In [2]:
```
#pip install numpy
#pip install pandas
#pip install matplotlib
#pip intsall seaborn
```

In [4]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [19]:
```
df = pd.read_csv(r'C:\Users\Kiran Singh\Downloads\Student_performance_data _.csv')
print(df.head())
```

```
   StudentID  Age  Gender  Ethnicity  ParentalEducation  StudyTimeWeekly  \
0       1001   17       1          0                  2        19.833723
1       1002   18       0          0                  1        15.408756
2       1003   15       0          2                  3         4.210570
3       1004   17       1          0                  3        10.028829
4       1005   17       1          0                  2         4.672495

   Absences  Tutoring  ParentalSupport  Extracurricular  Sports  Music  \
0         7         1                2                0       0      1
1         0         0                1                0       0      0
2        26         0                2                0       0      0
3        14         0                3                1       0      0
4        17         1                3                0       0      0

   Volunteering       GPA  GradeClass
0             0  2.929196         2.0
1             0  3.042915         1.0
2             0  0.112602         4.0
3             0  2.054218         3.0
4             0  1.288061         4.0
```

In [20]:
```
df.describe()
```

Out[20]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	A
count	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392
mean	2196.500000	16.468645	0.510870	0.877508	1.746237	9.771992	14
std	690.655244	1.123798	0.499986	1.028476	1.000411	5.652774	8
min	1001.000000	15.000000	0.000000	0.000000	0.000000	0.001057	0
25%	1598.750000	15.000000	0.000000	0.000000	1.000000	5.043079	7
50%	2196.500000	16.000000	1.000000	0.000000	2.000000	9.705363	15
75%	2794.250000	17.000000	1.000000	2.000000	2.000000	14.408410	22
max	3392.000000	18.000000	1.000000	3.000000	4.000000	19.978094	29

In [21]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2392 entries, 0 to 2391
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   StudentID          2392 non-null   int64  
 1   Age                2392 non-null   int64  
 2   Gender             2392 non-null   int64  
 3   Ethnicity          2392 non-null   int64  
 4   ParentalEducation  2392 non-null   int64  
 5   StudyTimeWeekly    2392 non-null   float64
 6   Absences           2392 non-null   int64  
 7   Tutoring           2392 non-null   int64  
 8   ParentalSupport    2392 non-null   int64  
 9   Extracurricular    2392 non-null   int64  
 10  Sports             2392 non-null   int64  
 11  Music              2392 non-null   int64  
 12  Volunteering       2392 non-null   int64  
 13  GPA                2392 non-null   float64
 14  GradeClass         2392 non-null   float64
dtypes: float64(3), int64(12)
memory usage: 280.4 KB
```

In [22]:
```
df.isnull().sum()
```

Out[22]:
```
StudentID            0
Age                  0
Gender               0
Ethnicity            0
ParentalEducation    0
StudyTimeWeekly      0
Absences             0
Tutoring             0
ParentalSupport      0
Extracurricular      0
Sports               0
Music                0
Volunteering         0
GPA                  0
GradeClass           0
dtype: int64
```

Drop Ethnicity column

In [26]:
```
df = df.drop("Ethnicity", axis = 1)
print(df.head())
```

```
   Age  Gender  ParentalEducation  StudyTimeWeekly  Absences  Tutoring  \
0   17       1                  2        19.833723         7         1
1   18       0                  1        15.408756         0         0
2   15       0                  3         4.210570        26         0
3   17       1                  3        10.028829        14         0
4   17       1                  2         4.672495        17         1

   ParentalSupport  Extracurricular  Sports  Music  Volunteering       GPA  \
0                2                0       0      1             0  2.929196
1                1                0       0      0             0  3.042915
2                2                0       0      0             0  0.112602
3                3                1       0      0             0  2.054218
4                3                0       0      0             0  1.288061

   GradeClass
0         2.0
1         1.0
2         4.0
3         3.0
4         4.0
```

Change Study time Weekly column

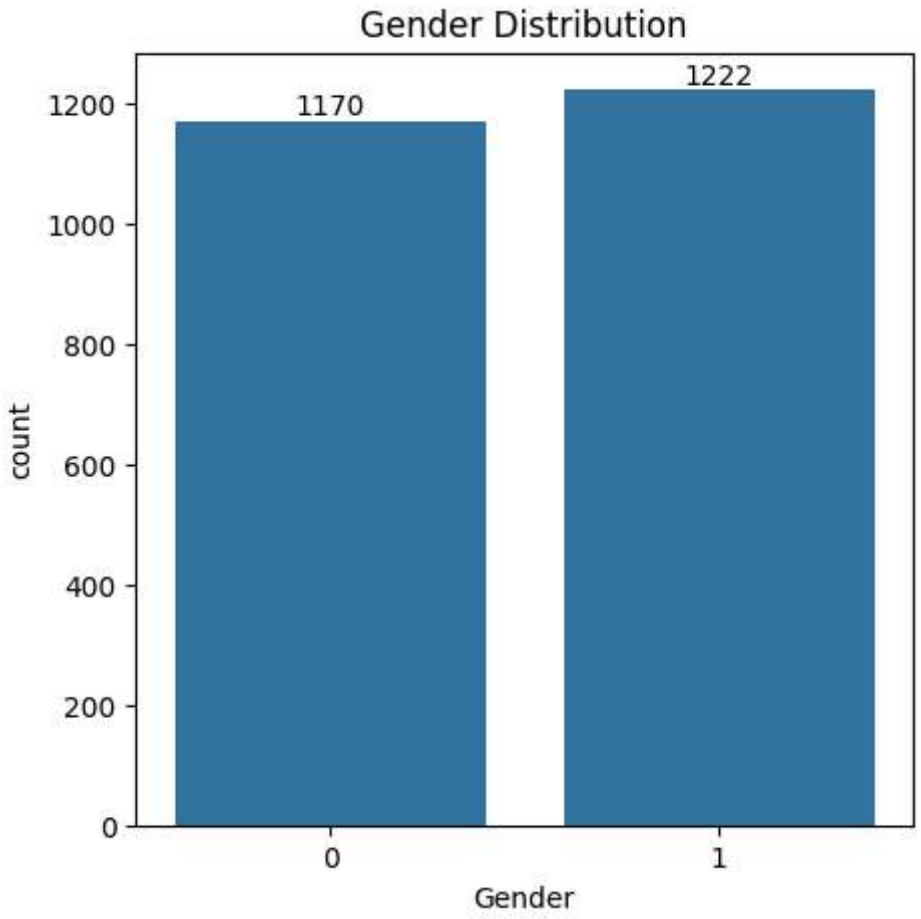
In [32]:
```
df["StudyTimeWeekly"] = df["StudyTimeWeekly"].astype(str).str.replace("4.210570","4.21
df.head()
```

Out[32]:

	Age	Gender	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracu
0	17	1	2	19.833722807854716	7	1	2	
1	18	0	1	15.40875605584674	0	0	1	
2	15	0	3	4.21056976881226	26	0	2	
3	17	1	3	10.028829473958217	14	0	3	
4	17	1	2	4.6724952729713305	17	1	3	

Gender Distribution

In [47]:
```
plt.figure(figsize= (5,5))
ax = sns.countplot(data = df, x = "Gender")
ax.bar_label(ax.containers[0])
plt.title("Gender Distribution")
plt.show()
```

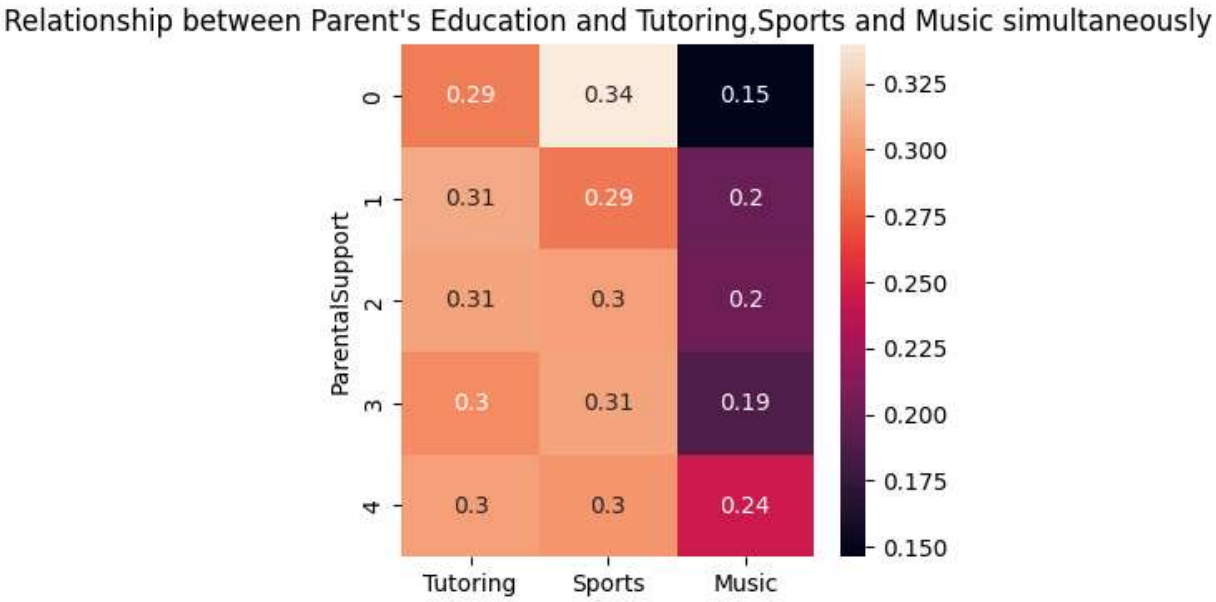


From the above chart we have analysed that: The number of males in the data is more than the number of females

In [37]:
```
gb = df.groupby("ParentalEducation").agg({"Tutoring":'mean',"Sports":'mean',"Music":'m
print(gb)
```

```
                    Tutoring    Sports     Music
ParentalEducation
0                   0.312757  0.304527  0.148148
1                   0.307692  0.304945  0.199176
2                   0.300857  0.306210  0.203426
3                   0.288828  0.272480  0.177112
4                   0.283333  0.366667  0.291667
```

In [48]:
```
plt.figure(figsize= (4,4))
sns.heatmap(gb, annot = True)
plt.title("Relationship between Parent's Education and Tutoring,Sports and Music simul
plt.show()
```

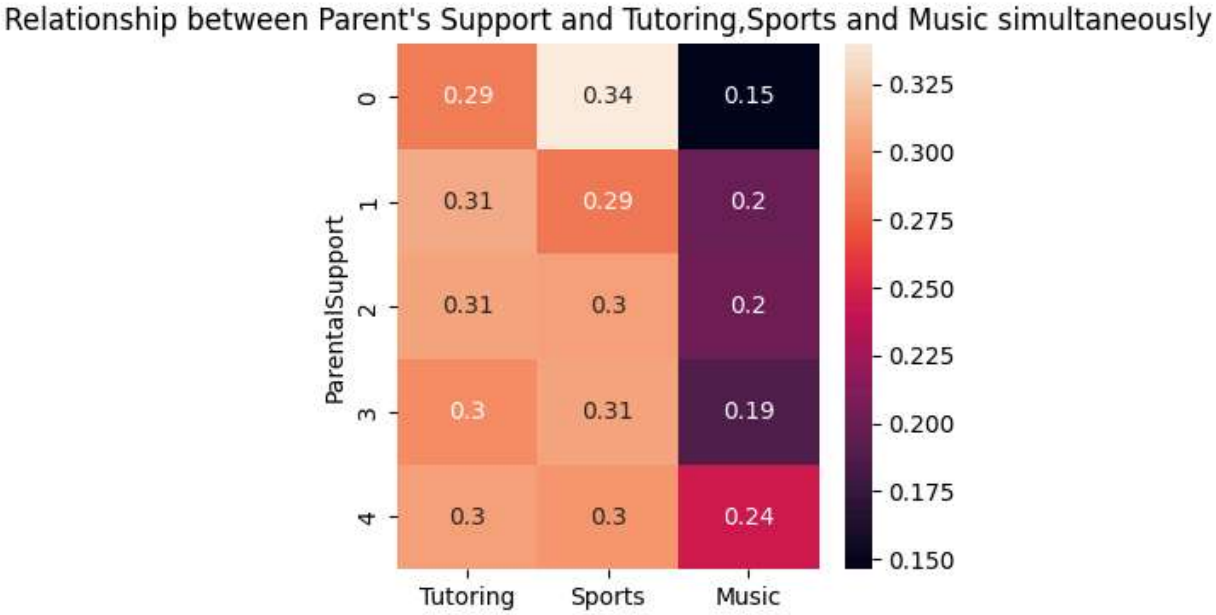


From the above chart we have concluded that the education of the parents have a good impact on their source

In [45]:
```
gb1 = df.groupby("ParentalSupport").agg({"Tutoring":'mean',"Sports":'mean',"Music":'me
print(gb1)
```

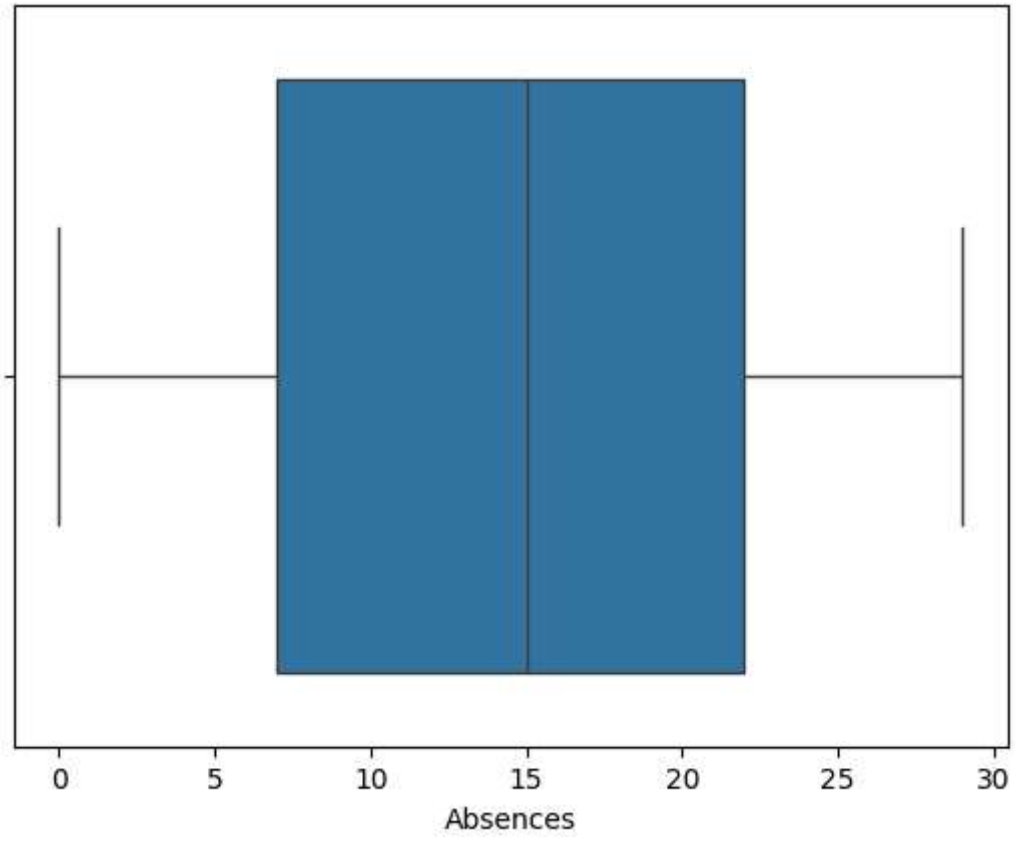
```
                  Tutoring    Sports     Music
ParentalSupport
0                 0.287736  0.339623  0.146226
1                 0.308793  0.286299  0.200409
2                 0.305405  0.304054  0.201351
3                 0.295552  0.305595  0.187948
4                 0.303150  0.299213  0.244094
```

In [49]:
```
plt.figure(figsize= (4,4))
sns.heatmap(gb1, annot = True)
plt.title("Relationship between Parent's Support and Tutoring,Sports and Music simulta
plt.show()
```

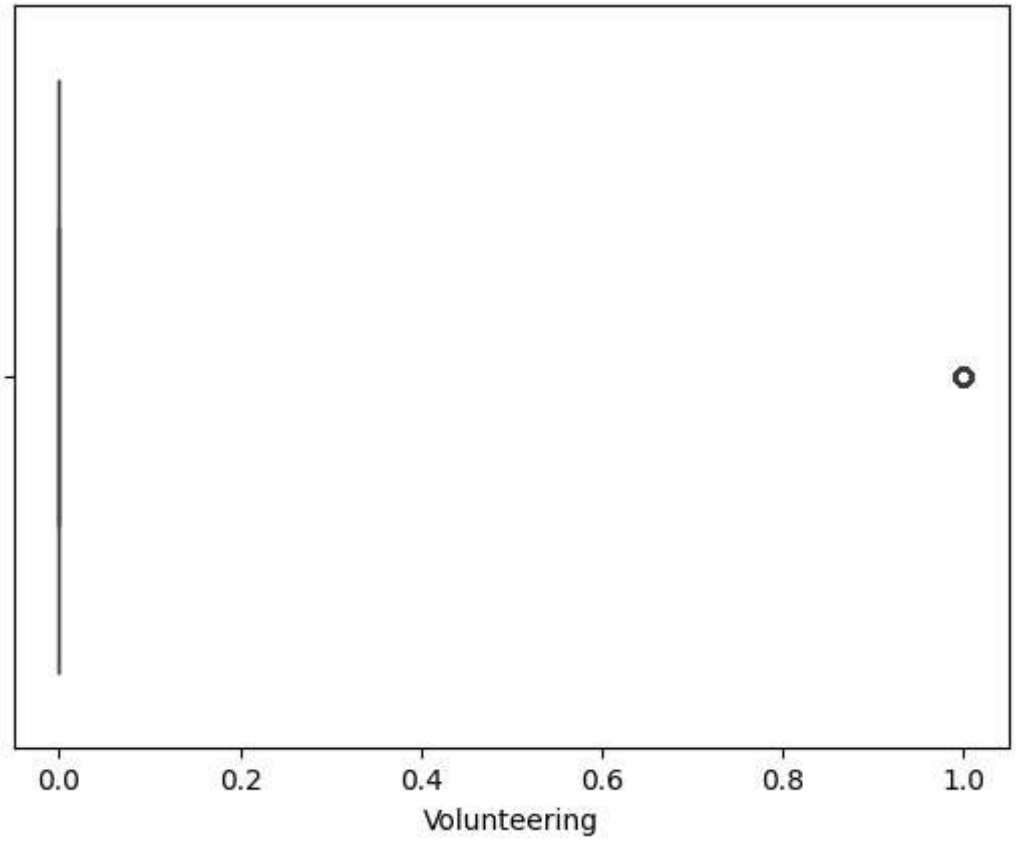


From the above chart we have concluded that there is major impact on the Tutoring and sports whereas no/negligible impact on the Music due to their parental support

In [50]:
```
sns.boxplot(data = df, x = "Absences")
plt.show()
```



In [51]:
```
sns.boxplot(data = df, x = "Volunteering")
plt.show()
```



In [53]:
```
sns.boxplot(data = df, x = "Extracurricular")
plt.show()
```