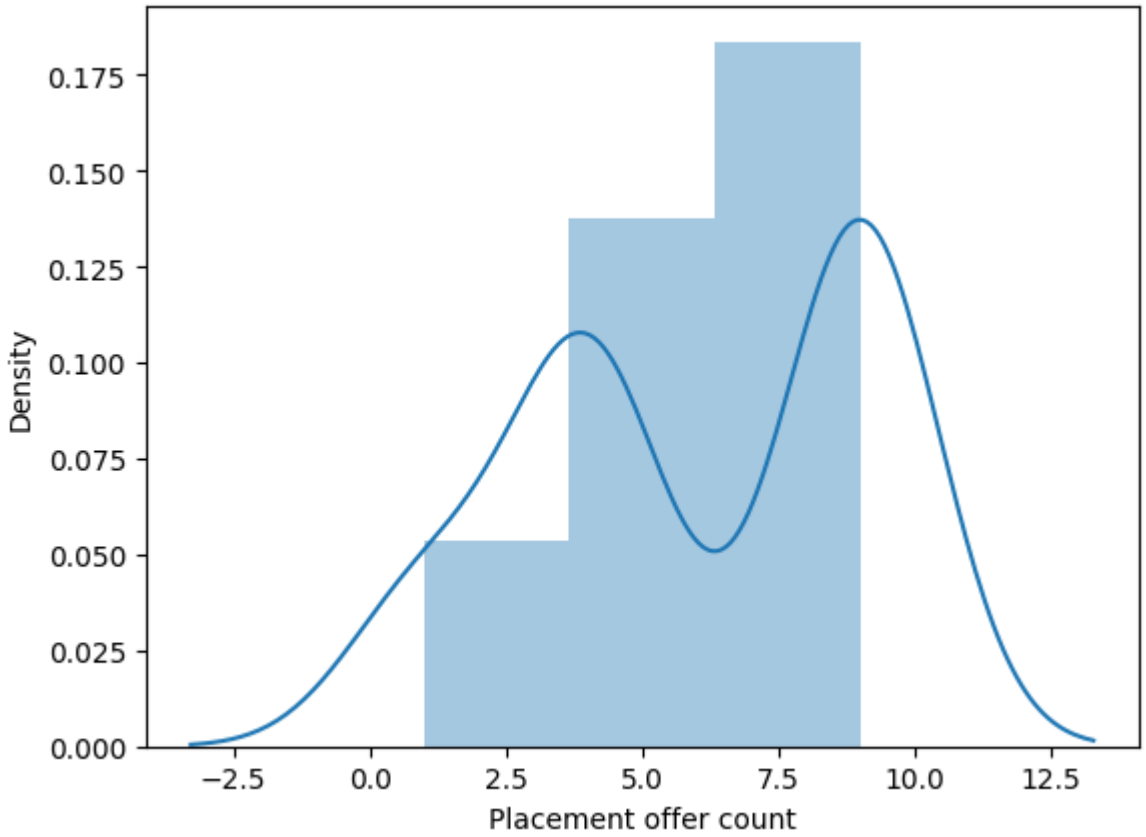
```
C:\Users\sneha\AppData\Local\Temp\ipykernel_11056\1051447204.py:4: UserWarni
ng:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["Placement offer count"])
```



In [ ]: