```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df= pd.read csv("expanded data with more Features.csv")
df.head(10)
   Gender EthnicGroup
                                 ParentEduc
                                                 LunchType
                                                              TestPrep \
   female
                   NaN
                         bachelor's degree
                                                  standard
                                                                  none
1
   female
                               some college
                                                  standard
               group C
                                                                   NaN
2
   female
               group B
                           master's degree
                                                  standard
                                                                  none
3
     male
                        associate's degree
                                             free/reduced
               group A
                                                                  none
4
     male
               group C
                               some college
                                                  standard
                                                                  none
5
   female
               group B associate's degree
                                                  standard
                                                                  none
6
   female
                               some college
                                                  standard
               group B
                                                             completed
7
     male
               group B
                               some college
                                              free/reduced
                                                                  none
8
               group D
                                high school
     male
                                              free/reduced
                                                             completed
9
  female
               group B
                                high school
                                             free/reduced
                                                                  none
  ParentMaritalStatus PracticeSport IsFirstChild NrSiblings
TransportMeans
               married
                            regularly
                                                             3.0
                                                yes
school bus
1
               married
                            sometimes
                                                             0.0
                                                yes
NaN
                                                             4.0
                single
                            sometimes
                                                yes
school bus
3
               married
                                                             1.0
                                never
                                                 no
NaN
               married
                            sometimes
                                                             0.0
4
                                                yes
school bus
                            regularly
                                                             1.0
               married
                                                yes
school bus
               widowed
                                                             1.0
                                never
                                                 no
private
               married
                            sometimes
                                                             1.0
                                                yes
private
                single
                            sometimes
                                                 no
                                                             3.0
private
               married
                            regularly
                                                             NaN
                                                yes
private
  WklyStudyHours
                   MathScore
                               ReadingScore
                                             WritingScore
0
              < 5
                          71
                                         71
                                                        74
                                         90
1
          5 - 10
                          69
                                                        88
2
                          87
                                         93
              < 5
                                                        91
3
          5 - 10
                          45
                                         56
                                                        42
4
          5 - 10
                          76
                                         78
                                                        75
```

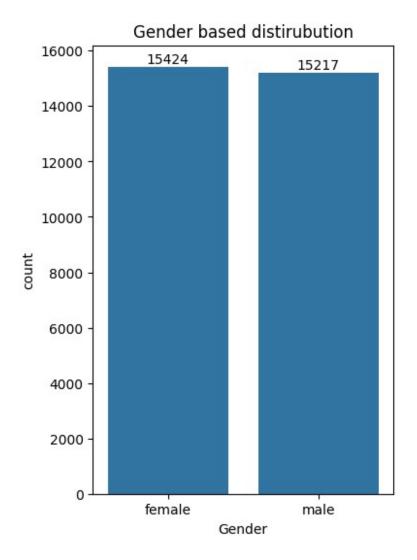
S - 10									
	5 6								
Section   Sect	7								
Unnamed: 0 Gender EthnicGroup ParentEduc unchType \ 10636 816 female group D high school tandard 10637 890 male group E high school tandard 10638 911 female NaN high school tree/reduced 10639 934 female group D associate's degree tandard 10640 960 male group B some college trandard 10637 none single sometimes no 10638 none single regularly no 10638 completed married sometimes no 1080638 completed married regularly no 1080638 completed married regularly no 1080639 completed married never no 1080640 none married never no 1080630 school_bus 5 - 10 59 61 105 10631 private 5 - 10 58 53 11 10639 school_bus 5 - 10 61 70 10639 school_bus 5 - 10 64 60 10640 school_bus 5 - 10 64 60 10640 school_bus 5 - 10 64 60 10640 school_bus 5 - 10 64 60	8 9								
Name	df.tail	.()							
10636				Gender I	EthnicGroup		Pare	entEduc	
10637	30636	•		female	group D		high	school	
934 female group D associate's degree trandard 9640 960 male group B some college trandard 10640 960 male group B some college trandard 1055blings \ 10636 none single sometimes no 10637 none single regularly no 10638 completed married sometimes no 10639 completed married regularly no 10639 completed married regularly no 10640 none married never no 10640 none married never no 1065 school_bus 5 - 10 59 61 15 16 16 170 170 170 170 170 170 170 170 170 170	30637		890	male	group E		high	school	
934	30638			female	NaN		high	school	
TransportMeans WklyStudyHours MathScore ReadingScore TransportMeans WklyStudyHours MathScore ReadingScore TransportMeans WklyStudyHours MathScore ReadingScore TransportMeans WklyStudyHours MathScore ReadingScore TransportMeans School_bus 5 - 10 58 53 TransportMeans WklyStudyHours MathScore ReadingScore TransportMeans WklyStudyHours MathScore Readin	30639			female	group D	associa	te's	degree	
TestPrep ParentMaritalStatus PracticeSport IsFirstChild 17Siblings \ 100636 none single sometimes no 100 10637 none single regularly no 100 10638 completed married sometimes no 100 10639 completed married regularly no 100 100 100 100 100 100 100 100 100 10	30640		960	male	group B	S	ome o	college	
10636   none   single   sometimes   no   10   10   10   10   10   10   10   1		Test	Prep P	arentMar	italStatus	PracticeS	port	IsFirst	Child
10   10   10   10   10   10   10   10		_	•				2		
10637   none   single   regularly   no   10638   completed   married   sometimes   no   10639   completed   married   regularly   no   10640   none   married   never   no   10640   none   married   never   no   10636   school_bus   5 - 10   59   61   10637   private   5 - 10   58   53   11   10638   private   5 - 10   61   70   10639   school_bus   5 - 10   82   90   13   10640   school_bus   5 - 10   64   60   88   88     10640   school_bus   5 - 10   64   60   88   88     10650   school_bus   5 - 10   64   60   60   88   80   80   80   80   80	2.0		none		single	somet	ımes		no
10   10   10   10   10   10   10   10	30637 1.0		none		single	regul	regularly		no
### TransportMeans WklyStudyHours MathScore ReadingScore ####################################	30638 1.0	completed			married	somet	sometimes		no
TransportMeans WklyStudyHours MathScore ReadingScore  0636 school_bus 5 - 10 59 61  0637 private 5 - 10 58 53  0638 private 5 - 10 61 70  07  0639 school_bus 5 - 10 82 90  03  0640 school_bus 5 - 10 64 60	30639 3.0	compl	eted		married	regul	regularly		no
### ### ##############################	30640 1.0		none		married	n	ever		no
80636       school_bus       5 - 10       59       61         80637       private       5 - 10       58       53         80638       private       5 - 10       61       70         80639       school_bus       5 - 10       82       90         83       school_bus       5 - 10       64       60         88				ns WklyS <sup>.</sup>	tudyHours I	MathScore	Rea	adingSco	re
90637     private     5 - 10     58     53       90638     private     5 - 10     61     70       97     90     90       93     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       98     90     90       99     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90       90     90     90    <	30636			us	5 - 10	59			61
01 00638 private 5 - 10 61 70 07 00639 school_bus 5 - 10 82 90 03 00640 school_bus 5 - 10 64 60	65 30637		priva	te	5 - 10	58			53
57 80639 school_bus 5 - 10 82 90 83 80640 school_bus 5 - 10 64 60	51								
80639 school_bus 5 - 10 82 90 93 80640 school_bus 5 - 10 64 60 88	30638 67		priva	те	5 - 10	61			70
80640 school_bus 5 - 10 64 60 8	30639 93	sc	hool_b	us	5 - 10	82			90
f.describe()	30640 58	sc	hool_b	us	5 - 10	64			60
	df.desc	ribe(	)						

```
NrSiblings
         Unnamed: 0
                                        MathScore
                                                   ReadingScore
WritingScore
count 30641.000000
                      29069.000000
                                     30641.000000
                                                   30641.000000
30641.000000
         499.556607
                          2.145894
                                        66.558402
                                                       69.377533
mean
68.418622
                          1,458242
                                        15.361616
                                                       14.758952
std
         288.747894
15.443525
                          0.000000
min
           0.000000
                                         0.000000
                                                       10.000000
4.000000
25%
         249.000000
                          1.000000
                                        56.000000
                                                       59.000000
58.000000
50%
         500.000000
                          2.000000
                                        67.000000
                                                       70.000000
69.000000
75%
         750.000000
                          3.000000
                                        78.000000
                                                       80.000000
79.000000
max
         999.000000
                          7.000000
                                       100.000000
                                                      100.000000
100.000000
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30641 entries, 0 to 30640
Data columns (total 15 columns):
#
     Column
                           Non-Null Count
                                            Dtype
- - -
 0
     Unnamed: 0
                           30641 non-null
                                            int64
 1
     Gender
                           30641 non-null
                                            object
 2
     EthnicGroup
                           28801 non-null
                                            object
     ParentEduc
 3
                           28796 non-null
                                            object
 4
     LunchType
                           30641 non-null
                                            object
 5
     TestPrep
                           28811 non-null
                                            object
 6
     ParentMaritalStatus
                           29451 non-null
                                            object
 7
     PracticeSport
                           30010 non-null
                                            object
 8
     IsFirstChild
                           29737 non-null
                                            object
 9
     NrSiblings
                           29069 non-null
                                            float64
    TransportMeans
 10
                           27507 non-null
                                            object
 11
     WklyStudyHours
                           29686 non-null
                                            object
     MathScore
                           30641 non-null
 12
                                            int64
 13
     ReadingScore
                           30641 non-null
                                            int64
14
     WritingScore
                           30641 non-null
                                            int64
dtypes: float64(1), int64(4), object(10)
memory usage: 3.5+ MB
df.isnull().sum()
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
# Drop unamed column
df = df.drop("Unnamed: 0", axis=1)
print(df.head())
   Gender EthnicGroup
                                ParentEduc
                                                LunchType TestPrep \
   female
                   NaN
                         bachelor's degree
                                                 standard
                                                               none
1
   female
                                                 standard
              group C
                              some college
                                                                NaN
2
              group B
   female
                           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
               \
              married
                           regularly
                                               yes
                                                            3.0
school bus
                           sometimes
                                                            0.0
1
              married
                                               yes
NaN
                                                            4.0
               single
                           sometimes
                                               yes
school bus
                                                            1.0
3
              married
                               never
                                                no
NaN
              married
                           sometimes
                                                            0.0
4
                                               yes
school bus
  WklyStudyHours
                  MathScore
                              ReadingScore
                                             WritingScore
0
             < 5
                          71
                                         71
                                                        74
          5 - 10
1
                          69
                                         90
                                                       88
2
             < 5
                          87
                                         93
                                                        91
3
          5 - 10
                          45
                                         56
                                                       42
4
          5 - 10
                          76
                                         78
                                                        75
```

# change wrong value (5 oct) in correct one(hours) in WklyStudyHours

```
df["WklyStudyHours"] = df["WklyStudyHours"].str.replace("05-0ct","5-
10")
df.head()
   Gender EthnicGroup
                                ParentEduc
                                                LunchType TestPrep
                                                 standard
0
   female
                   NaN
                         bachelor's degree
                                                               none
1
   female
              group C
                              some college
                                                 standard
                                                                NaN
2
   female
              group B
                           master's degree
                                                 standard
                                                               none
                        associate's degree
3
     male
                                             free/reduced
              group A
                                                               none
4
     male
              group C
                              some college
                                                 standard
                                                               none
  ParentMaritalStatus PracticeSport IsFirstChild
                                                    NrSiblings
TransportMeans
                           regularly
                                                            3.0
              married
                                               yes
school bus
                                                            0.0
1
              married
                           sometimes
                                               yes
NaN
               single
                           sometimes
                                                            4.0
                                               yes
school bus
3
              married
                                                            1.0
                               never
                                                no
NaN
                                                            0.0
4
              married
                           sometimes
                                               yes
school bus
  WklyStudyHours
                  MathScore
                              ReadingScore
                                             WritingScore
0
             < 5
                                                        74
                          71
                                         71
          5 - 10
                          69
                                         90
                                                        88
1
2
                                         93
             < 5
                          87
                                                        91
3
          5 - 10
                          45
                                         56
                                                        42
4
          5 - 10
                                                        75
                          76
                                         78
# Gender Distirbution
plt.figure(figsize = (4,6))
ax=sns.countplot(data = df,x="Gender")
plt.title("Gender based distirubution")
ax.bar label(ax.containers[0])
plt.show()
```

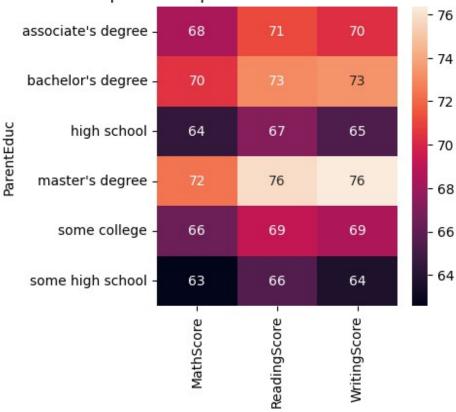


From the above chart we have found that females are more than male

```
af=
df.groupby("ParentEduc").agg({"MathScore": 'mean', "ReadingScore": 'mean'
, "WritingScore": 'mean'})
print(af)
                                ReadingScore WritingScore
                    MathScore
ParentEduc
associate's degree
                    68.365586
                                   71.124324
                                                  70.299099
bachelor's degree
                    70.466627
                                   73.062020
                                                  73.331069
high school
                                   67.213997
                                                  65.421136
                    64.435731
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
plt.figure(figsize = (4,4))
sns.heatmap(af,annot=True)
```

```
plt.title("Relationship between parents'education and student
education")
plt.show
<function matplotlib.pyplot.show(close=None, block=None)>
```

## Relationship between parents'education and student education

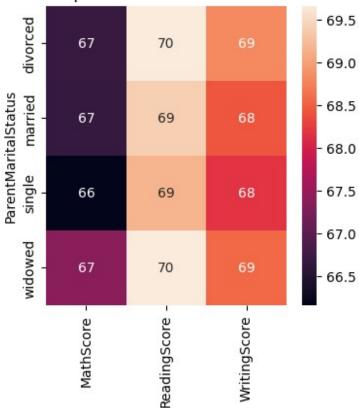


From the above chart we can see that parents'education have major imact on children'score

```
af1 =
df.groupby("ParentMaritalStatus").agg({"MathScore": 'mean', "ReadingScor
e":'mean',"WritingScore":'mean'})
print(af1)
                     MathScore ReadingScore WritingScore
ParentMaritalStatus
divorced
                     66.691197
                                                  68.799146
                                    69.655011
married
                     66.657326
                                    69.389575
                                                  68.420981
single
                     66.165704
                                   69.157250
                                                  68.174440
widowed
                     67.368866
                                   69.651438
                                                  68.563452
plt.figure(figsize = (4,4))
sns.heatmap(af1,annot = True)
plt.title("Relationship between parents'marital status and student
```

education")
plt.show()

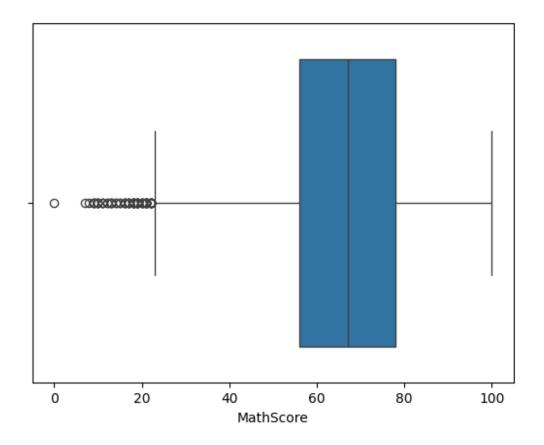
## Relationship between parents'marital status and student education



From above chart we can see that parent's martial has nothing to do with childern'score

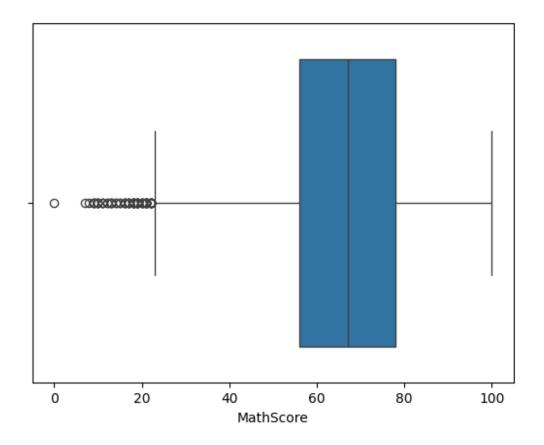
sns.boxplot(data =df,x= "MathScore")

<Axes: xlabel='MathScore'>



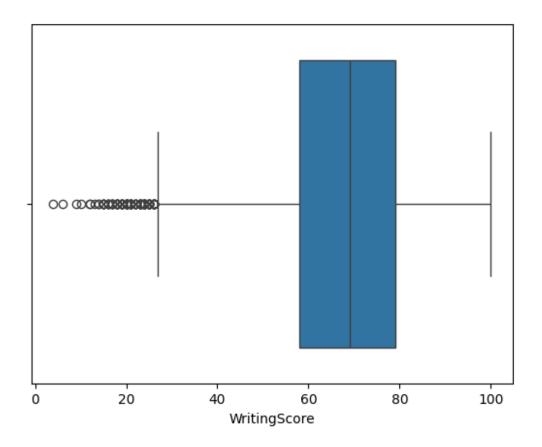
sns.boxplot(data =df,x= "ReadingScore")

<Axes: xlabel='MathScore'>



sns.boxplot(data =df,x= "WritingScore")

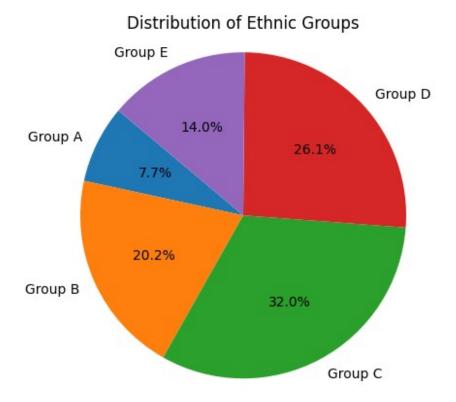
<Axes: xlabel='WritingScore'>



## disitribution of EthnicGroup

```
labels = ['Group A', 'Group B', 'Group C', 'Group D', 'Group E']
groupA = df[df['EthnicGroup'] == "group A"]['EthnicGroup'].count()
groupB = df[df['EthnicGroup'] == "group B"]['EthnicGroup'].count()
groupC = df[df['EthnicGroup'] == "group C"]['EthnicGroup'].count()
groupD = df[df['EthnicGroup'] == "group D"]['EthnicGroup'].count()
groupE = df[df['EthnicGroup'] == "group E"]['EthnicGroup'].count()
mylist = [groupA, groupB, groupC, groupD, groupE]

plt.pie(mylist, labels=labels, autopct='%1.1f%%', startangle=140)
plt.title("Distribution of Ethnic Groups")
plt.axis('equal') # Equal aspect ratio ensures pie is a circle.
plt.show()
```



The pie chart shows that Group C comprises 30% of the students, making it the largest ethnic group in the dataset, while Group E accounts for only 10%