PRACTICAL NO : 02

DATA WRANGING 2

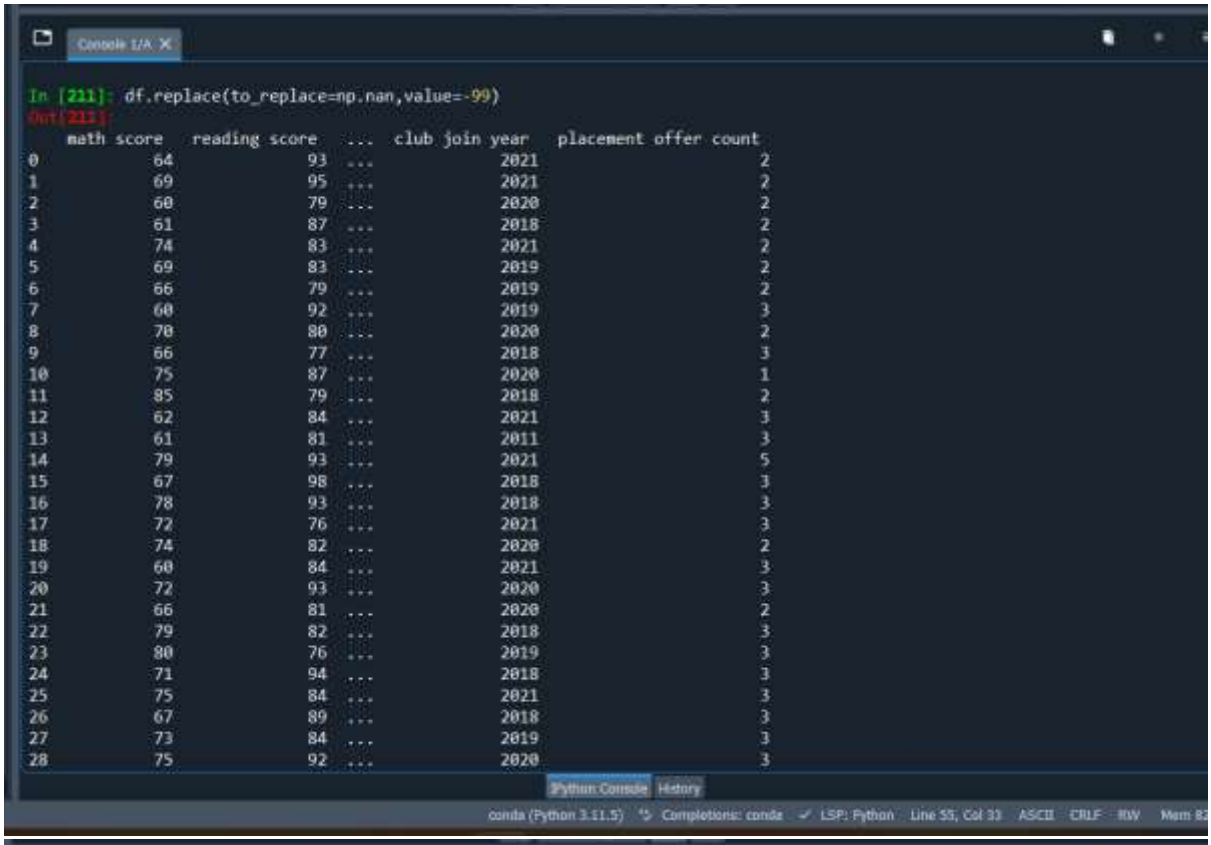
CODE :

```
 import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

df= pd.read_csv(r'E:\DSBDA\DSBDA PR 02\StudentPerformance.csv')
display(df)

#isnull
df.isnull()
data = pd.isnull(df['math score'])
display(data)

#notnull
df.notnull()
data = pd.notnull(df['math score'])
display(data)

#fillna
df.fillna(1)

#replace
```

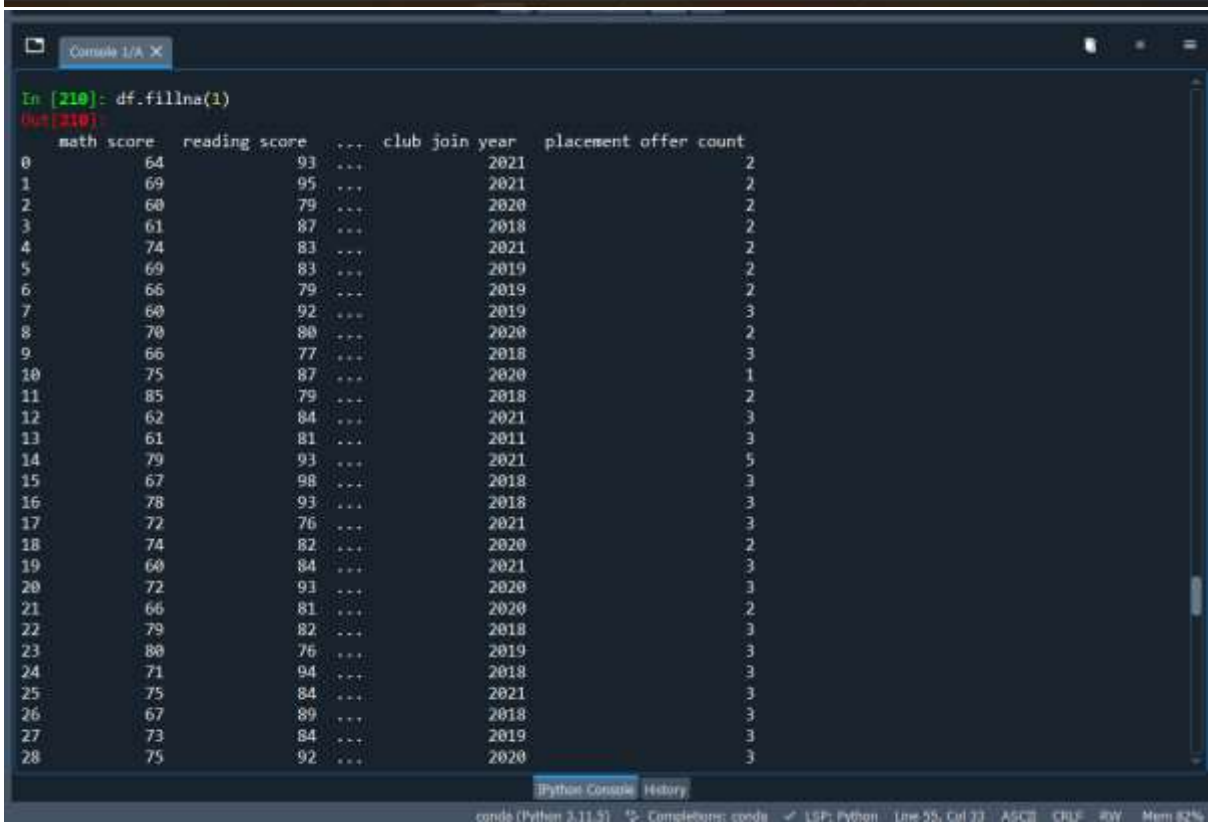
```
df.replace(to_replace=np.nan,value=-99)

#dropna
df.dropna()
df.dropna(axis = 1)
df.dropna(axis=0)

#Detecting outlier using Boxplot
col=['math score','reading score','writing score','placement score']
df.boxplot(col)
print(np.where(df['math score']>90))
print(np.where(df['reading score']<25))
print(np.where(df['writing score']<30))

#Detecting outlier using Scatterplot
fig, ax=plt.subplots(figsize=(18,10))
ax.scatter(df['placement score'],df['placement offer count'])
ax.set_xlabel('placement score')
ax.set_ylabel('placement offer count')
ax.set_title('scatter plot')
plt.show()

#Detecting outlier using Z-score
z=np.abs(stats.zscore(df['math score']))
print(z)
threshold = 0.18
```

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


#Histogram
df['math score'].plot(kind='hist')
df['log_math'] = np.log10(df['math score'])
df['log_math'].plot(kind='hist')
```

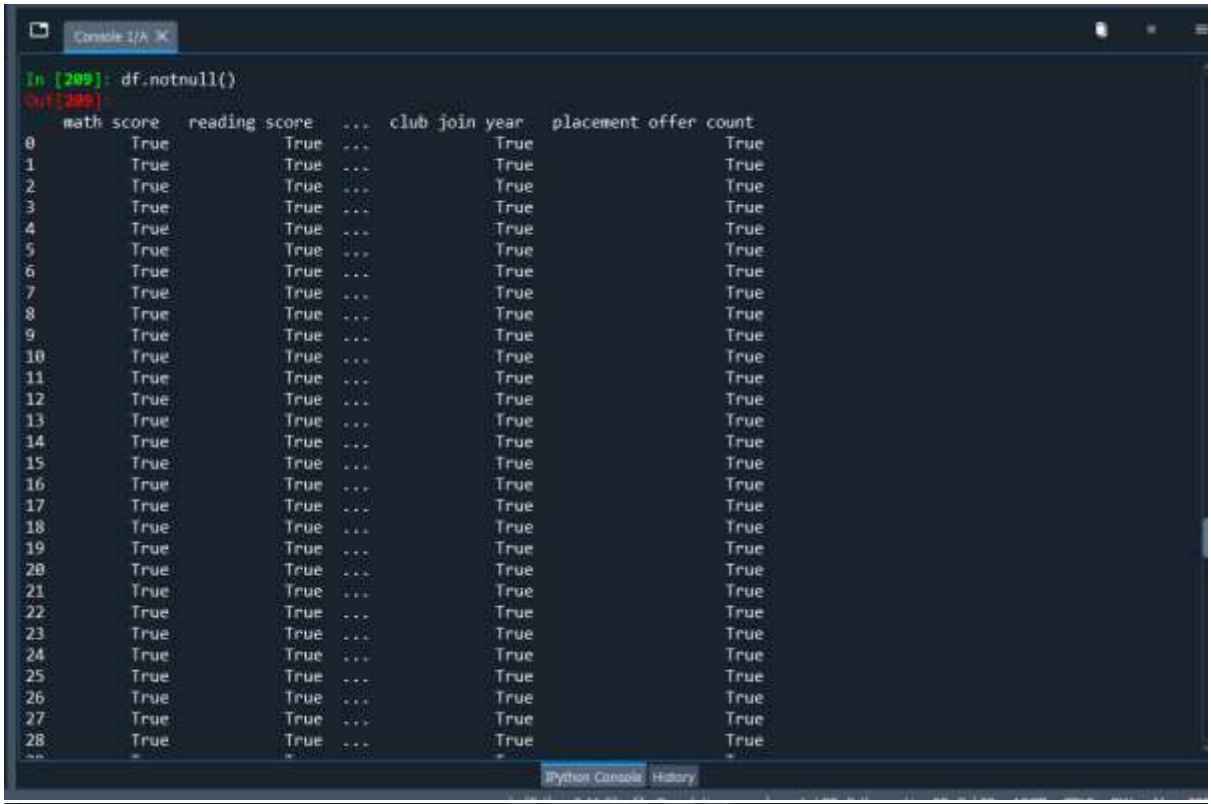
OUTPUT :

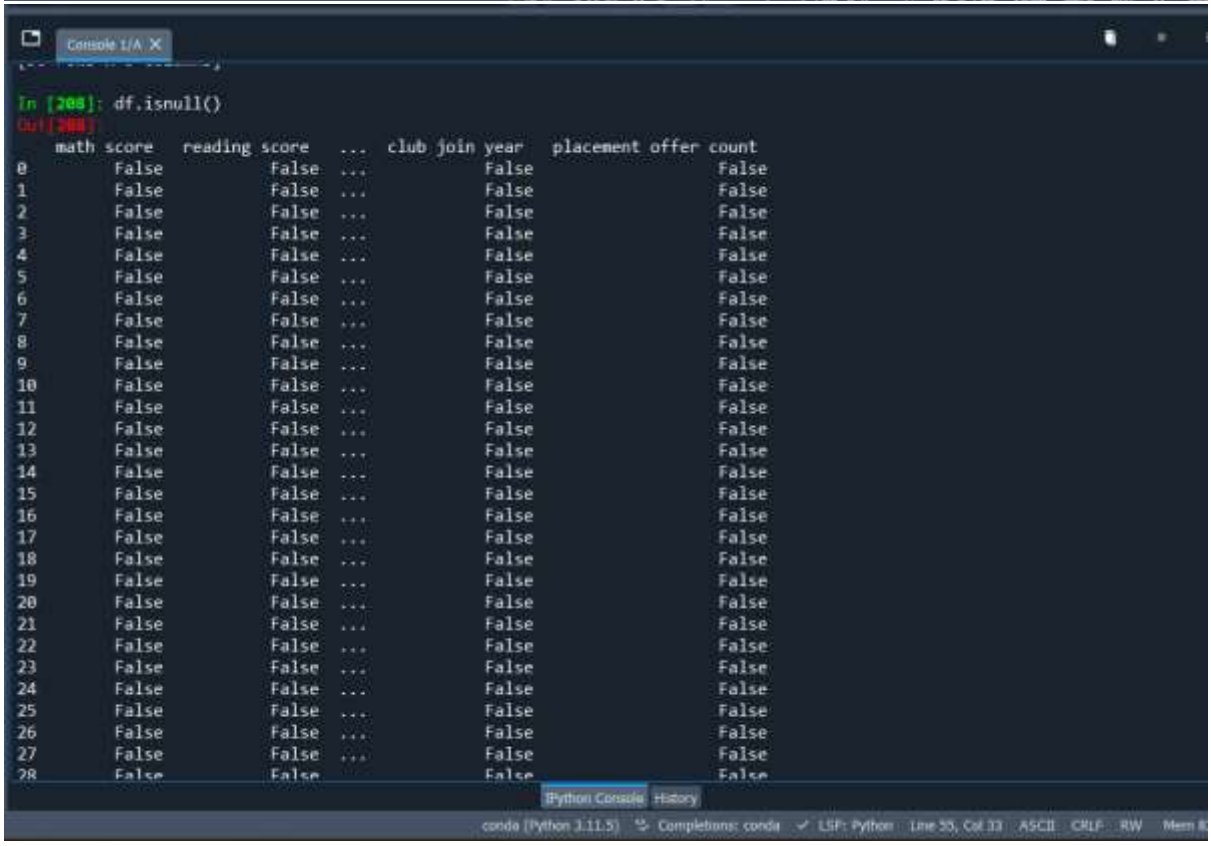


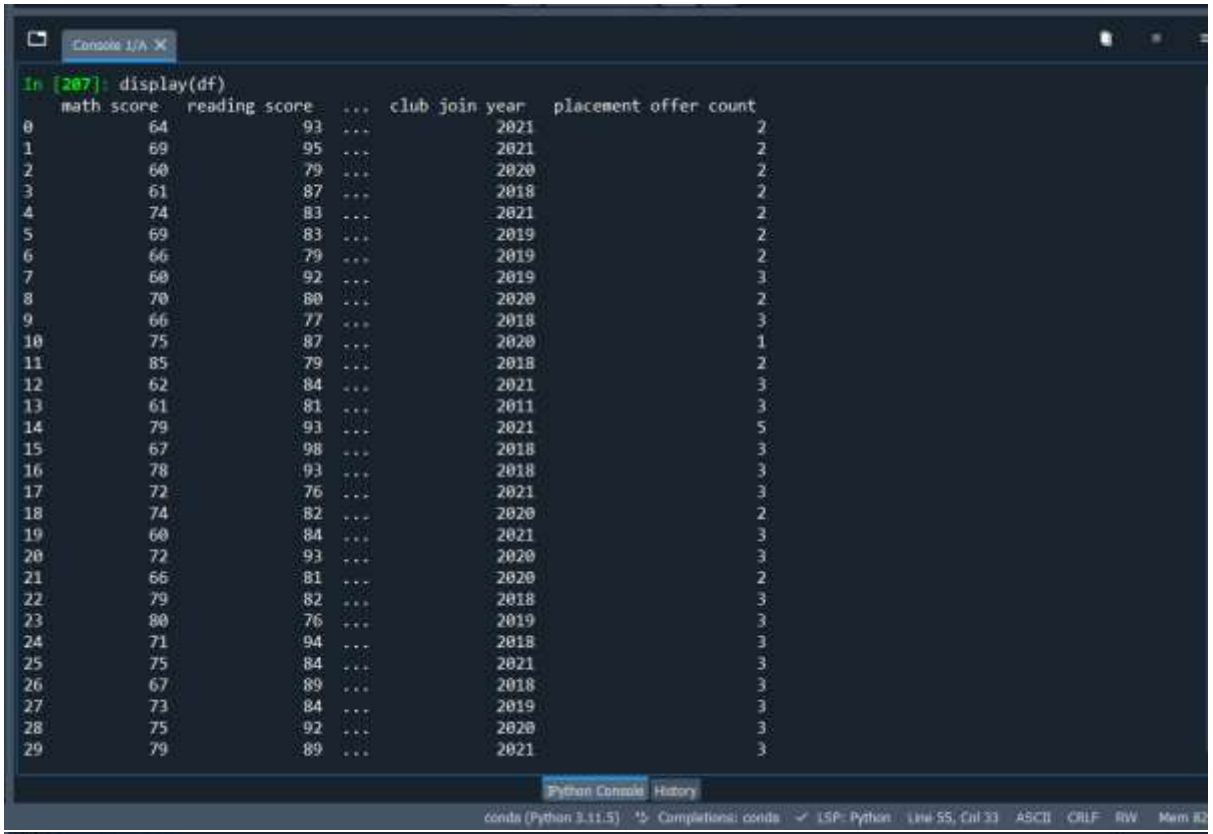
```
In [211]: df.replace(to_replace=np.nan,value=-99)
Out[211]:
    math score   reading score   ...  club join year   placement offer count
0           64              93  ...            2021                       2
1           69              95  ...            2021                       2
2           60              79  ...            2020                       2
3           61              87  ...            2018                       2
4           74              83  ...            2021                       2
5           69              83  ...            2019                       2
6           66              79  ...            2019                       2
7           60              92  ...            2019                       3
8           70              80  ...            2020                       2
9           66              77  ...            2018                       3
10          75              87  ...            2020                       1
11          85              79  ...            2018                       2
12          62              84  ...            2021                       3
13          61              81  ...            2011                       3
14          79              93  ...            2021                       5
15          67              98  ...            2018                       3
16          78              93  ...            2018                       3
17          72              76  ...            2021                       3
18          74              82  ...            2020                       2
19          60              84  ...            2021                       3
20          72              93  ...            2020                       3
21          66              81  ...            2020                       2
22          79              82  ...            2018                       3
23          80              76  ...            2019                       3
24          71              94  ...            2018                       3
25          75              84  ...            2021                       3
26          67              89  ...            2018                       3
27          73              84  ...            2019                       3
28          75              92  ...            2020                       3
```



```
In [210]: df.fillna(1)
Out[210]:
    math score   reading score   ...  club join year   placement offer count
0           64              93  ...            2021                       2
1           69              95  ...            2021                       2
2           60              79  ...            2020                       2
3           61              87  ...            2018                       2
4           74              83  ...            2021                       2
5           69              83  ...            2019                       2
6           66              79  ...            2019                       2
7           60              92  ...            2019                       3
8           70              80  ...            2020                       2
9           66              77  ...            2018                       3
10          75              87  ...            2020                       1
11          85              79  ...            2018                       2
12          62              84  ...            2021                       3
13          61              81  ...            2011                       3
14          79              93  ...            2021                       5
15          67              98  ...            2018                       3
16          78              93  ...            2018                       3
17          72              76  ...            2021                       3
18          74              82  ...            2020                       2
19          60              84  ...            2021                       3
20          72              93  ...            2020                       3
21          66              81  ...            2020                       2
22          79              82  ...            2018                       3
23          80              76  ...            2019                       3
24          71              94  ...            2018                       3
25          75              84  ...            2021                       3
26          67              89  ...            2018                       3
27          73              84  ...            2019                       3
28          75              92  ...            2020                       3
```


Console 1/A
In [209]: df.notnull()
Out[209]:
math score reading score ... club join year placement offer count
0 True True ... True True
1 True True ... True True
2 True True ... True True
3 True True ... True True
4 True True ... True True
5 True True ... True True
6 True True ... True True
7 True True ... True True
8 True True ... True True
9 True True ... True True
10 True True ... True True
11 True True ... True True
12 True True ... True True
13 True True ... True True
14 True True ... True True
15 True True ... True True
16 True True ... True True
17 True True ... True True
18 True True ... True True
19 True True ... True True
20 True True ... True True
21 True True ... True True
22 True True ... True True
23 True True ... True True
24 True True ... True True
25 True True ... True True
26 True True ... True True
27 True True ... True True
28 True True ... True True
IPython Console History

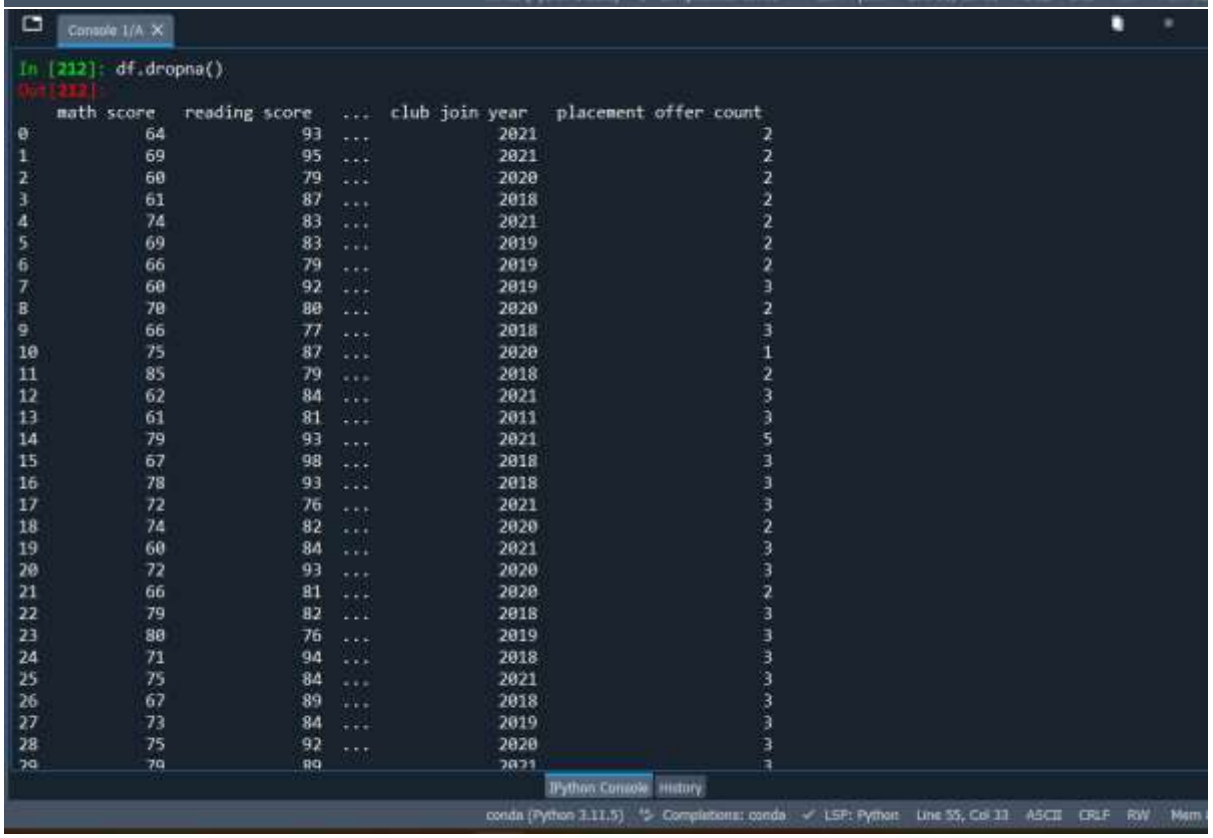

Console 1/A
In [208]: df.isnull()
Out[208]:
math score reading score ... club join year placement offer count
0 False False ... False False
1 False False ... False False
2 False False ... False False
3 False False ... False False
4 False False ... False False
5 False False ... False False
6 False False ... False False
7 False False ... False False
8 False False ... False False
9 False False ... False False
10 False False ... False False
11 False False ... False False
12 False False ... False False
13 False False ... False False
14 False False ... False False
15 False False ... False False
16 False False ... False False
17 False False ... False False
18 False False ... False False
19 False False ... False False
20 False False ... False False
21 False False ... False False
22 False False ... False False
23 False False ... False False
24 False False ... False False
25 False False ... False False
26 False False ... False False
27 False False ... False False
IPython Console History
conda (Python 3.11.5) Completions: conda LSP: Python Line 55, Col 33 ASCII CRLF RW



```
In [207]: display(df)
   math score   reading score   ...   club join year   placement offer count
0           64              93   ...             2021                       2
1           69              95   ...             2021                       2
2           60              79   ...             2020                       2
3           61              87   ...             2018                       2
4           74              83   ...             2021                       2
5           69              83   ...             2019                       2
6           66              79   ...             2019                       2
7           60              92   ...             2019                       3
8           70              80   ...             2020                       2
9           66              77   ...             2018                       3
10          75              87   ...             2020                       1
11          85              79   ...             2018                       2
12          62              84   ...             2021                       3
13          61              81   ...             2011                       3
14          79              93   ...             2021                       5
15          67              98   ...             2018                       3
16          78              93   ...             2018                       3
17          72              76   ...             2021                       3
18          74              82   ...             2020                       2
19          60              84   ...             2021                       3
20          72              93   ...             2020                       3
21          66              81   ...             2020                       2
22          79              82   ...             2018                       3
23          80              76   ...             2019                       3
24          71              94   ...             2018                       3
25          75              84   ...             2021                       3
26          67              89   ...             2018                       3
27          73              84   ...             2019                       3
28          75              92   ...             2020                       3
29          79              89   ...             2021                       3
```

IPython Console History

conda (Python 3.11.5) Completions: conda LSP: Python Line 55, Col 33 ASCII CRLF RW Mem 82



```
In [212]: df.dropna()
Out[212]:
   math score   reading score   ...   club join year   placement offer count
0           64              93   ...             2021                       2
1           69              95   ...             2021                       2
2           60              79   ...             2020                       2
3           61              87   ...             2018                       2
4           74              83   ...             2021                       2
5           69              83   ...             2019                       2
6           66              79   ...             2019                       2
7           60              92   ...             2019                       3
8           70              80   ...             2020                       2
9           66              77   ...             2018                       3
10          75              87   ...             2020                       1
11          85              79   ...             2018                       2
12          62              84   ...             2021                       3
13          61              81   ...             2011                       3
14          79              93   ...             2021                       5
15          67              98   ...             2018                       3
16          78              93   ...             2018                       3
17          72              76   ...             2021                       3
18          74              82   ...             2020                       2
19          60              84   ...             2021                       3
20          72              93   ...             2020                       3
21          66              81   ...             2020                       2
22          79              82   ...             2018                       3
23          80              76   ...             2019                       3
24          71              94   ...             2018                       3
25          75              84   ...             2021                       3
26          67              89   ...             2018                       3
27          73              84   ...             2019                       3
28          75              92   ...             2020                       3
29          79              89   ...             2021                       3
```

IPython Console History

conda (Python 3.11.5) Completions: conda LSP: Python Line 55, Col 33 ASCII CRLF RW Mem