

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [7]: df=pd.read_csv(r'C:\Users\lenovo\Downloads\data for python 2nd project.csv')
print(df.head())
```

	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	5 - 10	69	90	88
2	< 5	87	93	91
3	5 - 10	45	56	42
4	5 - 10	76	78	75

```
In [8]: df.describe()
```

```
Out[8]:
```

	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

```
In [9]: 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
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
```

```
In [10]: df.isnull().sum()
```

```
Out[10]: 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 [11]: df=df.drop("Unnamed: 0",axis=1)
print(df.head)
```

```

<bound method NDFrame.head of
nchType  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
...      ...      ...      ...      ...      ...
30636  female  group D    high school  standard  none
30637  male    group E    high school  standard  none
30638  female      NaN    high school  free/reduced  completed
30639  female  group D  associate's degree  standard  completed
30640  male    group B    some college  standard  none

      ParentMaritalStatus  PracticeSport  IsFirstChild  NrSiblings  \
0      married      regularly      yes      3.0
1      married      sometimes      yes      0.0
2      single      sometimes      yes      4.0
3      married      never      no      1.0
4      married      sometimes      yes      0.0
...      ...      ...      ...      ...
30636      single      sometimes      no      2.0
30637      single      regularly      no      1.0
30638      married      sometimes      no      1.0
30639      married      regularly      no      3.0
30640      married      never      no      1.0

      TransportMeans  WklyStudyHours  MathScore  ReadingScore  WritingScore
0      school_bus      < 5      71      71      74
1      NaN      5 - 10      69      90      88
2      school_bus      < 5      87      93      91
3      NaN      5 - 10      45      56      42
4      school_bus      5 - 10      76      78      75
...      ...      ...      ...      ...      ...
30636  school_bus      5 - 10      59      61      65
30637  private      5 - 10      58      53      51
30638  private      5 - 10      61      70      67
30639  school_bus      5 - 10      82      90      93
30640  school_bus      5 - 10      64      60      58

```

[30641 rows x 14 columns]>

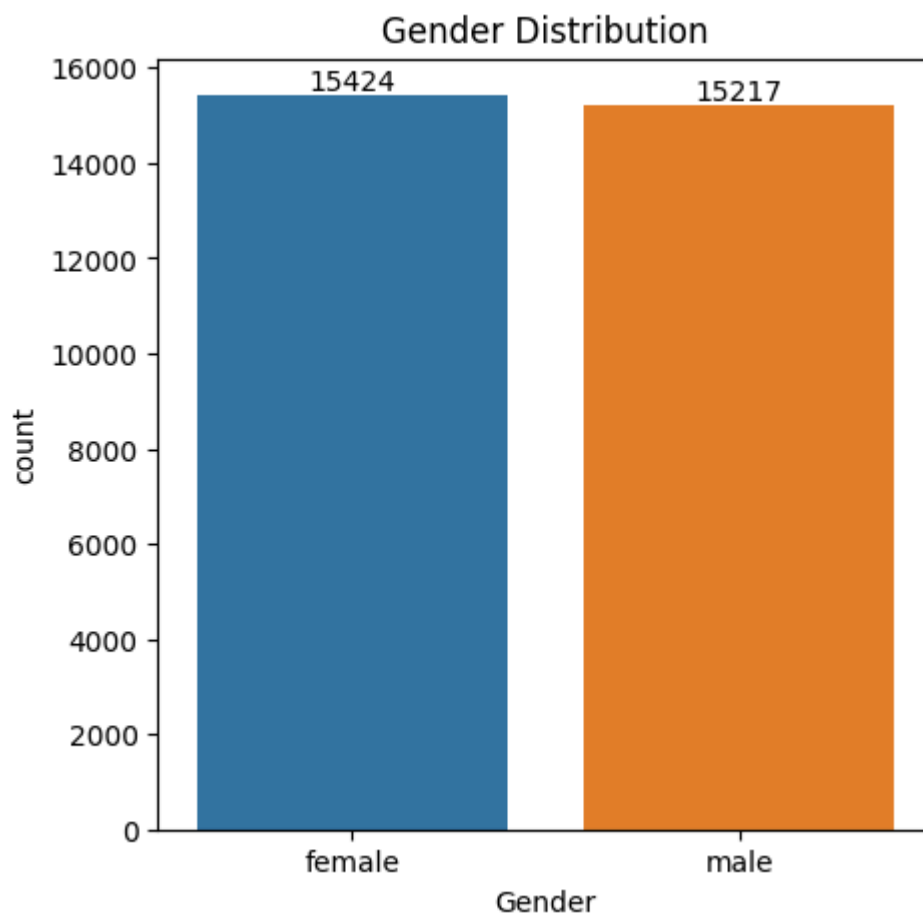
In [12]: `df.head()`

Out[12]:

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	Is
0	female	NaN	bachelor's degree	standard	none	married	regularly	
1	female	group C	some college	standard	NaN	married	sometimes	
2	female	group B	master's degree	standard	none	single	sometimes	
3	male	group A	associate's degree	free/reduced	none	married	never	
4	male	group C	some college	standard	none	married	sometimes	

GENDER DISTRIBUTION

```
In [35]: plt.figure(figsize=(5,5))
ax=sns.countplot(data=df,x="Gender",hue="Gender")
for bars in ax.containers:
    ax.bar_label(bars)
plt.title("Gender Distribution")
plt.show()
```



From above chart we can analyse that female in school is more than male student

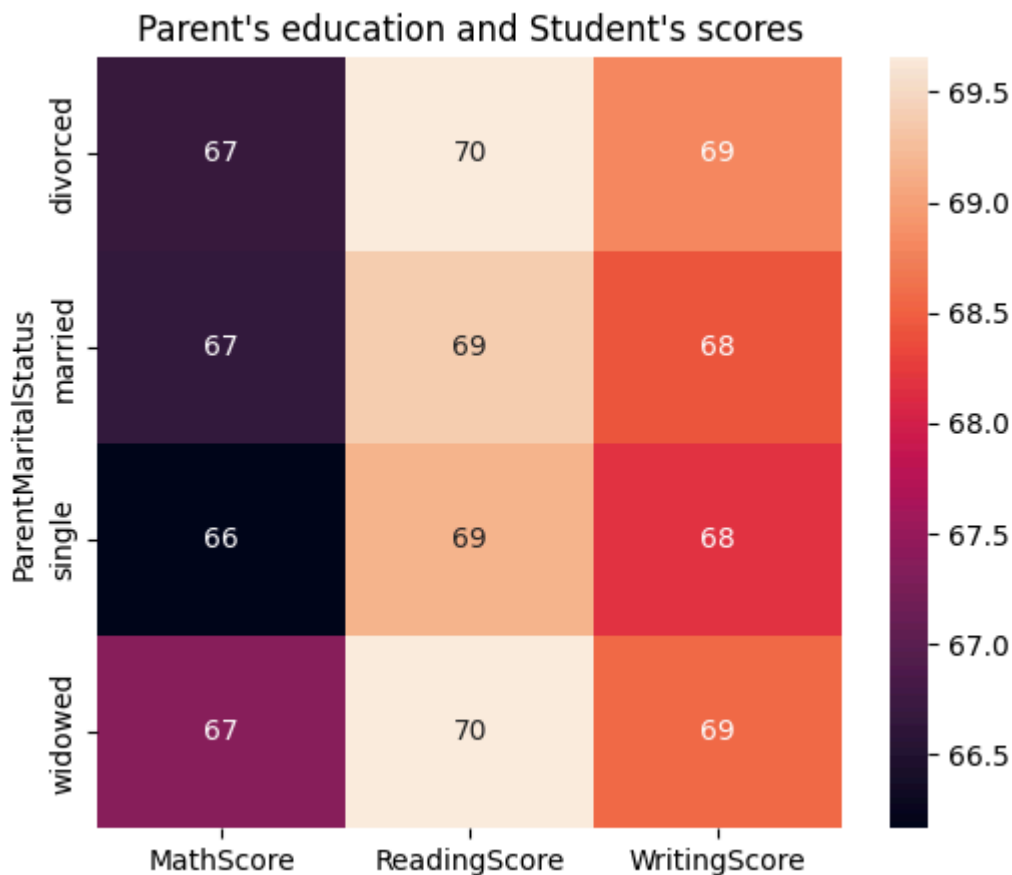
PARENTs EDUCATION AND STUDENT SCORES

```
In [28]: gb=df.groupby("ParentEduc").agg({"MathScore":'mean',"ReadingScore":'mean',"WritingScore":'mean'})
print(gb)
```

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

<Figure size 600x600 with 0 Axes>

```
In [38]: plt.figure(figsize=(6,5))
sns.heatmap(gb,annot=True)
plt.title("Parent's education and Student's scores")
plt.show()
```



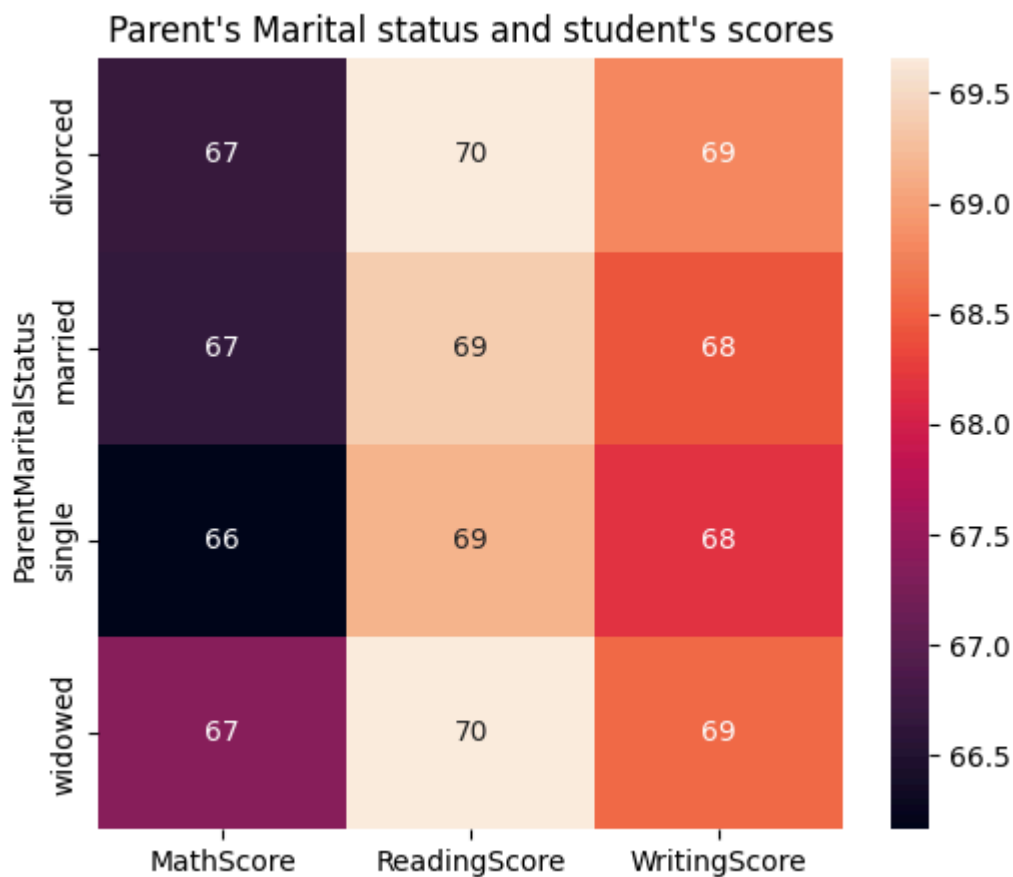
From the above chart we have concluded that impact of parent's education is directly proportional to student scores

PARENT MARITAL STATUS AND STUDENTS SCORES

```
In [33]: gm=df.groupby("ParentMaritalStatus").agg({"MathScore":'mean',"ReadingScore":'mean',
print(gm)
```

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

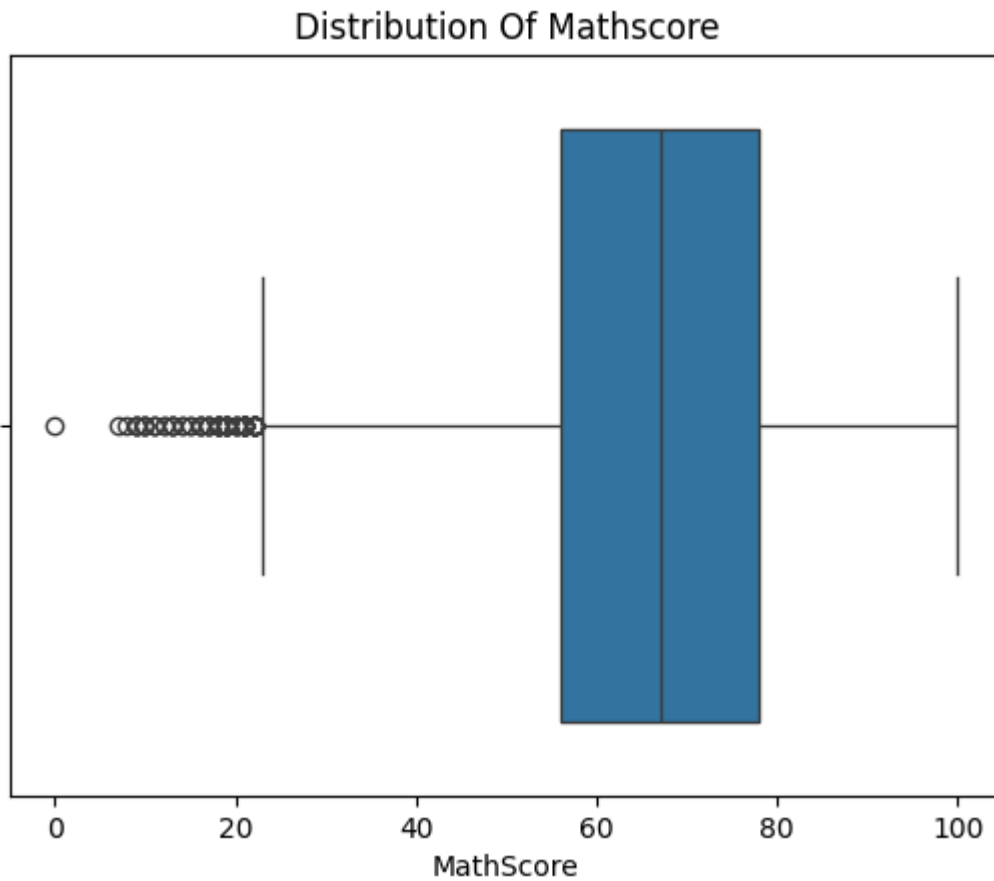
```
In [39]: plt.figure(figsize=(6,5))
sns.heatmap(gm,annot=True)
plt.title("Parent's Marital status and student's scores")
plt.show()
```



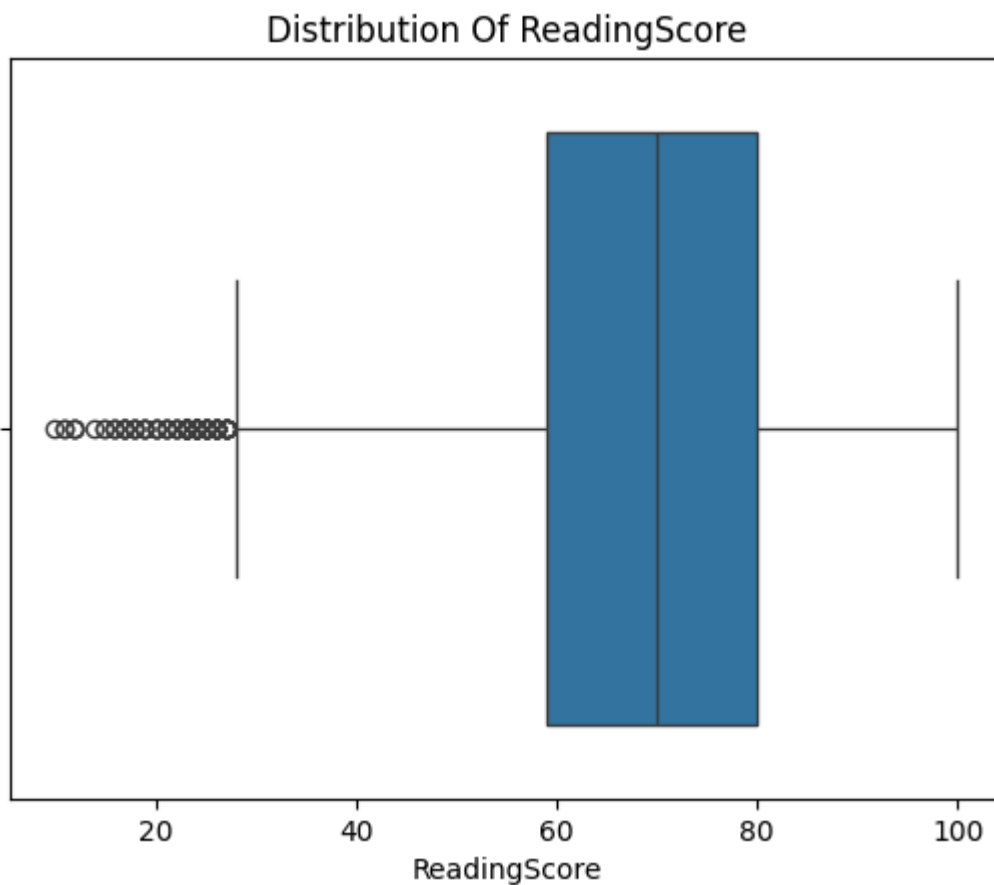
From the above chart, we can conclude that the impact of parents' marital status on student scores is negligible.

SCORE DISTRIBUTION

```
In [41]: sns.boxplot(data=df, x="MathScore")
plt.title("Distribution Of Maths score")
plt.show()
```

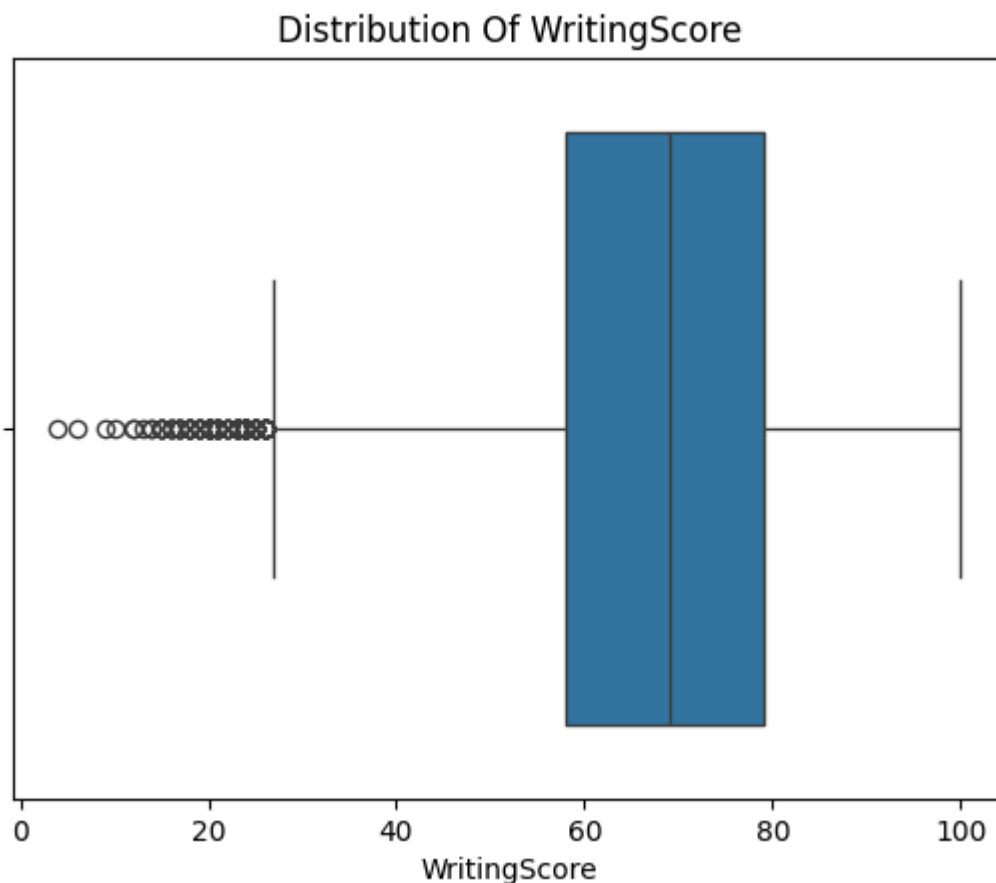


```
In [42]: sns.boxplot(data=df,x="ReadingScore")  
plt.title("Distribution Of ReadingScore")  
plt.show()
```



```
In [43]: sns.boxplot(data=df,x="WritingScore")  
plt.title("Distribution Of WritingScore")
```

```
plt.show()
```



From above bars we can conclude that Maths is the hardest as its outlier range lie between 0-20 whereas for others it is 20-40

DISTRIBUTION OF ETHNIC GROUPS

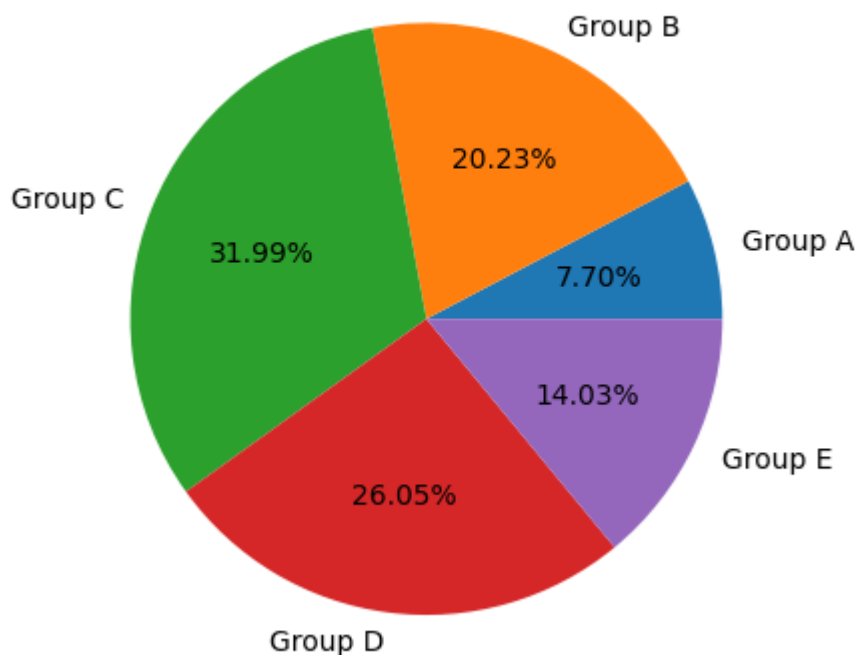
```
In [44]: print(df["EthnicGroup"].unique())
```

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

```
In [62]: 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()

l=["Group A", "Group B", "Group C", "Group D", "Group E"]
mlist=[groupA["EthnicGroup"],groupB["EthnicGroup"],groupC["EthnicGroup"],groupD["EthnicGroup"],groupE["EthnicGroup"]]
plt.pie(mlist, labels=l,autopct="%1.2f%%")
plt.title("Distribution Of Ethnic Groups")
plt.show()
```


Distribution Of Ethnic Groups



From above pie chart we can say that Ethnic group C is the primary contributor, with group D and group B being second and third, respectively.

Conclusion:

The student demographic data reveals that female students outnumber their male counterparts. Mathematics emerges as the most challenging subject for students in terms of securing high marks. Academic performance is notably influenced by the educational background of the parents, while the marital status of the parents has a negligible impact. Additionally, the majority of students belong to Group C, followed by Group D, with Group A having the smallest number of students."

In []: