

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('/content/StudentsPerformance.csv')
```

df

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
1	female	group C	some college	standard	completed	69.0	NaN	88.0
2	female	group B	master's degree	standard	none	90.0	95.0	93.0
3	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	male	group C	some college	standard	none	76.0	78.0	75.0
...
995	female	group E	master's degree	standard	completed	88.0	99.0	95.0

```
df.isnull()
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	True	False
2	False	False	False	False	False	False	False	False

```
df.isnull().sum()
```

```
gender          0
race/ethnicity  0
parental level of education  0
lunch           0
test preparation course      0
math score       1
reading score    2
writing score    1
dtype: int64
```

998	False	False	False	False	False	False	False	False
-----	-------	-------	-------	-------	-------	-------	-------	-------

```
df.fillna(df.max(),inplace=True)
```

```
df.isnull().sum()
```

```
gender          0
race/ethnicity  0
parental level of education  0
lunch           0
test preparation course      0
math score       0
reading score    0
writing score    0
dtype: int64
```

finding and handling outliers

```
df['math score']
```

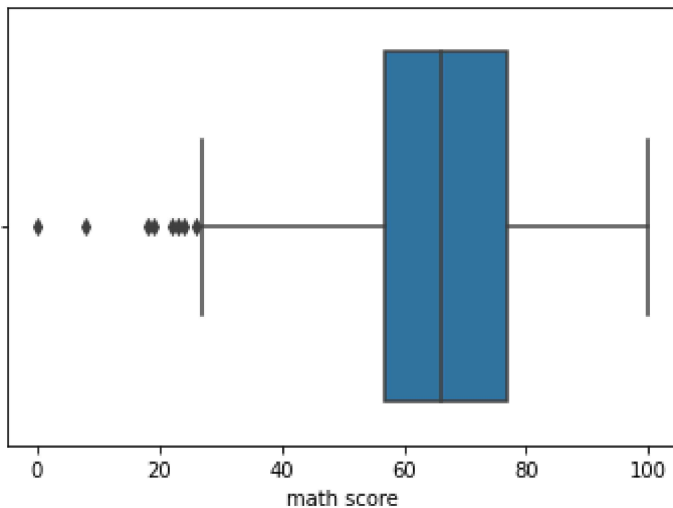
```
0    72.0
1    69.0
2    90.0
3    47.0
4    76.0
...
995   88.0
996   62.0
997   59.0
998   68.0
999   77.0
Name: math score, Length: 1000, dtype: float64
```

```
df['math score'].describe()
```

```
count    1000.000000
mean      66.116000
std       15.199394
min        0.000000
25%       57.000000
50%       66.000000
75%       77.000000
max      100.000000
Name: math score, dtype: float64
```

```
sns.boxplot(df['math score'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: P
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fdef5ed0550>
```



```
print("Highest Score : ",df['math score'].mean() + 3*df['math score'].std())
print("Lowest Score : ",df['math score'].mean() - 3*df['math score'].std())
```

```
Highest Score : 111.71418077610204
Lowest Score : 20.51781922389796
```

```
df[(df['math score']>111.714)|(df['math score']<20.517)]
```

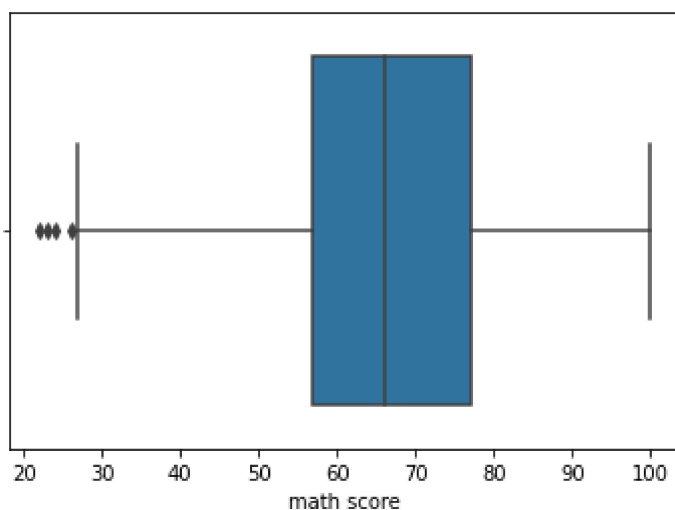
```
new_df=df[(df['math score']<111.714)&(df['math score']>20.517)]
```

```
new_df
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72.0	72.0	72.0
1	female	group C	some college	standard	completed	69.0	100.0	90.0
2	female	group B	master's degree	standard	none	90.0	95.0	93.0
3	male	group A	associate's degree	free/reduced	none	47.0	57.0	54.0
4	male	group C	some college	standard	none	76.0	78.0	77.0
...
995	female	group E	master's degree	standard	completed	88.0	99.0	97.0

```
sns.boxplot(new_df['math score'])
```

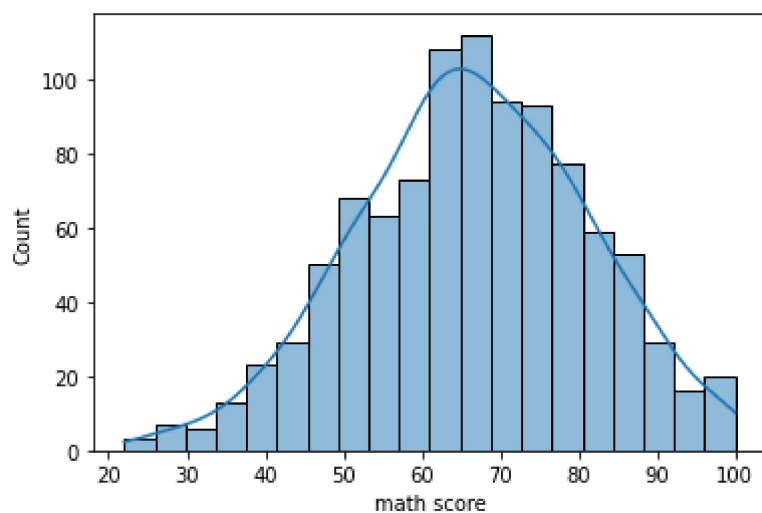
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyword arguments: {'x': 'math score'}.
FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7fdef4af9690>
```



to reduce skewness and normal distribution

```
sns.histplot(new_df['math score'],kde=True)
```

↪ <matplotlib.axes._subplots.AxesSubplot at 0x7fdef24891d0>



✓ 0s completed at 5:56 PM

● ✕