In [ ]:
```
import pandas as pd
import numpy as np
df = pd.read_csv("StudentPerformance.csv")
df
```

Out[3]:

	math score	reading score	writing score	placement score	club join year
0	63	84	64	84	2020
1	71	80	76	86	2018
2	64	81	66	81	2020
3	71	85	77	91	2018
4	68	86	76	92	2021
5	79	86	61	100	2019
6	75	79	66	76	2020
7	71	79	66	95	2019
8	66	88	66	88	2020
9	70	79	61	87	2021
10	78	80	65	85	2021
11	76	884	73	92	2020
12	74	79	79	98	2019

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

```
    math score  reading score  writing score  placement score  club joi
n year
0        False          False          False            False
False
1        False          False          False            False
False
2        False          False          False            False
False
3        False          False          False            False
False
4        False          False          False            False
False
5        False          False          False            False
False
6         True          False          False            False
False
7        False          False          False            False
False
8        False          False          False            False
False
9        False          False          False            False
False
10       False          False          False            False
False
11        True          False          False            False
False
12       False          False          False            False
False
```

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

Out[7]:

	math score	reading score	writing score	placement score	club join year
6	NaN	79	66	76	2020
11	NaN	884	73	92	2020

In [ ]:
```
print(df.columns)
df.columns = df.columns.str.strip()
print(df.head())
```

```
Index(['gender', 'math score', 'reading score', 'writing score',
       'placement score', 'club join year'],
      dtype='object')
   gender  math score  reading score  writing score  placement score  \
0  female        63.0             84             64               84
1  female        71.0             80             76               86
2    male        64.0             81             66               81
3    male        71.0             85             77               91
4  female        68.0             86             76               92

   club join year
0            2020
1            2018
2            2020
3            2018
4            2021
```

In [ ]:
```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
df = pd.read_csv("StudentPerformance.csv")
df.columns = df.columns.str.strip()
if 'gender' in df.columns:
    le = LabelEncoder()
    df['gender'] = le.fit_transform(df['gender'])
    print("Gender column encoded successfully!")
else:
    print("Error: 'gender' column not found in the dataset!")
print(df.head())
```

```
Gender column encoded successfully!
   gender  math score  reading score  writing score  placement score  \
0       0        63.0             84             64               84
1       0        71.0             80             76               86
2       1        64.0             81             66               81
3       1        71.0             85             77               91
4       0        68.0             86             76               92

   club join year
0            2020
1            2018
2            2020
3            2018
4            2021
```

In [ ]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['gender'] = le.fit_transform(df['gender'])
df
```

Out[18]:

	gender	math score	reading score	writing score	placement score	club join year
0	0	63.0	84	64	84	2020
1	0	71.0	80	76	86	2018
2	1	64.0	81	66	81	2020
3	1	71.0	85	77	91	2018
4	0	68.0	86	76	92	2021
5	1	79.0	86	61	100	2019
6	1	NaN	79	66	76	2020
7	1	71.0	79	66	95	2019
8	0	66.0	88	66	88	2020
9	1	70.0	79	61	87	2021
10	0	78.0	80	65	85	2021
11	0	NaN	884	73	92	2020
12	1	74.0	79	79	98	2019

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

Out[21]:

	gender	math score	reading score	writing score	placement score	club join year
0	female	63.0	84	64	84	2020
1	female	71.0	80	76	86	2018
2	male	64.0	81	66	81	2020
3	male	71.0	85	77	91	2018
4	female	68.0	86	76	92	2021
5	male	79.0	86	61	100	2019
6	male	NaN	79	66	76	2020
7	male	71.0	79	66	95	2019
8	female	66.0	88	66	88	2020
9	male	70.0	79	61	87	2021
10	female	78.0	80	65	85	2021
11	female	NaN	884	73	92	2020
12	male	74.0	79	79	98	2019

In [ ]:
```
df
d=df
d.fillna(50)
```

Out[24]:

	gender	math score	reading score	writing score	placement score	club join year
0	female	63.0	84	64	84	2020
1	female	71.0	80	76	86	2018
2	male	64.0	81	66	81	2020
3	male	71.0	85	77	91	2018
4	female	68.0	86	76	92	2021
5	male	79.0	86	61	100	2019
6	male	50.0	79	66	76	2020
7	male	71.0	79	66	95	2019
8	female	66.0	88	66	88	2020
9	male	70.0	79	61	87	2021
10	female	78.0	80	65	85	2021
11	female	50.0	884	73	92	2020
12	male	74.0	79	79	98	2019

In [ ]:
```
import pandas as pd
missing_values = ["Na", "na"]
df = pd.read_csv("StudentPerformance.csv", na_values=missing_values)
mean_value = df['math score'].mean()
median_value = df['math score'].median()
mode_value = df['math score'].mode()[0]
print("Mean:", mean_value)
print("Median:", median_value)
print("Mode:", mode_value)
```

```
Mean: 70.45454545454545
Median: 71.0
Mode: 71.0
```

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

Out[31]:

	gender	reading score	writing score	placement score	club join year
0	female	84	64	84	2020
1	female	80	76	86	2018
2	male	81	66	81	2020
3	male	85	77	91	2018
4	female	86	76	92	2021
5	male	86	61	100	2019
6	male	79	66	76	2020
7	male	79	66	95	2019
8	female	88	66	88	2020
9	male	79	61	87	2021
10	female	80	65	85	2021
11	female	884	73	92	2020
12	male	79	79	98	2019

In [ ]:
```
new_d=df.dropna(axis=0,how='any')
new_d
```

Out[32]:

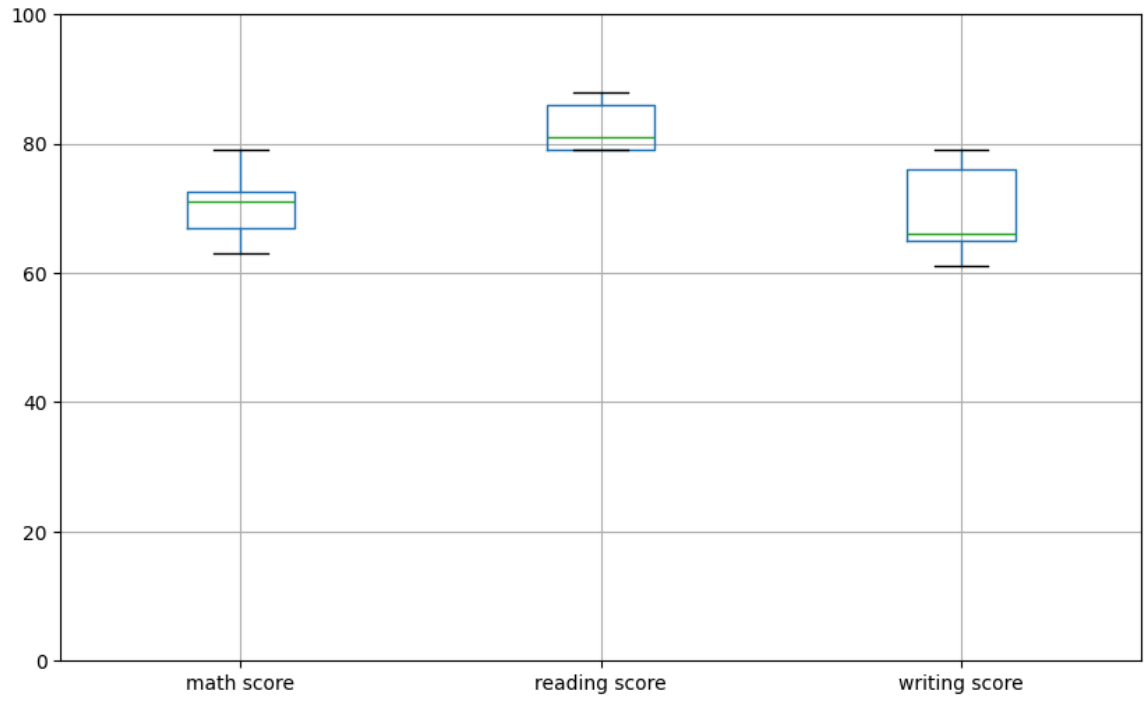
	gender	math score	reading score	writing score	placement score	club join year
0	female	63.0	84	64	84	2020
1	female	71.0	80	76	86	2018
2	male	64.0	81	66	81	2020
3	male	71.0	85	77	91	2018
4	female	68.0	86	76	92	2021
5	male	79.0	86	61	100	2019
7	male	71.0	79	66	95	2019
8	female	66.0	88	66	88	2020
9	male	70.0	79	61	87	2021
10	female	78.0	80	65	85	2021
12	male	74.0	79	79	98	2019

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

missing_values = ["Na", "na"]
df = pd.read_csv("StudentPerformance.csv", na_values=missing_values)
num_bins=5
df['math_binned'] = pd.cut(df['math score'], bins=num_bins, labels=False
print(df[['math score', 'math_binned']].head())
```

```
   math score  math_binned
0        63.0          0.0
1        71.0          2.0
2        64.0          0.0
3        71.0          2.0
4        68.0          1.0
```

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df= pd.read_csv("StudentPerformance.csv")
df
plt.figure(figsize=(10, 6))
df.boxplot(column=['math score', 'reading score', 'writing score'])
plt.ylim(0,100)
plt.show()
```



In [ ]: