project of student_score

In [229]:

```
import numpy as np
import pandas as pd
df=pd.read_csv('student_scores.csv')
df['WklyStudyHours']=df['WklyStudyHours'].replace('05-Oct','5-10')
df
```

Out[229]:

	Unnamed: 0	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	IsFirstChild
0	0	female	NaN	bachelor's degree	standard	none	married	regularly	yes
1	1	female	group C	some college	standard	NaN	married	sometimes	yes
2	2	female	group B	master's degree	standard	none	single	sometimes	yes
3	3	male	group A	associate's degree	free/reduced	none	married	never	no
4	4	male	group C	some college	standard	none	married	sometimes	yes
...	...	...	...	...	...	...	...	...	...
30636	816	female	group D	high school	standard	none	single	sometimes	no
30637	890	male	group E	high school	standard	none	single	regularly	no
30638	911	female	NaN	high school	free/reduced	completed	married	sometimes	no
30639	934	female	group D	associate's degree	standard	completed	married	regularly	no
30640	960	male	group B	some college	standard	none	married	never	no

30641 rows × 15 columns

In [230]:

```
print(df.describe())
print(df['MathScore'].sum())
```

```
         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
2039416
```

In [ ]:

In [231]:

```
# df=pd.read_csv('student_scores.csv')
# print(df)
```

```
df
print(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
```

In [232]:

```
print(df['NrSiblings'].mean())
df
df['NrSiblings']=df['NrSiblings'].fillna(2.14)
print(df.isnull().sum())
```

```
2.1458942516082424
Unnamed: 0                0
Gender                    0
EthnicGroup            1840
ParentEduc             1845
LunchType                 0
TestPrep               1830
ParentMaritalStatus    1190
PracticeSport           631
IsFirstChild            904
NrSiblings                0
TransportMeans         3134
WklyStudyHours          955
MathScore                 0
ReadingScore              0
WritingScore              0
dtype: int64
```

In [233]:

```
df
df['Gender'] = df['Gender'].str.lower()
print(df)
gb=df.groupby('ParentEduc').agg({'MathScore':'mean','ReadingScore':'mean','WritingScore'
:'mean'})
print(gb)
```

```
       Unnamed: 0  Gender EthnicGroup          ParentEduc     LunchType  \
0               0  female         NaN   bachelor's degree      standard
1               1  female     group C        some college      standard
2               2  female     group B     master's degree      standard
3               3    male     group A  associate's degree  free/reduced
4               4    male     group C        some college      standard
...           ...     ...         ...                 ...           ...
30636         816  female     group D         high school      standard
30637         890    male     group E         high school      standard
30638         911  female         NaN         high school  free/reduced
30639         934  female     group D  associate's degree      standard
30640         960    male     group B        some college      standard

        TestPrep ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings  \
0           none             married     regularly          yes         3.0
1            NaN             married     sometimes          yes         0.0
2           none              single     sometimes          yes         4.0
3           none             married         never           no         1.0
```

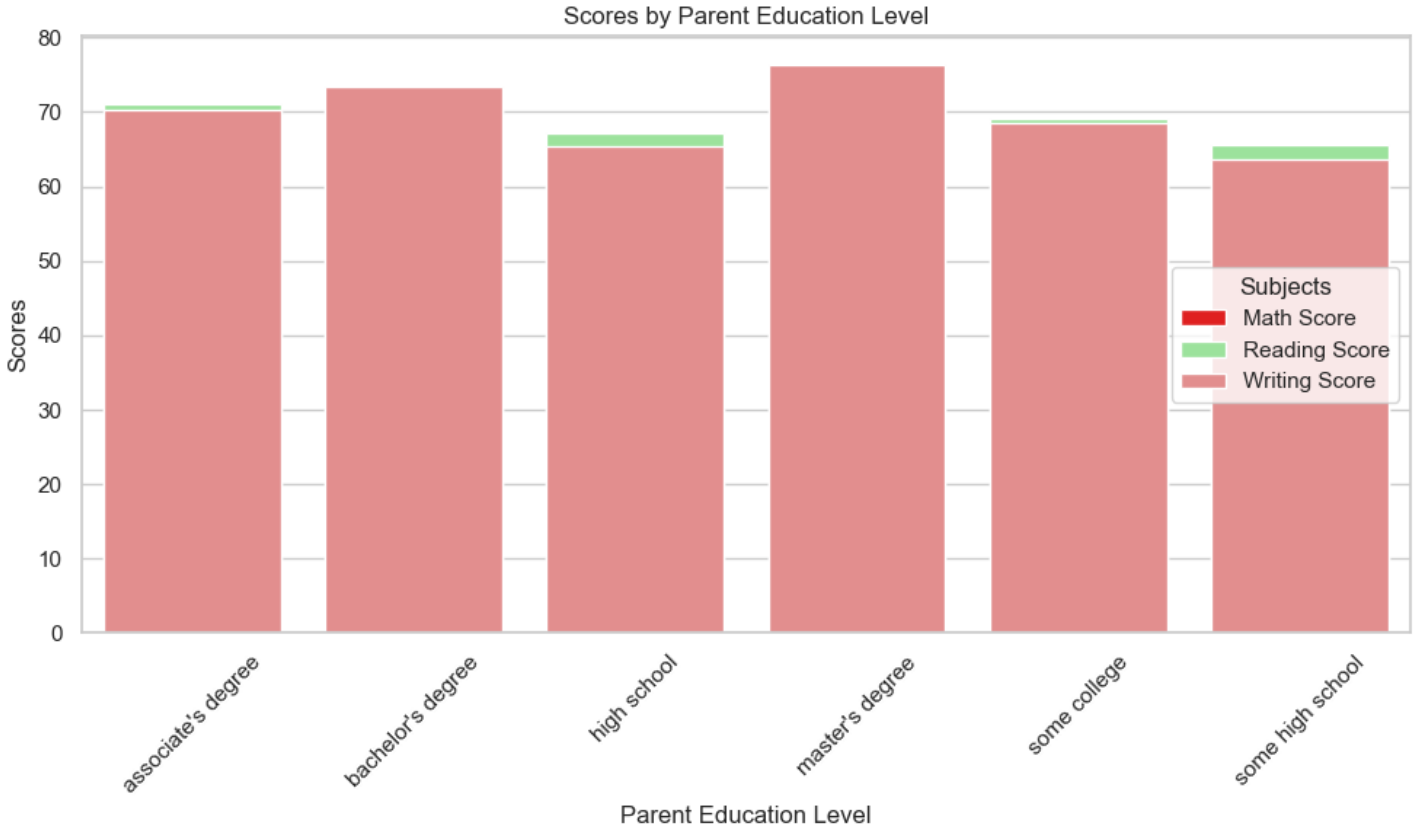
```
4           none              married     sometimes          yes         0.0
...          ...                  ...           ...          ...         ...
30636       none               single     sometimes           no         2.0
30637       none               single     regularly           no         1.0
30638  completed              married     sometimes           no         1.0
30639  completed              married     regularly           no         3.0
30640       none              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 15 columns]
                   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 [234]:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Your data (assuming it's already in a DataFrame)
data2 = {
    'ParentEduc': ["associate's degree", "bachelor's degree", "high school", "master's de
gree", "some college", "some high school"],
    'MathScore': [68.365586, 70.466627, 64.435731, 72.336134, 66.390472, 62.584013],
    'ReadingScore': [71.124324, 73.062020, 67.213997, 75.832921, 69.179708, 65.510785],
    'WritingScore': [70.299099, 73.331069, 65.421136, 76.356896, 68.501432, 63.632409]
}

# Create a DataFrame
df2 = pd.DataFrame(data2)
print(df2)

# Set a seaborn style
sns.set(style="whitegrid")

# Create a grouped bar chart
plt.figure(figsize=(10, 6))
sns.barplot(x='ParentEduc', y='MathScore', data=df2, color='red', label='Math Score')
sns.barplot(x='ParentEduc', y='ReadingScore', data=df2, color='lightgreen', label='Readi
ng Score')
sns.barplot(x='ParentEduc', y='WritingScore', data=df2, color='lightcoral', label='Writi
ng Score')

# Customize the plot
plt.xlabel("Parent Education Level")
plt.ylabel("Scores")
plt.title("Scores by Parent Education Level")
plt.xticks(rotation=45)
plt.legend(title='Subjects',loc=5)

# Show the plot
plt.tight_layout()
plt.show()
```

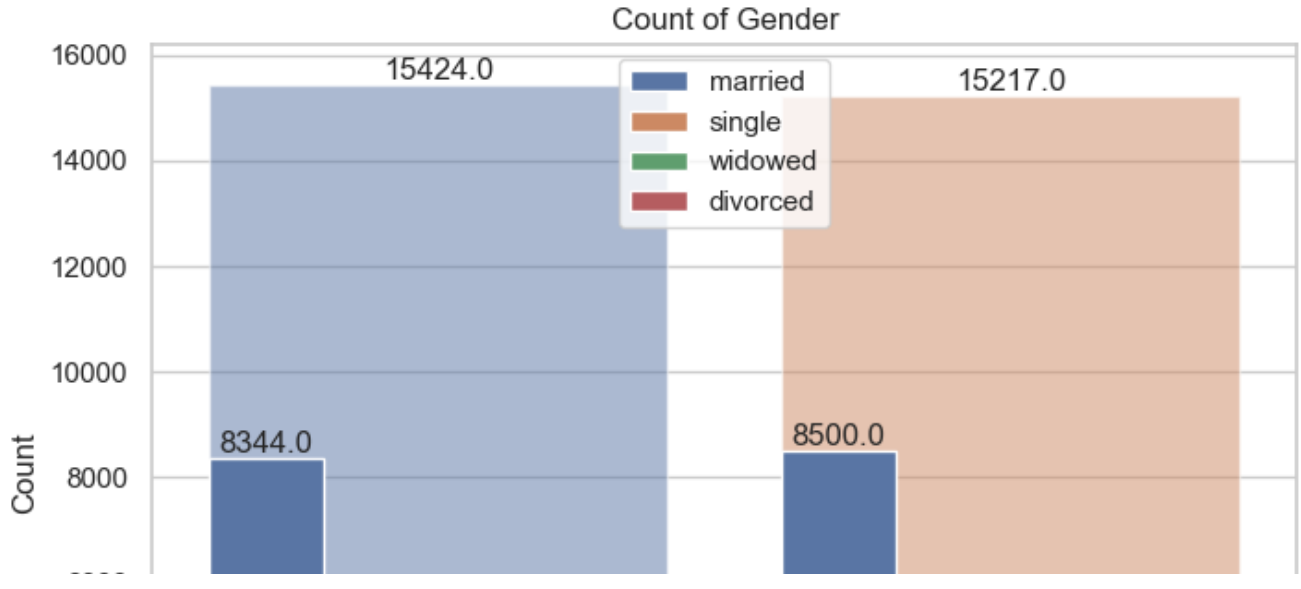
```
           ParentEduc  MathScore  ReadingScore  WritingScore
```

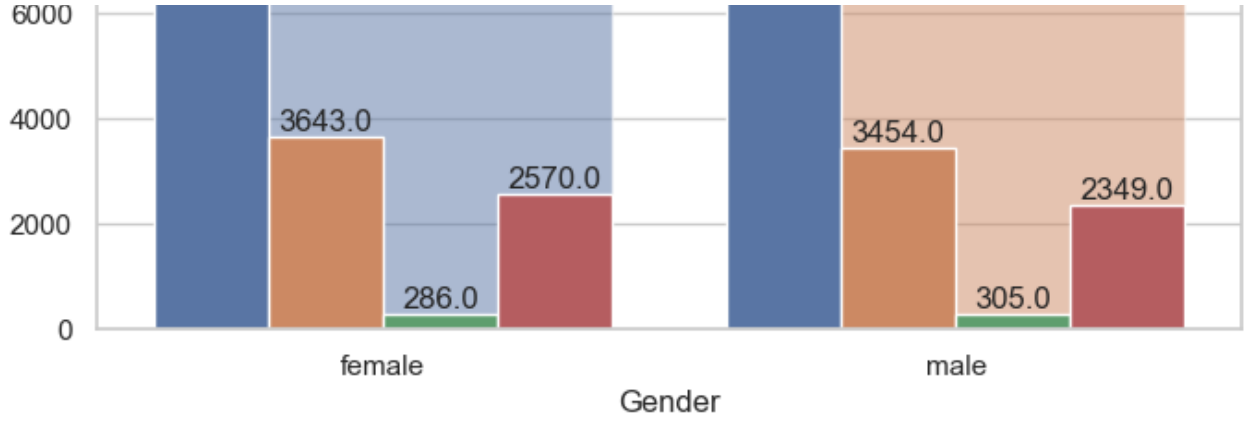
```
0  associate's degree  68.365586     71.124324     70.299099
1   bachelor's degree  70.466627     73.062020     73.331069
2         high school  64.435731     67.213997     65.421136
3     master's degree  72.336134     75.832921     76.356896
4        some college  66.390472     69.179708     68.501432
5    some high school  62.584013     65.510785     63.632409
```



In [287]:

```
# print(df)
plt.figure(figsize=(8, 6))
ax1=sns.countplot(data=df, x='Gender',alpha=.5)
ax1=sns.countplot(data=df, x='Gender',hue='ParentMaritalStatus')
for p in ax1.patches:
    ax1.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2., p.get_height()),
                 ha='center', va='bottom', fontsize=12)


plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Count of Gender')
plt.legend(loc=9)

plt.show()
```





In [242]:

```
print(df)
```

```
       Unnamed: 0  Gender EthnicGroup          ParentEduc     LunchType  \
0               0  female          NaN   bachelor's degree      standard
1               1  female      group C        some college      standard
2               2  female      group B     master's degree      standard
3               3    male      group A  associate's degree  free/reduced
4               4    male      group C        some college      standard
...           ...     ...          ...                 ...           ...
30636         816  female      group D         high school      standard
30637         890    male      group E         high school      standard
30638         911  female          NaN         high school  free/reduced
30639         934  female      group D  associate's degree      standard
30640         960    male      group B        some college      standard

        TestPrep ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings  \
0           none             married     regularly          yes         3.0
1            NaN             married     sometimes          yes         0.0
2           none              single     sometimes          yes         4.0
3           none             married         never           no         1.0
4           none             married     sometimes          yes         0.0
...          ...                 ...           ...         ...         ...
30636       none              single     sometimes           no         2.0
30637       none              single     regularly           no         1.0
30638  completed             married     sometimes           no         1.0
30639  completed             married     regularly           no         3.0
30640       none             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 15 columns]
Unnamed: 0             2219
Gender                 2219
EthnicGroup            2219
ParentEduc             2078
LunchType              2219
TestPrep               2081
ParentMaritalStatus    2121
PracticeSport          2167
IsFirstChild           2168
NrSiblings             2219
TransportMeans         1999
WklyStudyHours         2146
MathScore              2219
```

```
ReadingScore           2219
WritingScore           2219
dtype: int64
```

In [248]:

```
# df1=df.loc[(df['Category'])=='Bills']
x=df.loc[(df['EthnicGroup']=='group A')]
print(x)
```

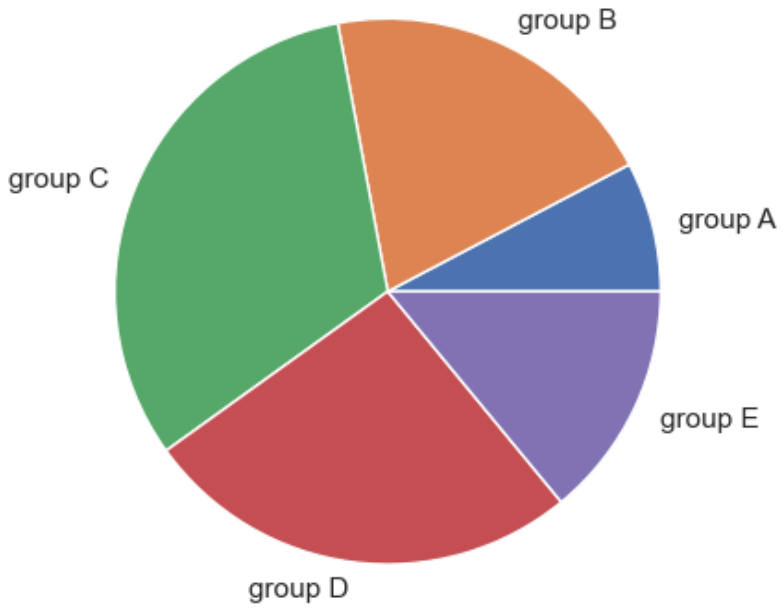
```
       Unnamed: 0  Gender EthnicGroup          ParentEduc     LunchType  \
3               3    male     group A  associate's degree  free/reduced
13             13    male     group A        some college      standard
14             14  female     group A     master's degree      standard
25             25    male     group A     master's degree  free/reduced
56             61    male     group A    some high school  free/reduced
...           ...     ...         ...                 ...           ...
30603         281    male     group A         high school      standard
30621         638  female     group A   bachelor's degree      standard
30622         640    male     group A  associate's degree  free/reduced
30627         730  female     group A         high school      standard
30634         785    male     group A  associate's degree  free/reduced

        TestPrep ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings  \
3           none             married         never           no        1.00
13     completed              single     sometimes          yes        1.00
14          none            divorced     sometimes          yes        2.00
25          none             married     regularly          yes        1.00
56          none             married     sometimes          yes        2.14
...          ...                 ...           ...          ...         ...
30603       none              single     regularly           no        2.00
30621       none              single     regularly           no        2.00
30622  completed            divorced     regularly           no        3.00
30627  completed             married         never           no        2.14
30634  completed                 NaN     sometimes           no        2.00

      TransportMeans WklyStudyHours  MathScore  ReadingScore  WritingScore
3                NaN           5-10         45            56            42
13           private          > 10         80            73            71
14           private           < 5         48            53            58
25        school_bus           5-10         75            76            76
56        school_bus           5-10         39            39            34
...              ...            ...        ...           ...           ...
30603     school_bus           5-10         71            63            65
30621        private           5-10         66            80            78
30622        private           5-10         53            53            53
30627     school_bus          > 10         58            77            82
30634     school_bus           5-10         65            60            60

[2219 rows x 15 columns]
```

In [277]:

```
print(df['EthnicGroup'].unique())
ga=df.loc[(df['EthnicGroup']=='group A')].count()[2]
print(ga)
gb=df.loc[(df['EthnicGroup']=='group B')].count()[2]
print(gb)
gc=df.loc[(df['EthnicGroup']=='group C')].count()[2]
print(gc)
gd=df.loc[(df['EthnicGroup']=='group D')].count()[2]
print(gd)
ge=df.loc[(df['EthnicGroup']=='group E')].count()[2]
print(ge)
plt.pie([ga,gb,gc,gd,ge], labels=['group A','group B','group C','group D','group E'])
plt.show()
```

```
[nan 'group C' 'group B' 'group A' 'group D' 'group E']
2219
5826
9212
7503
```

4041



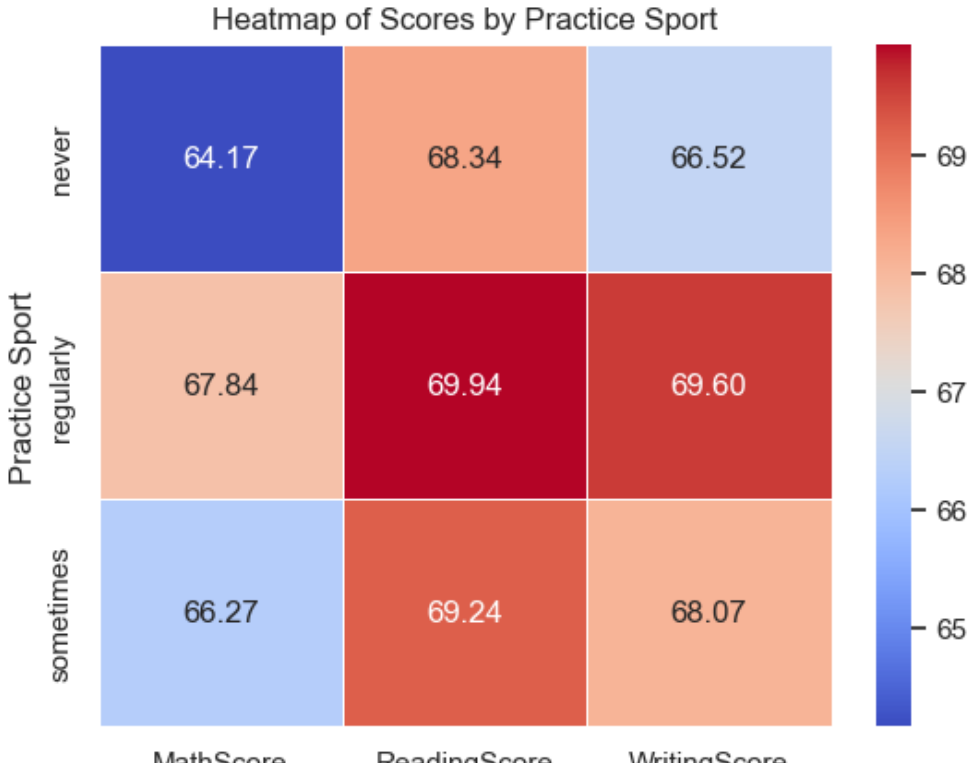
In [296]:

```
df
gb=df.groupby('PracticeSport').agg({'MathScore':'mean','ReadingScore':'mean','WritingSco
re':'mean'})
print(gb)
```

```
               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 [299]:

```
sns.heatmap(data=gb, annot=True, cmap='coolwarm', linewidths=0.5, fmt=".2f")
plt.xlabel('Subject')
plt.ylabel('Practice Sport')
plt.title('Heatmap of Scores by Practice Sport')
plt.show()
```



MathScore ReadingScore WritingScore

Subject