student-results-analysis

August 26, 2024

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
data=pd.read_csv(r"C:\Preet\student_scores.csv")
```

[3]:
```
data.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 
 3   ParentEduc           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
 10  TransportMeans       27507 non-null  object 
 11  WklyStudyHours       29686 non-null  object 
 12  MathScore            30641 non-null  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
```

[4]:
```
data.head()
```

[4]:

	Unnamed: 0	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	\
0	0	female	NaN	bachelor's degree	standard	none	
1	1	female	group C	some college	standard	NaN	
2	2	female	group B	master's degree	standard	none	

```
3           3    male     group A  associate's degree  free/reduced     none
4           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         05-Oct         69            90            88
2            < 5         87            93            91
3         05-Oct         45            56            42
4         05-Oct         76            78            75
```

[5]:
```
data.describe()
```

[5]:
```
         Unnamed: 0    NrSiblings     MathScore  ReadingScore  WritingScore
count  30641.000000  29069.000000  30641.000000  30641.000000  30641.000000
mean     499.556607      2.145894     66.558402     69.377533     68.418622
std      288.747894      1.458242     15.361616     14.758952     15.443525
min        0.000000      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
```

[6]:
```
data.isnull().sum()
```

[6]:
```
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
```

1 Drop unamed columns

[7]:
```
data = data.drop("Unnamed: 0",axis =1)
```

[8]:
```
data.head()
```

[8]:
```
   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         05-Oct         69            90            88
2            < 5         87            93            91
3         05-Oct         45            56            42
4         05-Oct         76            78            75
```

2 change weekly study hours column

[9]:
```
data["WklyStudyHours"]=data["WklyStudyHours"].str.replace("05-Oct","5-10")
data.head()
```

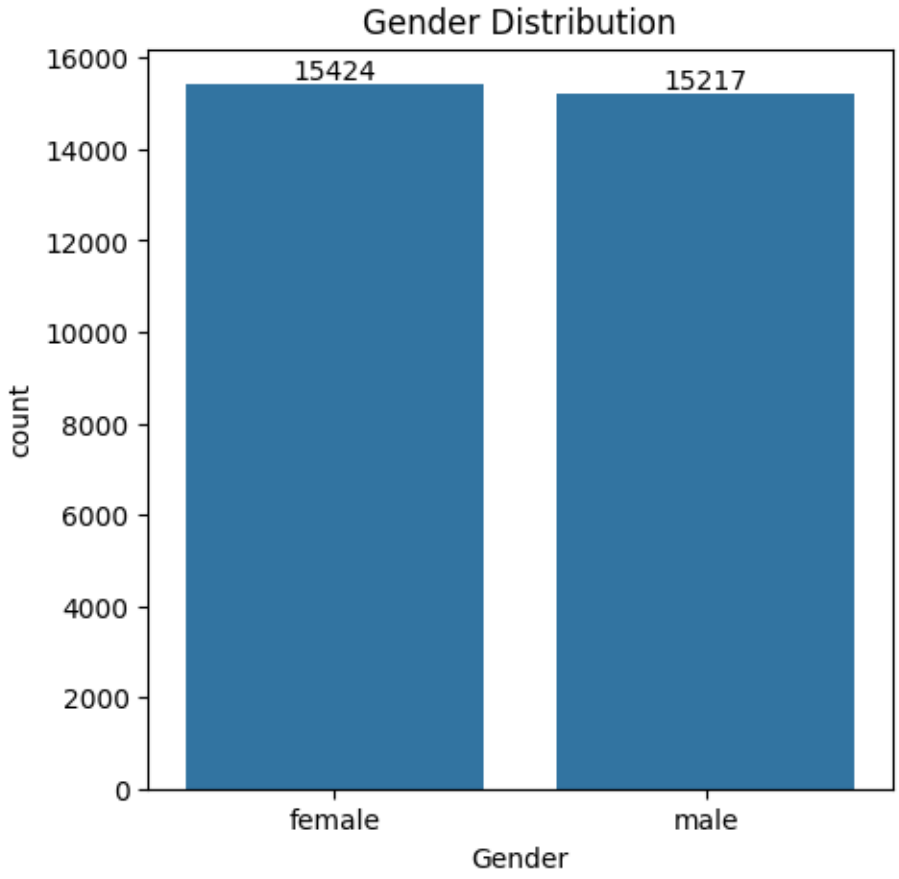
[9]:
```
   Gender EthnicGroup          ParentEduc     LunchType TestPrep  \
0  female         NaN   bachelor's degree      standard     none
1  female     group C        some college      standard      NaN
2  female     group B     master's degree      standard     none
3    male     group A  associate's 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

3 Gender Distribution

[10]:
```
plt.figure(figsize =(5,5))
ax=sns.countplot(data=data,x="Gender")
plt.title("Gender Distribution")
ax.bar_label(ax.containers[0])
plt.show()
```



4 From the above chart we analyzed that:

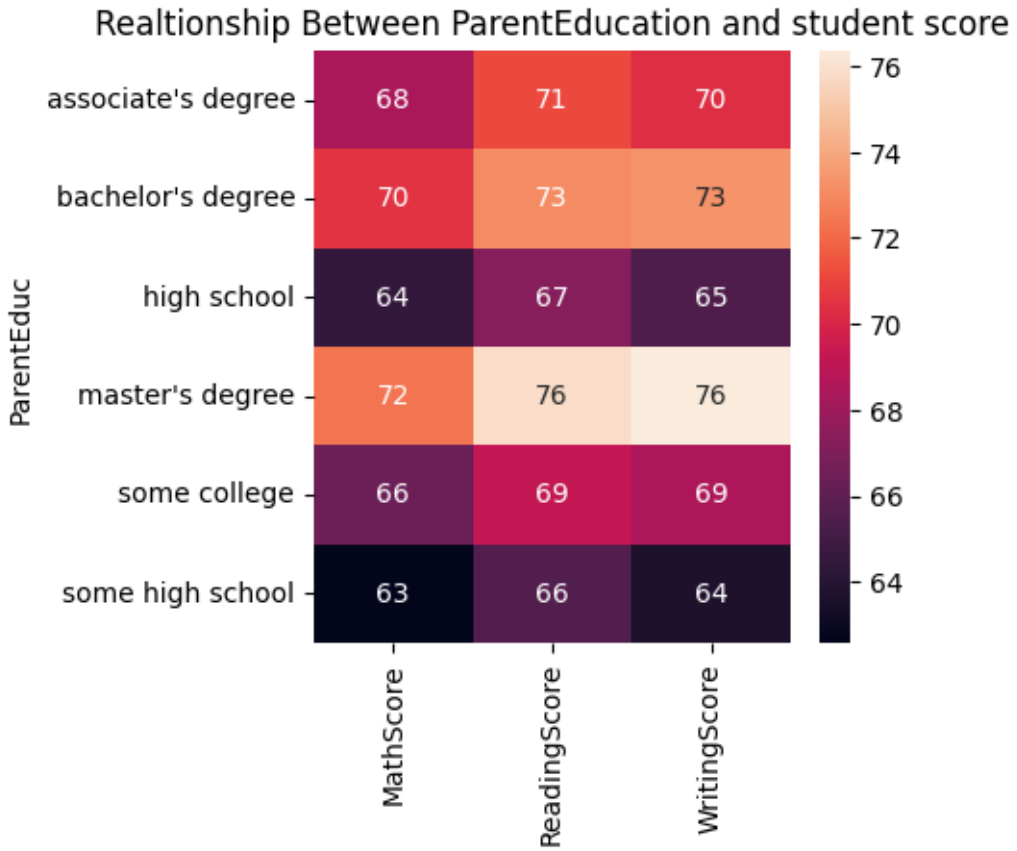
more females in the data is more than number of males

[11]:
```
gb =data.groupby("ParentEduc").agg({"MathScore":'mean',"ReadingScore":
  ↪'mean',"WritingScore":'mean'})
gb
```

[11]:

ParentEduc	MathScore	ReadingScore	WritingScore
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

[12]:
```
plt.figure(figsize =(4,4))
sns.heatmap(gb,annot=True)
plt.title("Realtionship Between ParentEducation and student score")
plt.show()
```



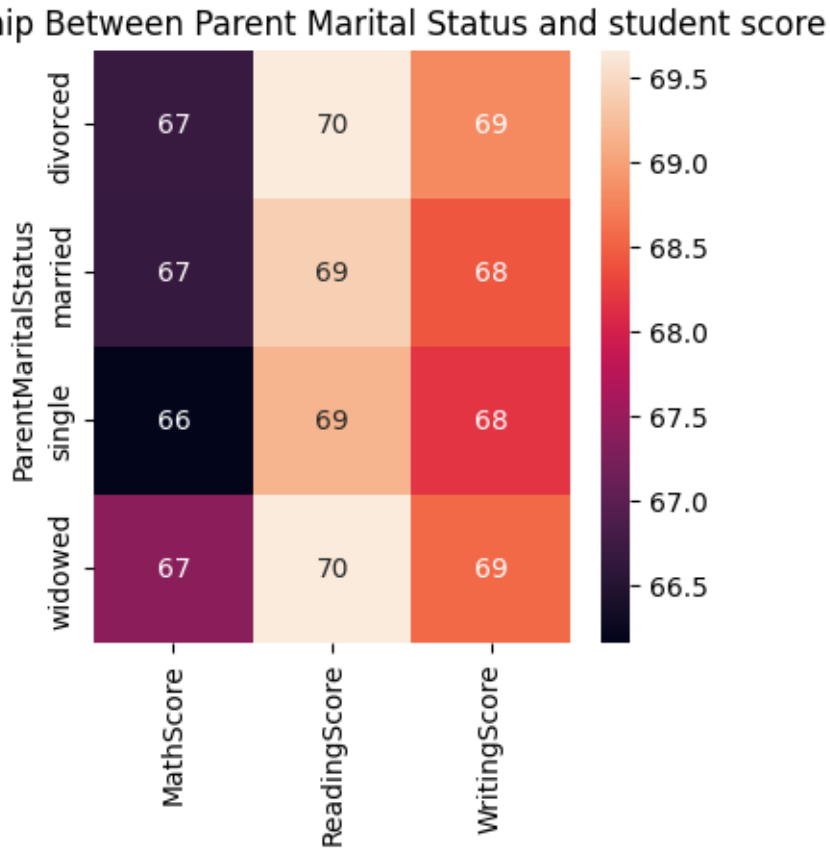
5 from the above heatmap parents' education has a direct relationship with student's education

[13]:
```
gb1 =data.groupby("ParentMaritalStatus").agg({"MathScore":'mean',"ReadingScore":
  ↪'mean',"WritingScore":'mean'})
gb1
```

[13]:

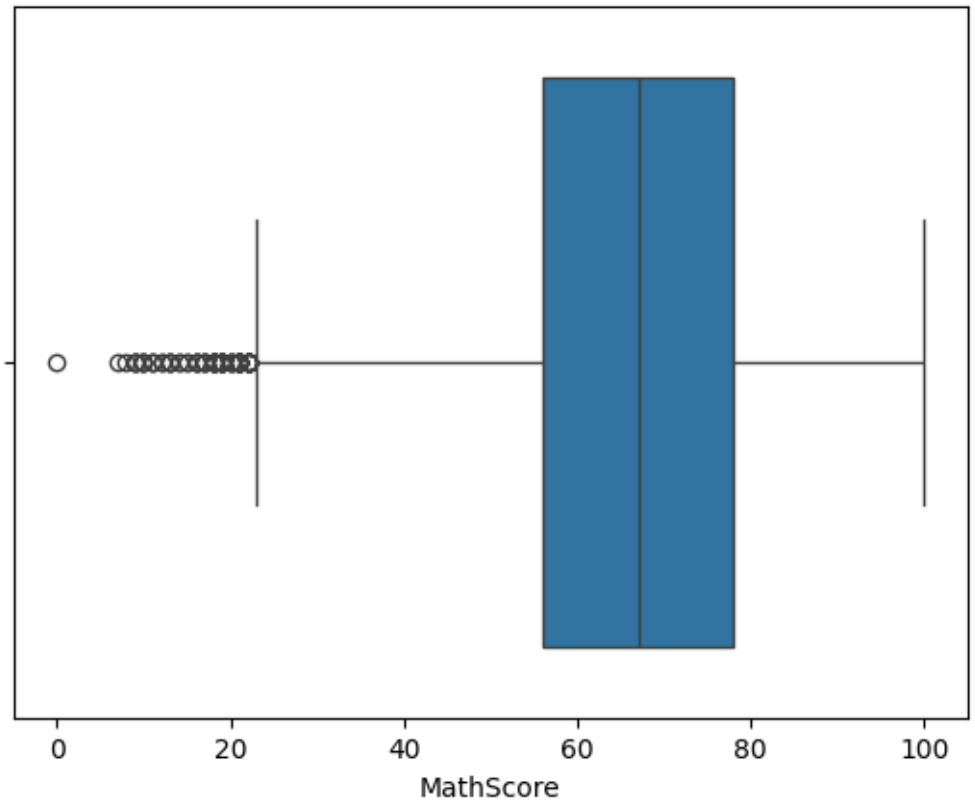
ParentMaritalStatus	MathScore	ReadingScore	WritingScore
divorced	66.691197	69.655011	68.799146
married	66.657326	69.389575	68.420981
single	66.165704	69.157250	68.174440
widowed	67.368866	69.651438	68.563452

[14]:
```
plt.figure(figsize =(4,4))
sns.heatmap(gb1,annot=True)
plt.title("Realtionship Between Parent Marital Status and student score")
plt.show()
```

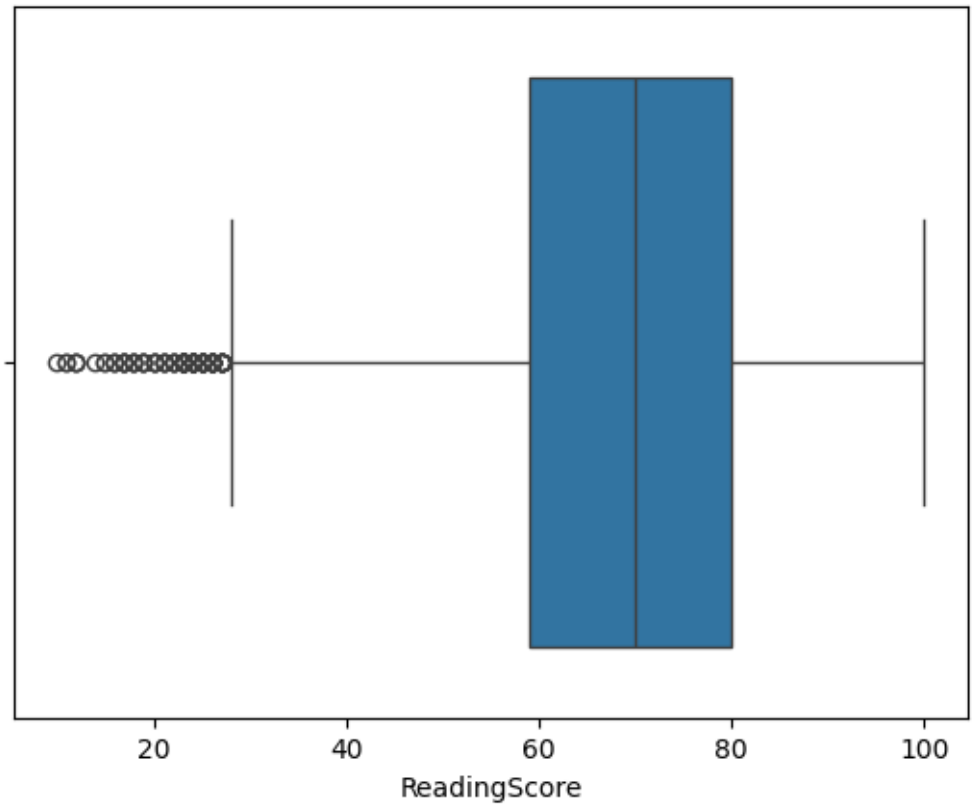


6 From the above heatmap analysis parent marital status does not have an impact on student education

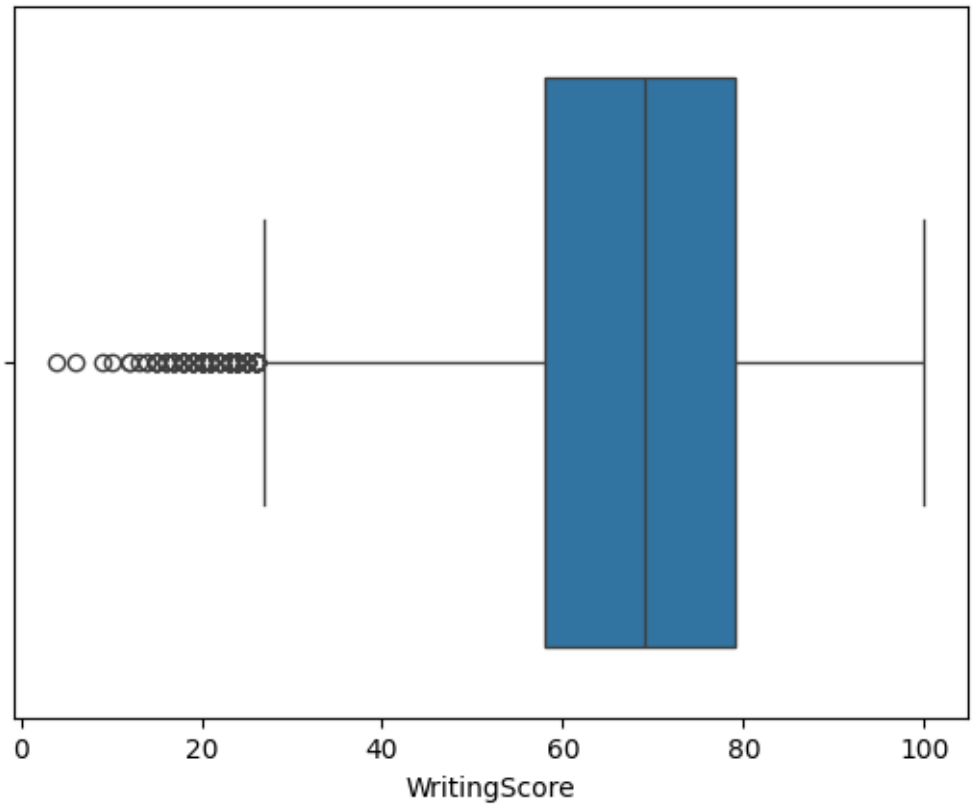
[15]:
```
sns.boxplot(data=data,x="MathScore")
plt.show()
```



[16]:
```
sns.boxplot(data=data,x="ReadingScore")
plt.show()
```



[17]:
```
sns.boxplot(data=data,x="WritingScore")
plt.show()
```

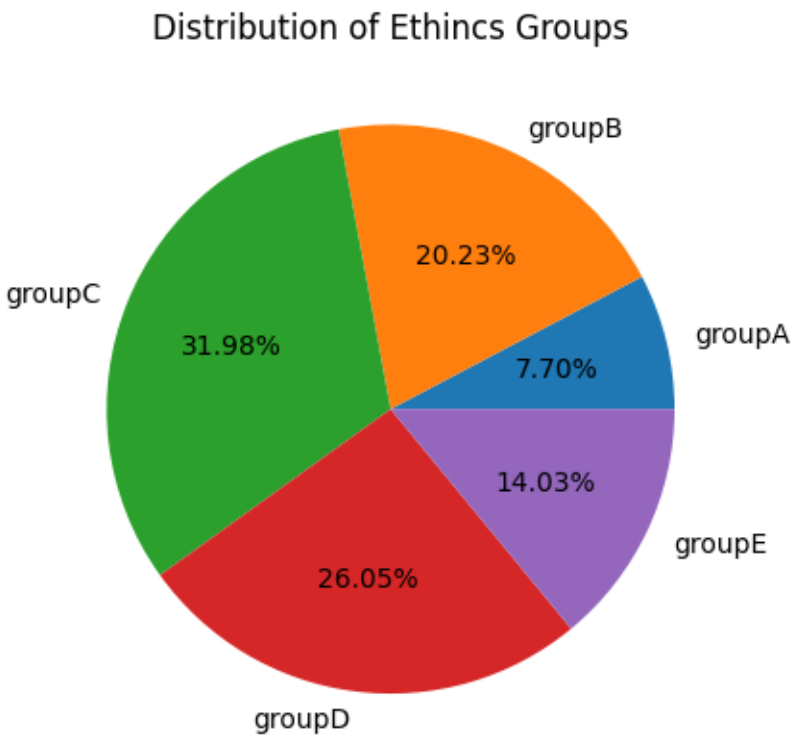


[18]:
```
print(data["EthnicGroup"].unique())
```

```
[nan 'group C' 'group B' 'group A' 'group D' 'group E']
```

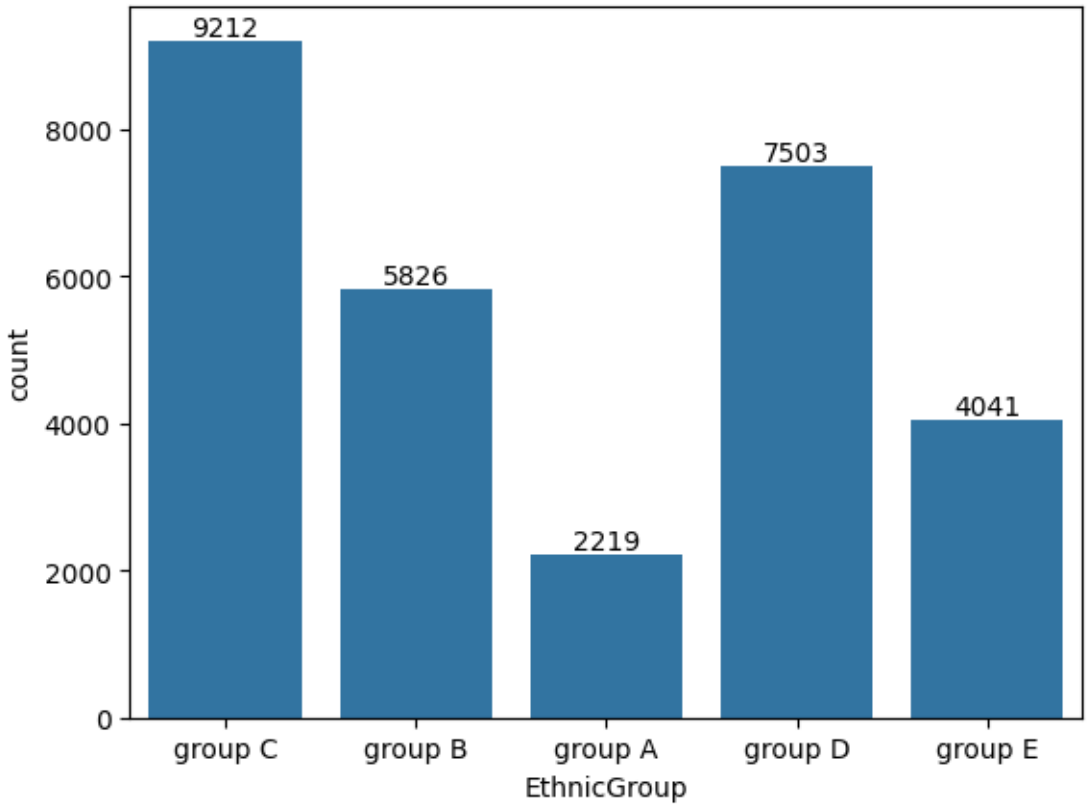
7 Distribution of Ethinic Group

[29]:
```
groupA=data.loc[(data["EthnicGroup"] =="group A")].count()
groupB=data.loc[(data["EthnicGroup"] =="group B")].count()
groupC=data.loc[(data["EthnicGroup"] =="group C")].count()
groupD=data.loc[(data["EthnicGroup"] =="group D")].count()
groupE=data.loc[(data["EthnicGroup"] =="group E")].count()
mylist = [groupA["EthnicGroup"], groupB["EthnicGroup"], groupC["EthnicGroup"],␣
 ↪groupD["EthnicGroup"], groupE["EthnicGroup"]]
l=["groupA","groupB","groupC","groupD","groupE"]
plt.pie(mylist,labels=l,autopct="%1.2f%%")
plt.title("Distribution of Ethincs Groups")
plt.show()
```



[30]:
```
ax= sns.countplot(data=data,x='EthnicGroup')
ax.bar_label(ax.containers[0])
```

[30]:
```
[Text(0, 0, '9212'),
 Text(0, 0, '5826'),
 Text(0, 0, '2219'),
 Text(0, 0, '7503'),
 Text(0, 0, '4041')]
```



[22]:

```
Gender                  5826
EthnicGroup             5826
ParentEduc              5470
LunchType               5826
TestPrep                5488
ParentMaritalStatus     5605
PracticeSport           5704
IsFirstChild            5649
NrSiblings              5546
TransportMeans          5238
WklyStudyHours          5642
MathScore               5826
ReadingScore            5826
WritingScore            5826
dtype: int64
```

[33]:

```
Gender                  9212
EthnicGroup             9212
```

```
ParentEduc             8652
LunchType              9212
TestPrep               8652
ParentMaritalStatus    8858
PracticeSport          9050
IsFirstChild           8929
NrSiblings             8763
TransportMeans         8280
WklyStudyHours         8933
MathScore              9212
ReadingScore           9212
WritingScore           9212
dtype: int64
```

[34]:

```
Gender                 7503
EthnicGroup            7503
ParentEduc             7056
LunchType              7503
TestPrep               7070
ParentMaritalStatus    7218
PracticeSport          7343
IsFirstChild           7285
NrSiblings             7106
TransportMeans         6713
WklyStudyHours         7270
MathScore              7503
ReadingScore           7503
WritingScore           7503
dtype: int64
```

[ ]:

[ ]:

[ ]: