

```
In [141]: !pip install matplotlib
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: matplotlib in c:\users\shrut\appdata\roaming\python\python311\site-packages (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy<2,>=1.21 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=8 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

```
In [143]: !pip install seaborn
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: seaborn in c:\programdata\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\shrut\appdata\roaming\python\python311\site-packages (from seaborn) (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=8 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

```
In [144]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [145]: df=pd.read_csv("student_score.csv")
```

```
In [146]: df.shape
```

```
Out[146]: (30641, 15)
```

```
In [19]: print(df.head())
```

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	\
0	female	NaN	bachelor's degree	standard	none	
1	female	group C	some college	standard	NaN	
2	female	group B	master's degree	standard	none	
3	male	group A	associate's degree	free/reduced	none	
4	male	group C	some college	standard	none	

	ParentMaritalStatus	PracticeSport	IsFirstChild	NrSiblings	TransportMeans	\
0	married	regularly	yes	3.0	school_bus	
1	married	sometimes	yes	0.0	NaN	
2	single	sometimes	yes	4.0	school_bus	
3	married	never	no	1.0	NaN	
4	married	sometimes	yes	0.0	school_bus	

	WklyStudyHours	MathScore	ReadingScore	WritingScore
0	< 5	71	71	74
1	5 - 10	69	90	88
2	< 5	87	93	91
3	5 - 10	45	56	42
4	5 - 10	76	78	75

```
In [18]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30641 entries, 0 to 30640
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Gender                 30641 non-null object
1   EthnicGroup            28801 non-null object
2   ParentEduc             28796 non-null object
3   LunchType              30641 non-null object
4   TestPrep               28811 non-null object
5   ParentMaritalStatus    29451 non-null object
6   PracticeSport          30010 non-null object
7   IsFirstChild           29737 non-null object
8   NrSiblings             29069 non-null float64
9   TransportMeans         27507 non-null object
10  WklyStudyHours         29686 non-null object
11  MathScore              30641 non-null int64
12  ReadingScore           30641 non-null int64
13  WritingScore           30641 non-null int64
dtypes: float64(1), int64(3), object(10)
memory usage: 3.3+ MB
```

```
In [147]: pd.isnull(df).sum()
```

```
Out[147]: Unnamed: 0      0
Gender      0
EthnicGroup 1840
ParentEduc  1845
LunchType   0
TestPrep    1830
ParentMaritalStatus 1190
PracticeSport 631
IsFirstChild 904
NrSiblings  1572
TransportMeans 3134
WklyStudyHours 955
MathScore   0
ReadingScore 0
WritingScore 0
dtype: int64
```

```
In [148]: df.drop(["Unnamed: 0"],axis=1,inplace=True)
```

```
In [149]: df.columns
```

```
Out[149]: Index(['Gender', 'EthnicGroup', 'ParentEduc', 'LunchType', 'TestPrep',
      'ParentMaritalStatus', 'PracticeSport', 'IsFirstChild', 'NrSiblings',
      'TransportMeans', 'WklyStudyHours', 'MathScore', 'ReadingScore',
      'WritingScore'],
      dtype='object')
```

```
In [150]: df.describe()
```

```
Out[150]:
```

	NrSiblings	MathScore	ReadingScore	WritingScore
count	29069.000000	30641.000000	30641.000000	30641.000000
mean	2.145894	66.558402	69.377533	68.418622
std	1.458242	15.361616	14.758952	15.443525
min	0.000000	0.000000	10.000000	4.000000
25%	1.000000	56.000000	59.000000	58.000000
50%	2.000000	67.000000	70.000000	69.000000
75%	3.000000	78.000000	80.000000	79.000000
max	7.000000	100.000000	100.000000	100.000000

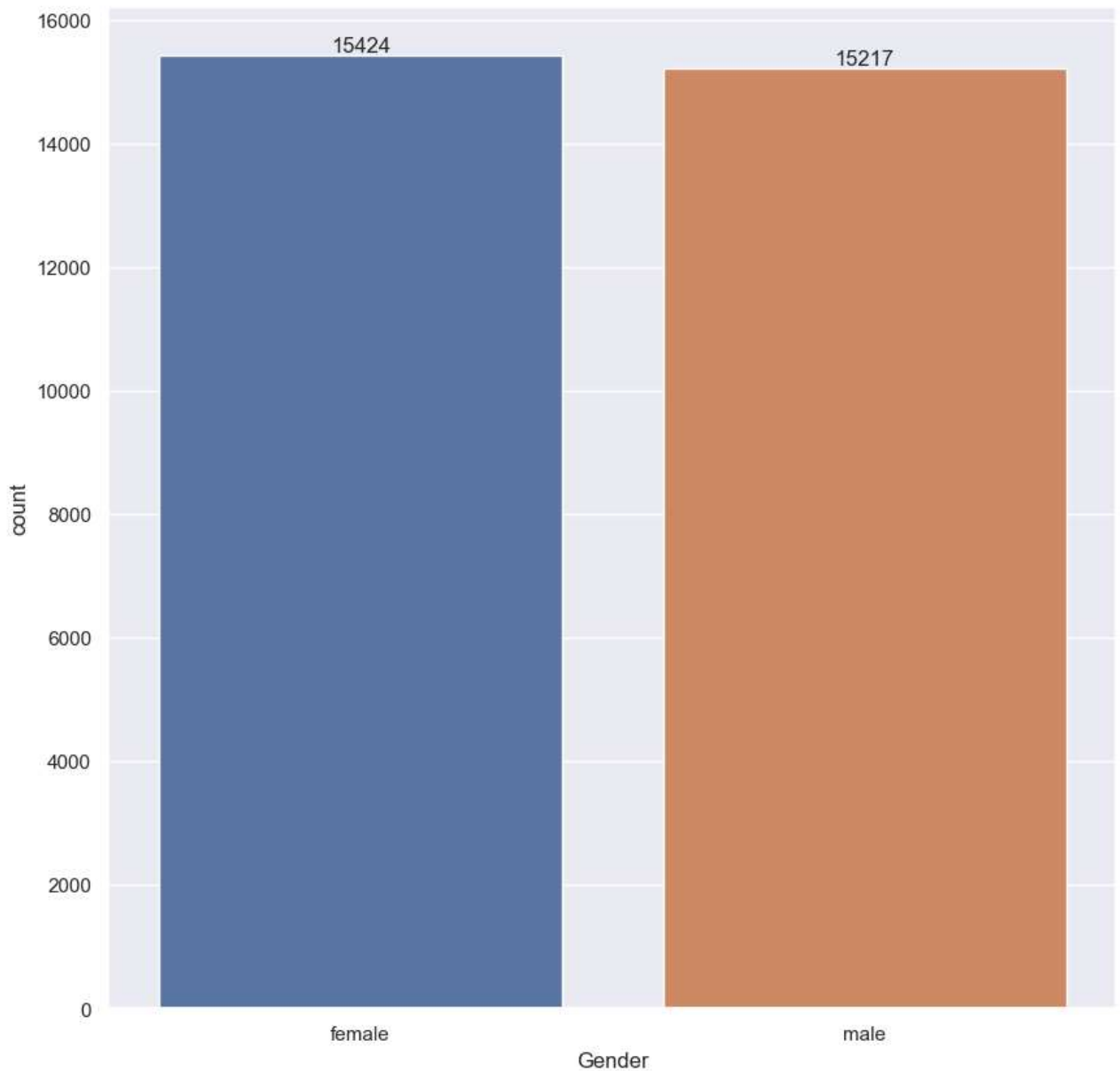
```
In [55]: df.head(5)
```

```
Out[55]:
```

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	IsFirstChild	NrSiblings	Trans
0	female	NaN	bachelor's degree	standard	none	married	regularly	yes	3.0	
1	female	group C	some college	standard	NaN	married	sometimes	yes	0.0	
2	female	group B	master's degree	standard	none	single	sometimes	yes	4.0	
3	male	group A	associate's degree	free/reduced	none	married	never	no	1.0	
4	male	group C	some college	standard	none	married	sometimes	yes	0.0	

gender distribution

```
In [152]: ax=sns.countplot(data=df,x='Gender')
sns.set(rc={"figure.figsize":(9,10)})
for bars in ax.containers:
    ax.bar_label(bars)
```



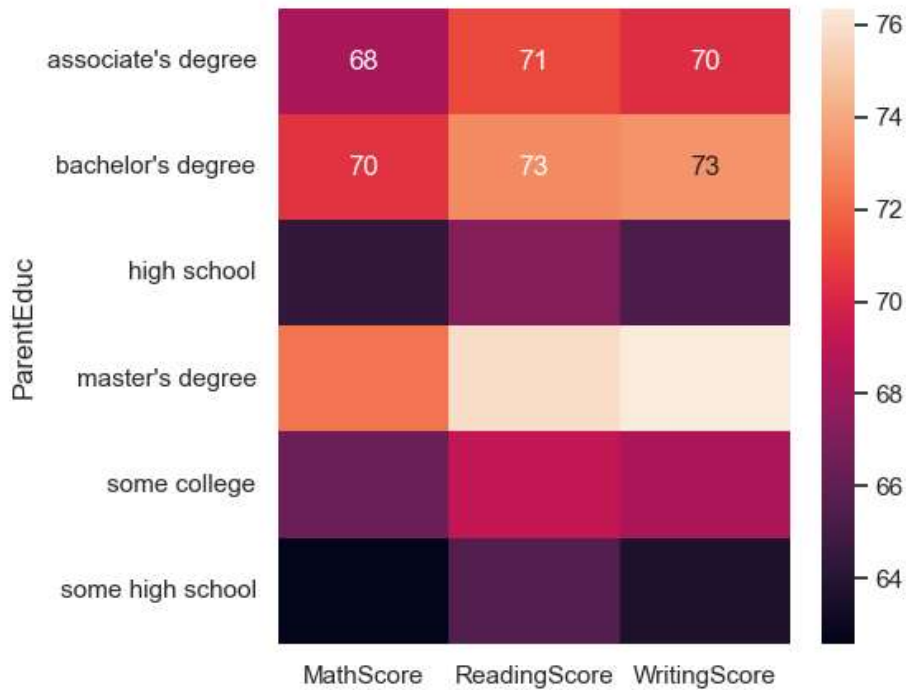
```
In [66]: # FOR Above chart we have analysed that number of female data more than male data    Gender
```

ParentEduc impact

```
In [99]: pe=df.groupby("ParentEduc").agg({"MathScore":'mean',"ReadingScore":'mean',"WritingScore":'mean'})
print(pe)
```

	MathScore	ReadingScore	WritingScore
ParentEduc			
associate's degree	68.365586	71.124324	70.299099
bachelor's degree	70.466627	73.062020	73.331069
high school	64.435731	67.213997	65.421136
master's degree	72.336134	75.832921	76.356896
some college	66.390472	69.179708	68.501432
some high school	62.584013	65.510785	63.632409

```
In [114]: sns.heatmap(pe , annot=True)
plt.pyplot.show()
sns.set(rc={"figure.figsize":(5,5)})
```



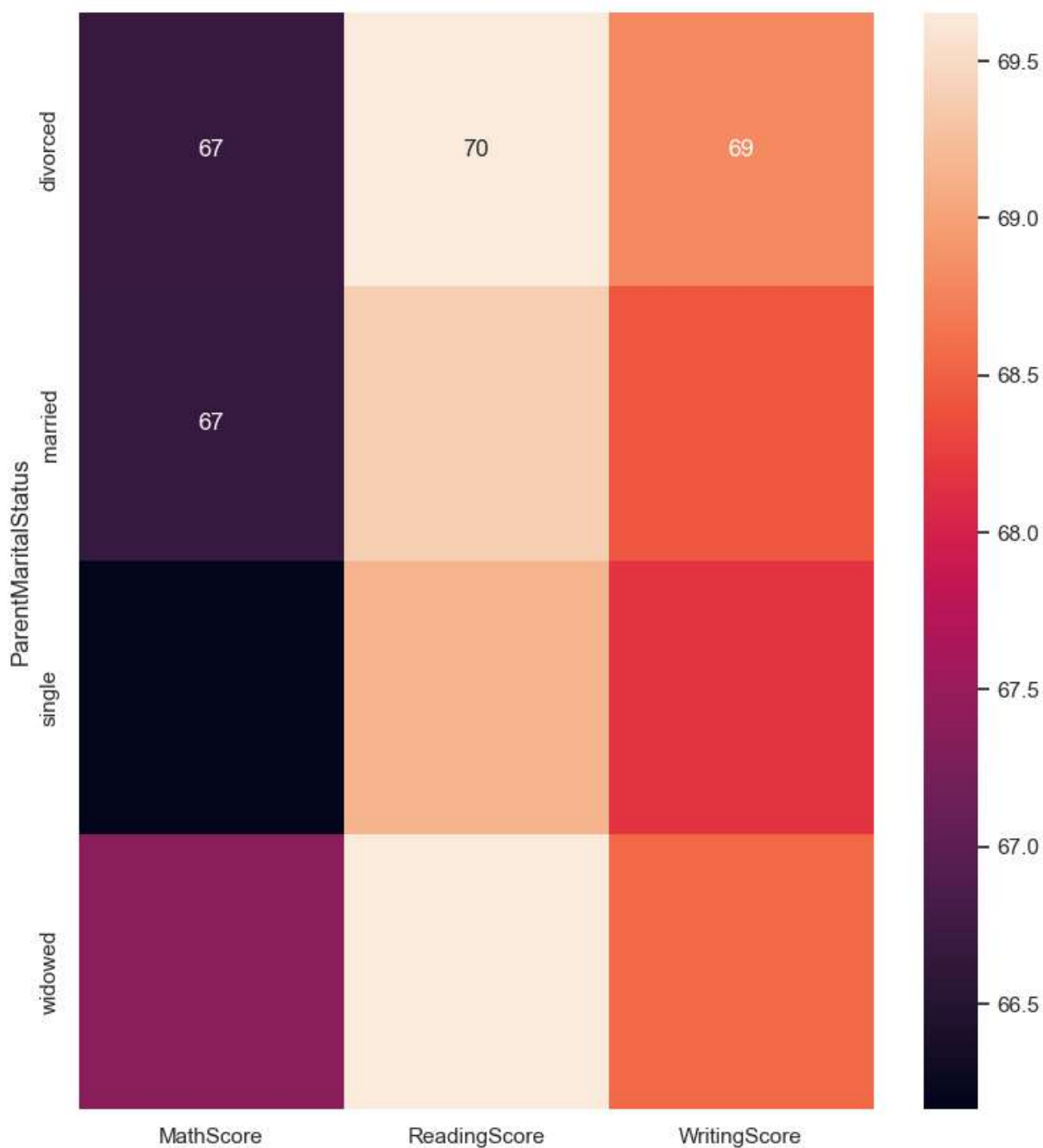
```
In [ ]: # FOR Above chart we have analysed that parent education is good impact for student studies.
```

ParentMaritalStatus¶

```
In [101]: pe1=df.groupby("ParentMaritalStatus").agg({"MathScore":'mean',"ReadingScore":'mean',"WritingScore":'m
print(pe1)
```

	MathScore	ReadingScore	WritingScore
ParentMaritalStatus			
divorced	66.691197	69.655011	68.799146
married	66.657326	69.389575	68.420981
single	66.165704	69.157250	68.174440
widowed	67.368866	69.651438	68.563452

```
In [154]: sns.heatmap(pe1, annot=True)
sns.set(rc={"figure.figsize":(5,5)})
```

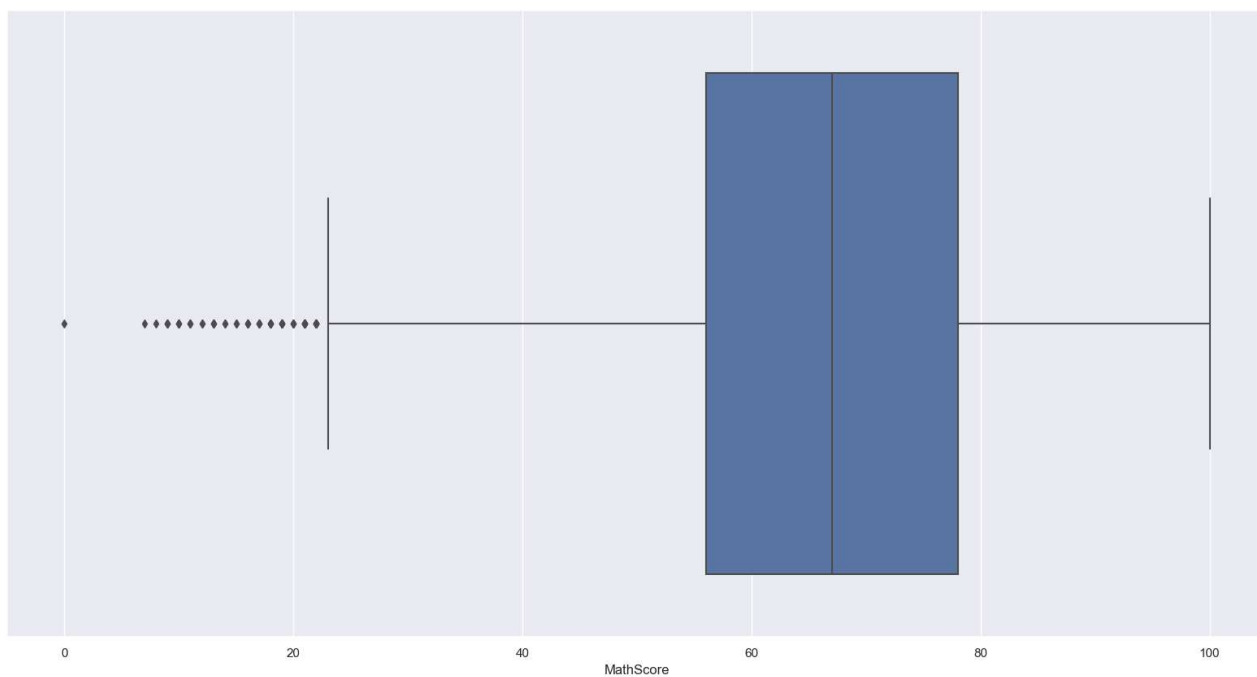


```
In [ ]: # FOR Above chart we have analysed that parent maritalstatus not that much impact for student studies
```

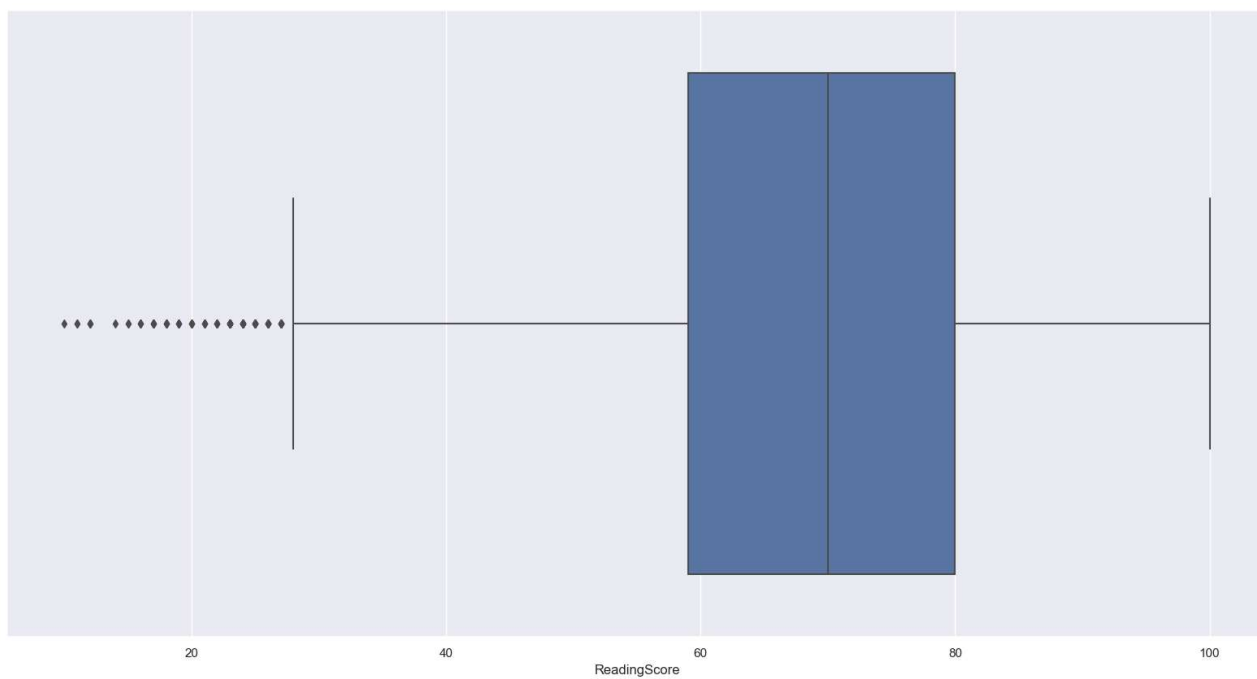
```
In [155]: pe2=df.groupby("PracticeSport").agg({"MathScore":'mean',"ReadingScore":'mean',"WritingScore":'mean'})
print(pe2)
```

	MathScore	ReadingScore	WritingScore
PracticeSport			
never	64.171079	68.337662	66.522727
regularly	67.839155	69.943019	69.604003
sometimes	66.274831	69.241307	68.072438

```
In [130]: sns.boxplot(data=df, x="MathScore")  
sns.set(rc={"figure.figsize":(20,10)})
```

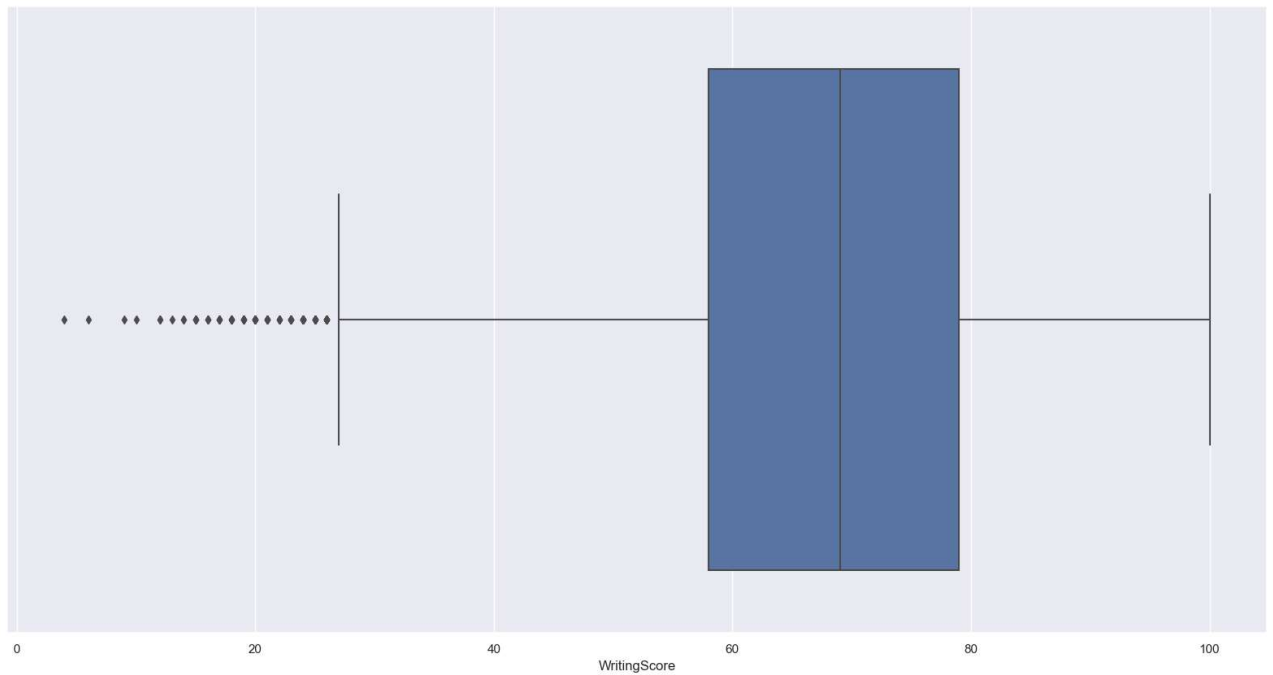


```
In [129]: sns.boxplot(data=df, x="ReadingScore")  
sns.set(rc={"figure.figsize":(20,10)})
```



```
In [128]: sns.boxplot(data=df,x="WritingScore")

sns.set(rc={"figure.figsize":(20,10)})
```



```
In [133]: df["EthnicGroup"].unique()
```

```
Out[133]: array([nan, 'group C', 'group B', 'group A', 'group D', 'group E'],
              dtype=object)
```

```
In [156]: groupA=df.loc[(df["EthnicGroup"] == 'group A')].count()
print(groupA)
```

```
Gender          2219
EthnicGroup     2219
ParentEduc      2078
LunchType       2219
TestPrep        2081
ParentMaritalStatus 2121
PracticeSport   2167
IsFirstChild    2168
NrSiblings      2096
TransportMeans   1999
WklyStudyHours  2146
MathScore       2219
ReadingScore    2219
WritingScore    2219
dtype: int64
```



```
In [181]: groupA=df.loc[(df["EthnicGroup"] == 'group A')].count()
groupB=df.loc[(df["EthnicGroup"] == 'group B')].count()
groupC=df.loc[(df["EthnicGroup"] == 'group C')].count()
groupD=df.loc[(df["EthnicGroup"] == 'group D')].count()
groupE=df.loc[(df["EthnicGroup"] == 'group E')].count()

A=['group A','group B','group C','group D','group E']

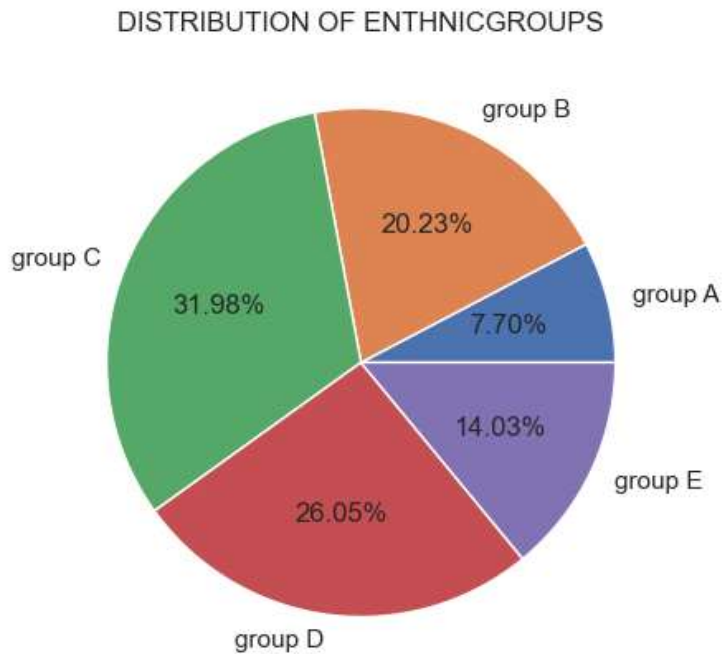
mylist=[groupA["EthnicGroup"],groupB["EthnicGroup"],groupC["EthnicGroup"],groupD["EthnicGroup"],groupE["EthnicGroup"]]

plt.pie(mylist, labels=A, autopct='%1.2f%%')

plt.title("DISTRIBUTION OF ENTHNICGROUPS")

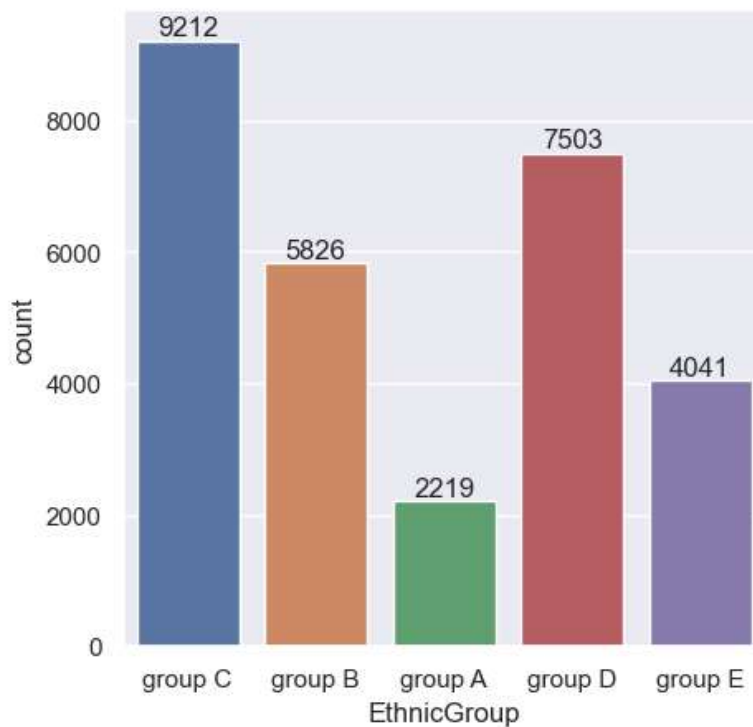
plt.figure(figsize=(10, 10))
```

Out[181]: <Figure size 1000x1000 with 0 Axes>



<Figure size 1000x1000 with 0 Axes>

```
In [179]: ax = sns.countplot(data=df, x="EthnicGroup")
for bars in ax.containers:
    ax.bar_label(bars)
```



```
In [180]: ax = sns.countplot(data=df, x="PracticeSport")
for bars in ax.containers:
    ax.bar_label(bars)
```

