Project Name

Relationship between Parent Education Level and Student Study

Name of the University/School

University Teknikal Malaysia Melaka (UTeM)

Team Member(s)

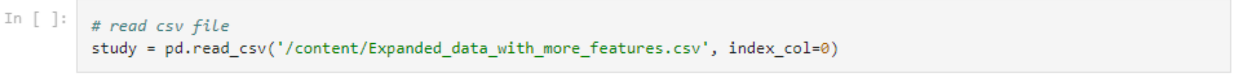
Ethan Lew (B032110072)
Low Pei Zuo (B032010013)
Kaw Zi Jian (B032010029)

1.0 Aims of AI

This study aims to investigate the relationship between parent education levels and student study behavior, exploring whether higher levels of parental education positively impact student academic performance. It seeks to identify the influence of parental involvement, expectations, and aspirations on student study habits and learning outcomes. The research will assess potential mediating factors, such as socioeconomic status and access to educational resources, to gain a comprehensive understanding of the parent-student educational connection.

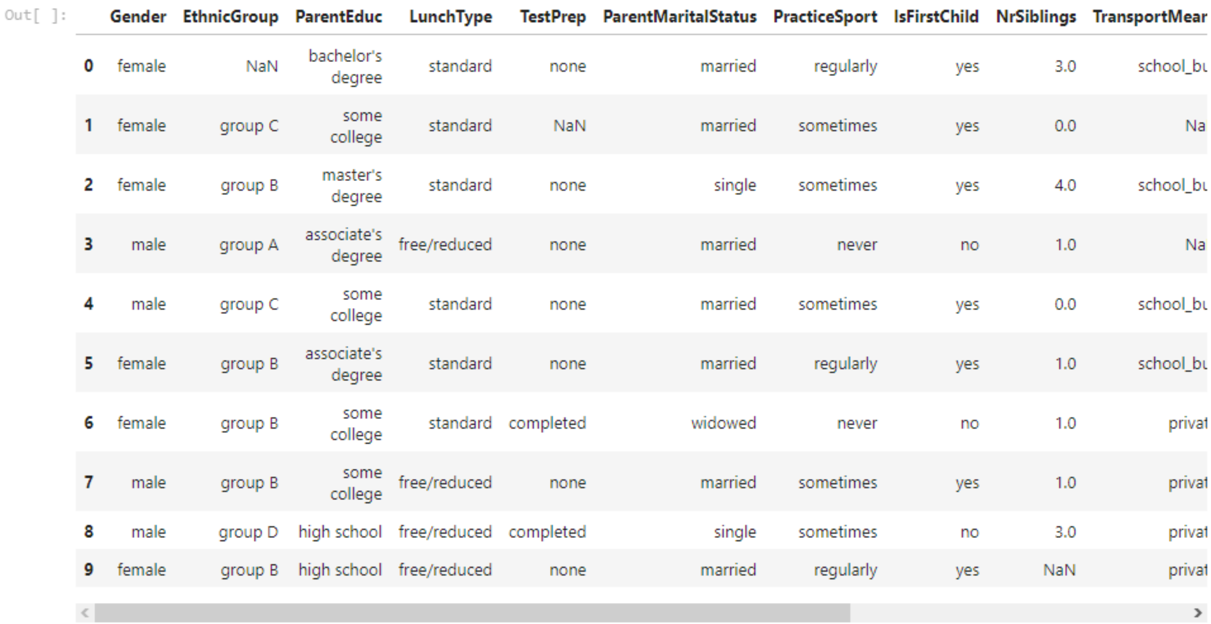
2.0 How to Deploy Our Prediction

1. Read the dataset.



```
In [ ]:  # read csv file
         study = pd.read_csv('/content/Expanded_data_with_more_features.csv', index_col=0)
```

Output:

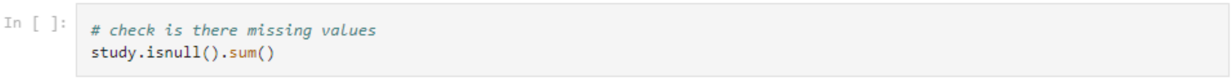


Out[ ]:

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	IsFirstChild	NrSiblings	TransportMear
0	female	NaN	bachelor's degree	standard	none	married	regularly	yes	3.0	school_bu
1	female	group C	some college	standard	NaN	married	sometimes	yes	0.0	Na
2	female	group B	master's degree	standard	none	single	sometimes	yes	4.0	school_bu
3	male	group A	associate's degree	free/reduced	none	married	never	no	1.0	Na
4	male	group C	some college	standard	none	married	sometimes	yes	0.0	school_bu
5	female	group B	associate's degree	standard	none	married	regularly	yes	1.0	school_bu
6	female	group B	some college	standard	completed	widowed	never	no	1.0	privat
7	male	group B	some college	free/reduced	none	married	sometimes	yes	1.0	privat
8	male	group D	high school	free/reduced	completed	single	sometimes	no	3.0	privat
9	female	group B	high school	free/reduced	none	married	regularly	yes	NaN	privat

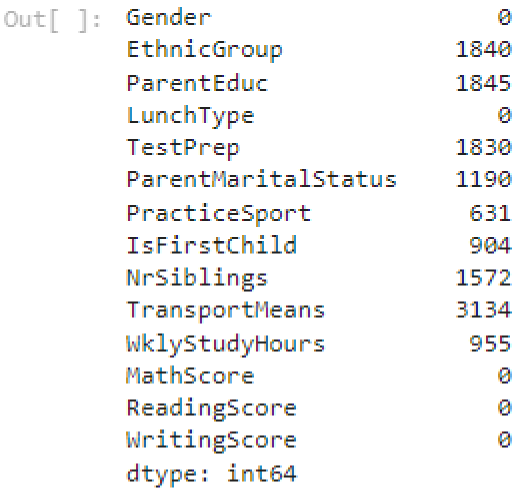
2. Data Cleaning

i. Check the missing values.



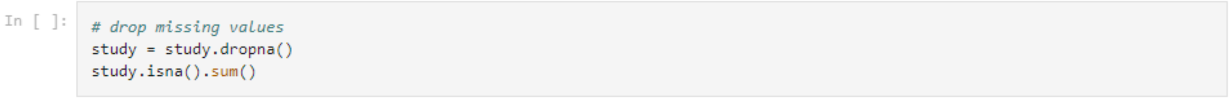
```
In [ ]: # check is there missing values
        study.isnull().sum()
```

Output:



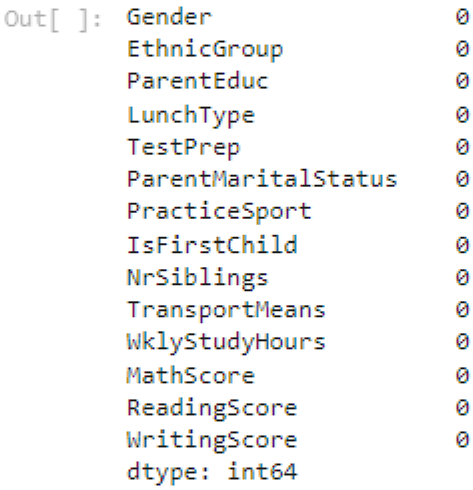
```
Out[ ]: Gender                    0
        EthnicGroup            1840
        ParentEduc             1845
        LunchType                 0
        TestPrep               1830
        ParentMaritalStatus    1190
        PracticeSport           631
        IsFirstChild            904
        NrSiblings             1572
        TransportMeans         3134
        WklyStudyHours          955
        MathScore                 0
        ReadingScore              0
        WritingScore              0
        dtype: int64
```

ii. Drop the missing values.



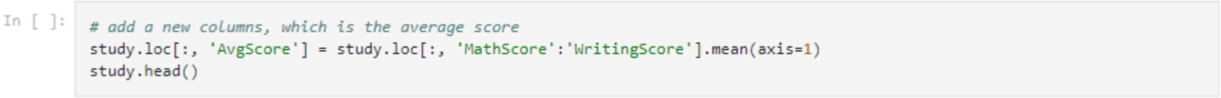
```
In [ ]: # drop missing values
        study = study.dropna()
        study.isna().sum()
```

Output:



```
Out[ ]: Gender                 0
        EthnicGroup            0
        ParentEduc             0
        LunchType              0
        TestPrep               0
        ParentMaritalStatus    0
        PracticeSport          0
        IsFirstChild           0
        NrSiblings             0
        TransportMeans         0
        WklyStudyHours         0
        MathScore              0
        ReadingScore           0
        WritingScore           0
        dtype: int64
```

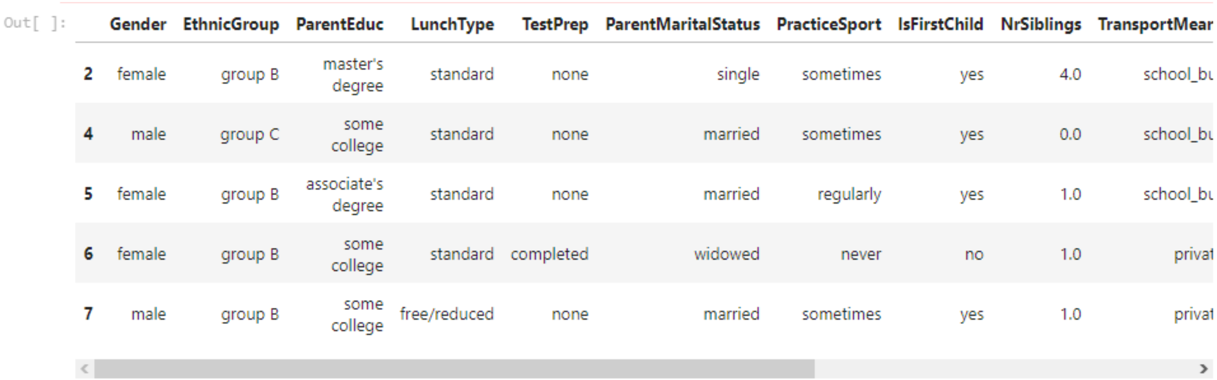
iii. Add new column into the dataset.



```
In [ ]: # add a new columns, which is the average score
        study.loc[:, 'AvgScore'] = study.loc[:, 'MathScore':'WritingScore'].mean(axis=1)
        study.head()
```

a

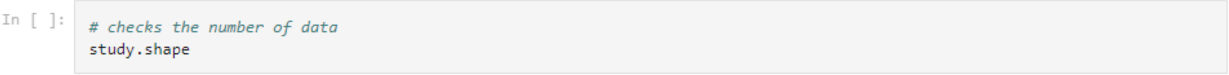
Output:



Out[ ]:

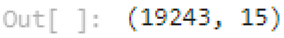
	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	PracticeSport	IsFirstChild	NrSiblings	TransportMear
2	female	group B	master's degree	standard	none	single	sometimes	yes	4.0	school_bu
4	male	group C	some college	standard	none	married	sometimes	yes	0.0	school_bu
5	female	group B	associate's degree	standard	none	married	regularly	yes	1.0	school_bu
6	female	group B	some college	standard	completed	widowed	never	no	1.0	privat
7	male	group B	some college	free/reduced	none	married	sometimes	yes	1.0	privat

iv. Check the number of data.



```
In [ ]: # checks the number of data
        study.shape
```

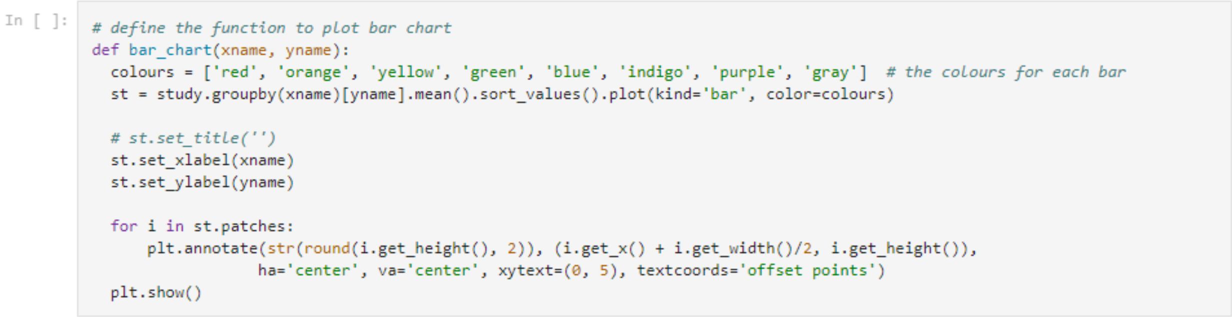
Output:



```
Out[ ]: (19243, 15)
```

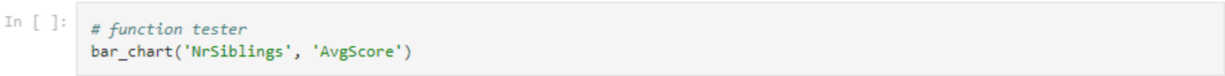
3. Data Analysis

i. Create a bar chart to analysis the column dataset.a



In [ ]:

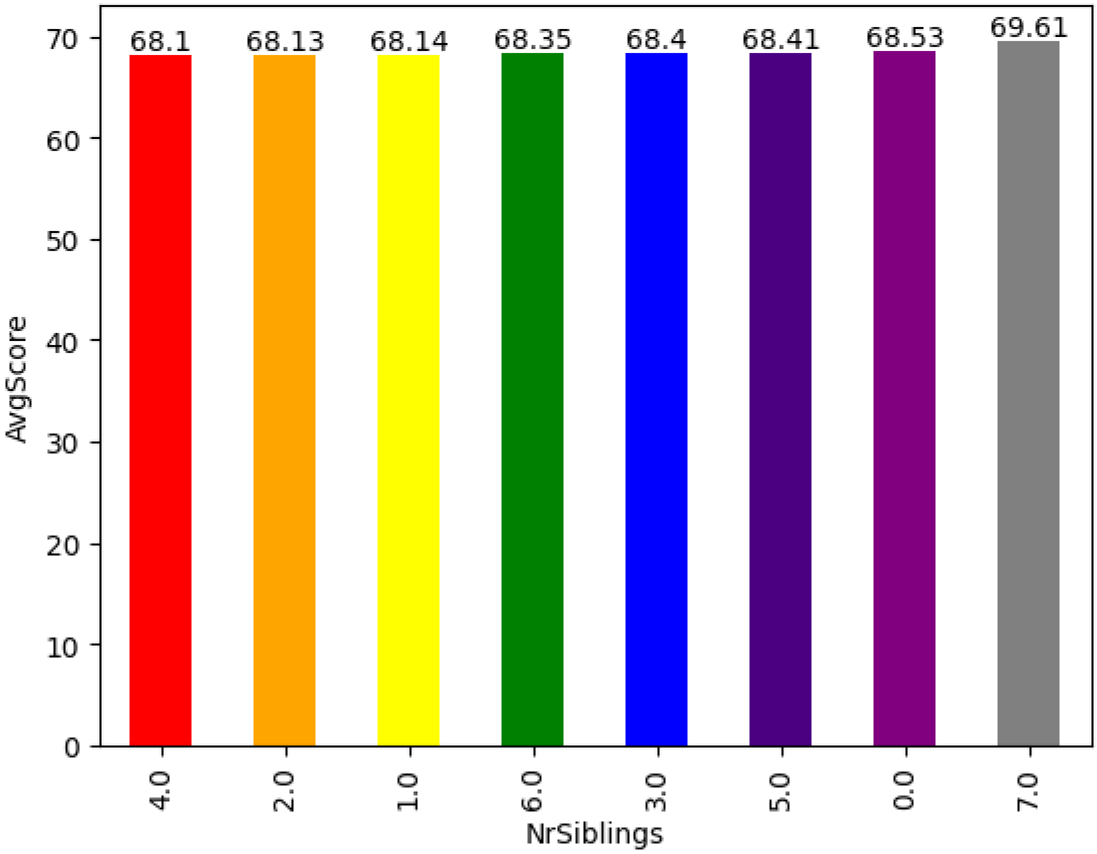
```
# define the function to plot bar chart
def bar_chart(xname, yname):
  colours = ['red', 'orange', 'yellow', 'green', 'blue', 'indigo', 'purple', 'gray']  # the colours for each bar
  st = study.groupby(xname)[yname].mean().sort_values().plot(kind='bar', color=colours)

  # st.set_title('')
  st.set_xlabel(xname)
  st.set_ylabel(yname)

  for i in st.patches:
      plt.annotate(str(round(i.get_height(), 2)), (i.get_x() + i.get_width()/2, i.get_height()),
                  ha='center', va='center', xytext=(0, 5), textcoords='offset points')
  plt.show()
```



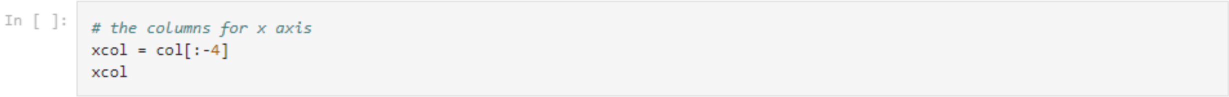
In [ ]:

```
# function tester
bar_chart('NrSiblings', 'AvgScore')
```

Output:

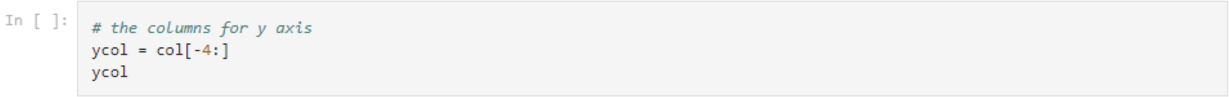


ii. Save the column for x and y axis.



In [ ]:

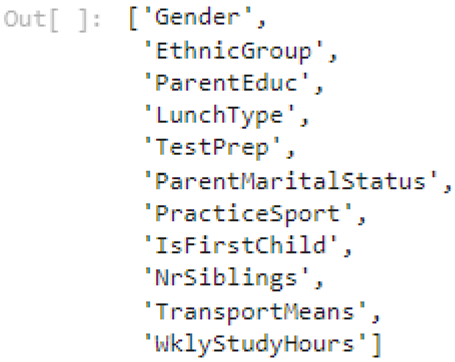
```
# the columns for x axis
xcol = col[:-4]
xcol
```



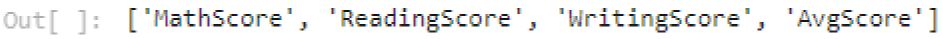
In [ ]:

```
# the columns for y axis
ycol = col[-4:]
ycol
```

Output:

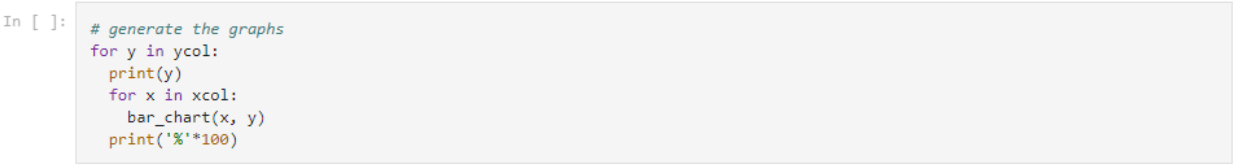


```
Out[ ]: ['Gender',
          'EthnicGroup',
          'ParentEduc',
          'LunchType',
          'TestPrep',
          'ParentMaritalStatus',
          'PracticeSport',
          'IsFirstChild',
          'NrSiblings',
          'TransportMeans',
          'WklyStudyHours']
```



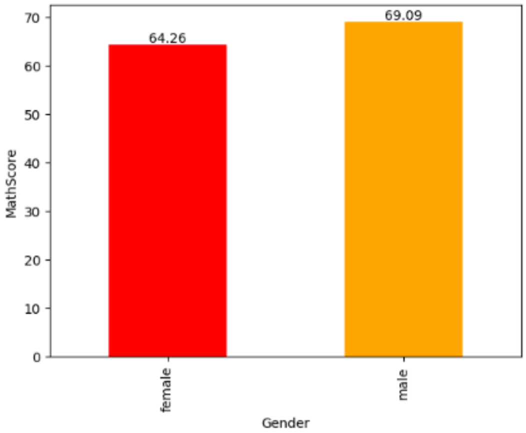
```
Out[ ]: ['MathScore', 'ReadingScore', 'WritingScore', 'AvgScore']
```

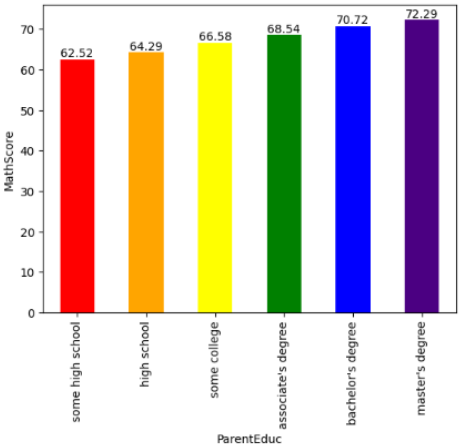
iii. Generate the graph based on Math Score, Reading Score, Writing Score and Average Score

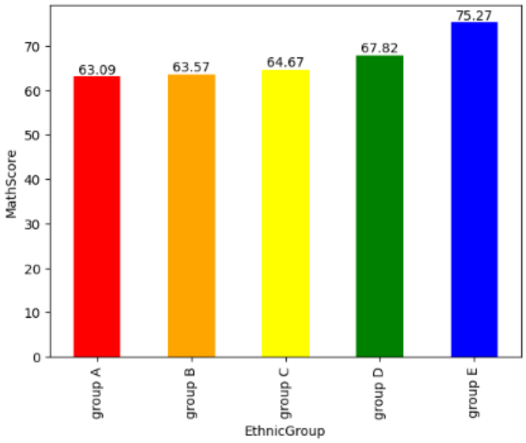


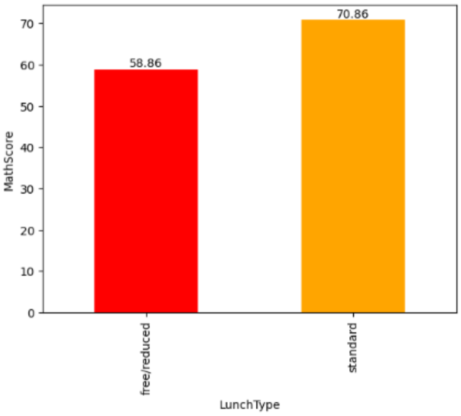
```
In [ ]: # generate the graphs
        for y in ycol:
          print(y)
          for x in xcol:
            bar_chart(x, y)
          print('%'*100)
```

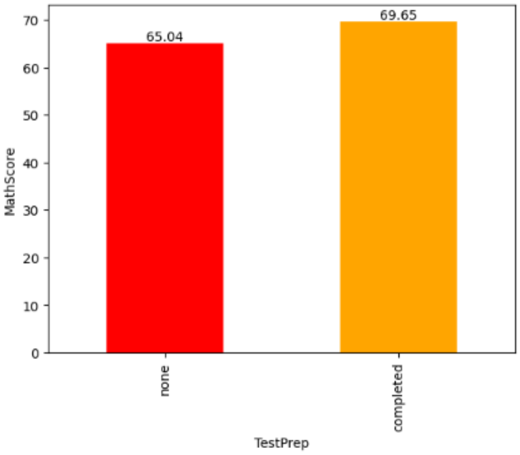
1. Output for Math Score:

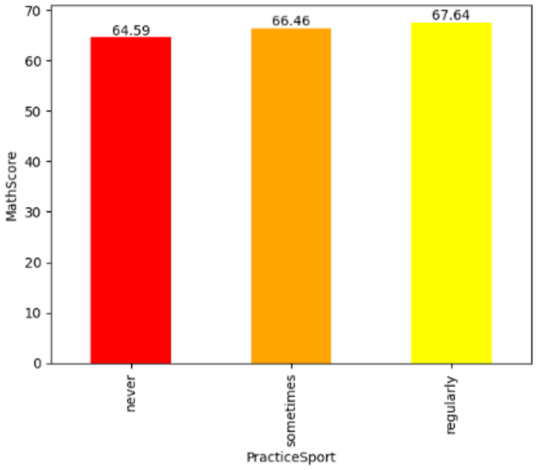

MathScore
Gender
female
male
64.26
69.09
0
10
20
30
40
50
60
70

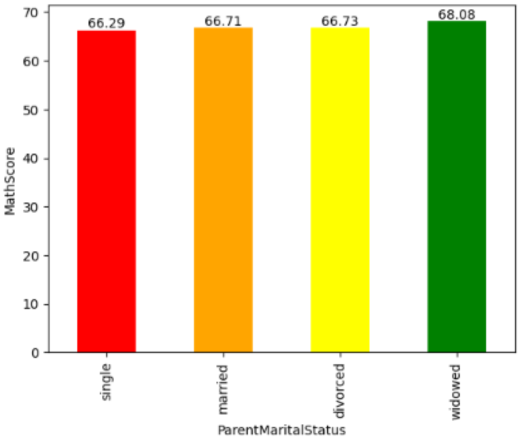

MathScore
ParentEduc
some high school
high school
some college
associate's degree
bachelor's degree
master's degree
62.52
64.29
66.58
68.54
70.72
72.29
0
10
20
30
40
50
60
70

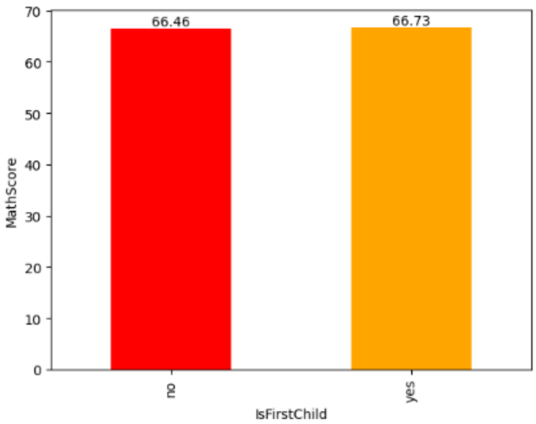

MathScore
EthnicGroup
group A
group B
group C
group D
group E
63.09
63.57
64.67
67.82
75.27
0
10
20
30
40
50
60
70

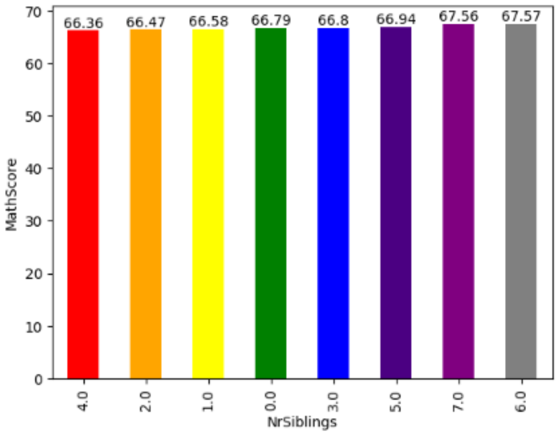

MathScore
LunchType
free/reduced
standard
58.86
70.86
0
10
20
30
40
50
60
70

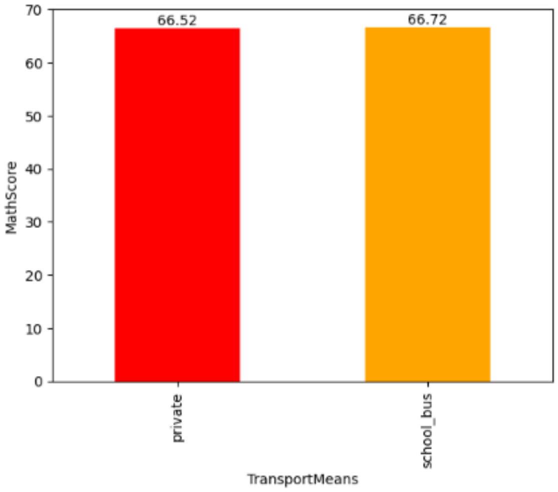

MathScore
TestPrep
none
completed
65.04
69.65
0
10
20
30
40
50
60
70

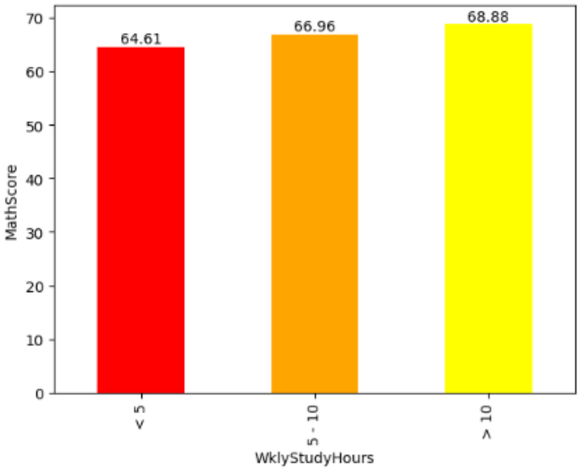

MathScore
PracticeSport
never
sometimes
regularly
64.59
66.46
67.64
0
10
20
30
40
50
60
70


MathScore
ParentMaritalStatus
single
married
divorced
widowed
66.29
66.71
66.73
68.08
0
10
20
30
40
50
60
70

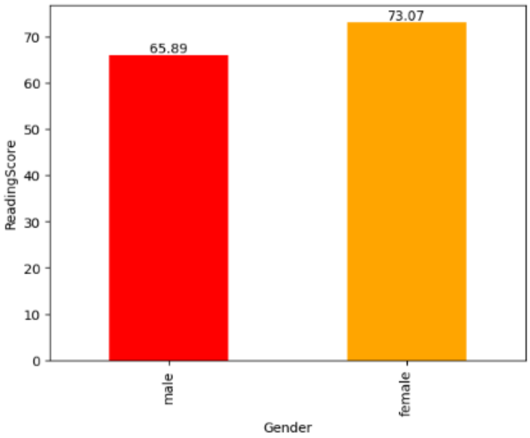

MathScore
IsFirstChild
no
yes
66.46
66.73
0
10
20
30
40
50
60
70

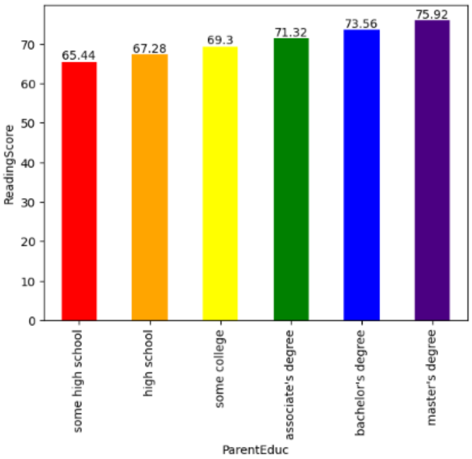

MathScore
0
10
20
30
40
50
60
70
66.36
66.47
66.58
66.79
66.8
66.94
67.56
67.57
4.0
2.0
1.0
0.0
3.0
5.0
7.0
6.0
NrSiblings

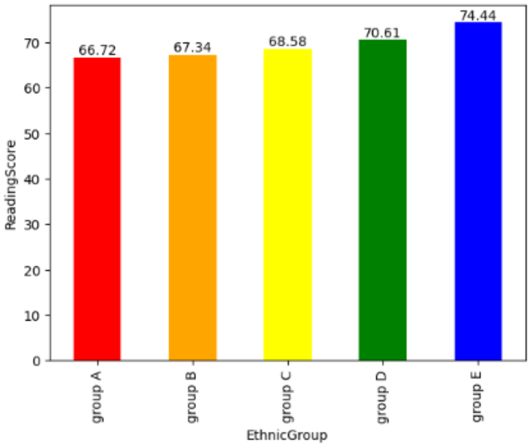

MathScore
0
10
20
30
40
50
60
70
66.52
66.72
private
school_bus
TransportMeans

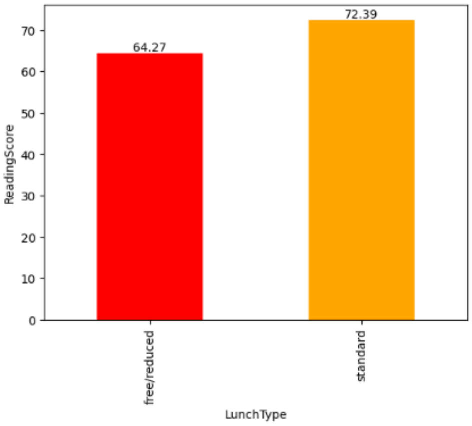

MathScore
0
10
20
30
40
50
60
70
64.61
66.96
68.88
< 5
5 - 10
> 10
WklyStudyHours

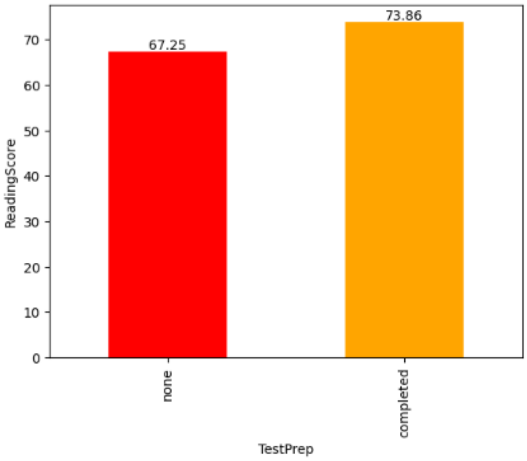
2. Output for Reading Score:

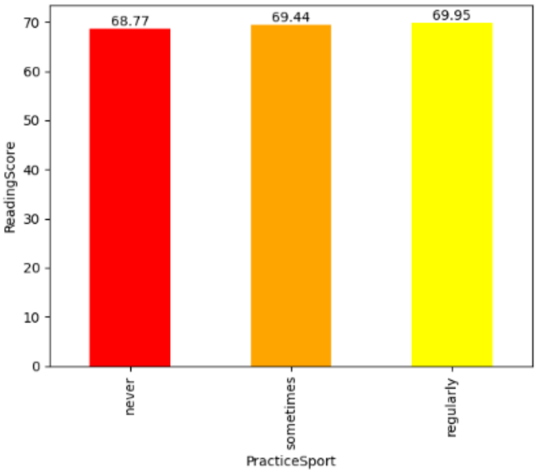


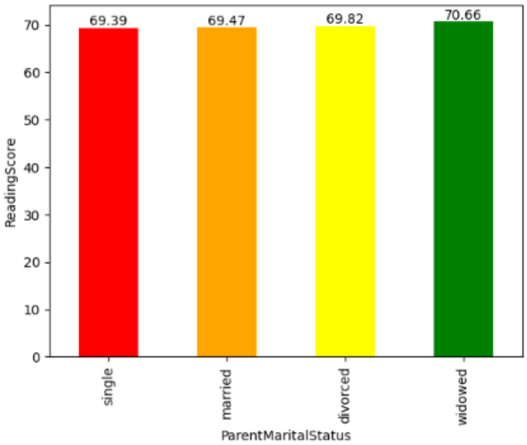


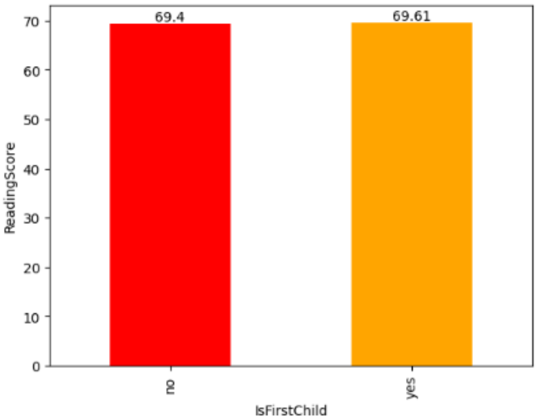


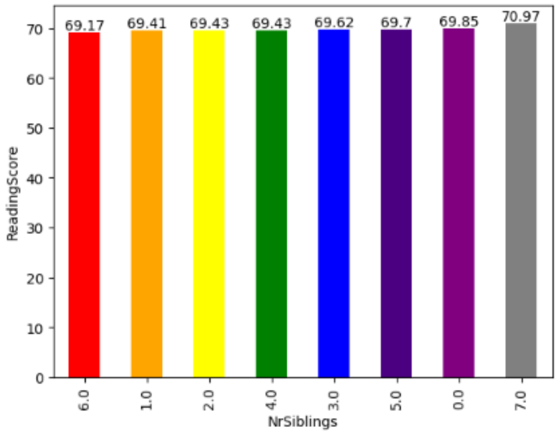


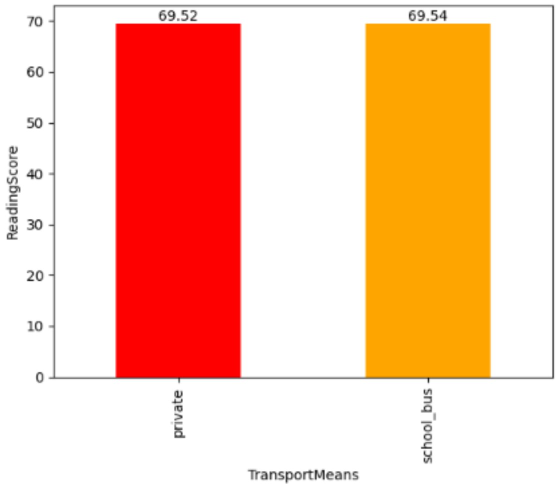


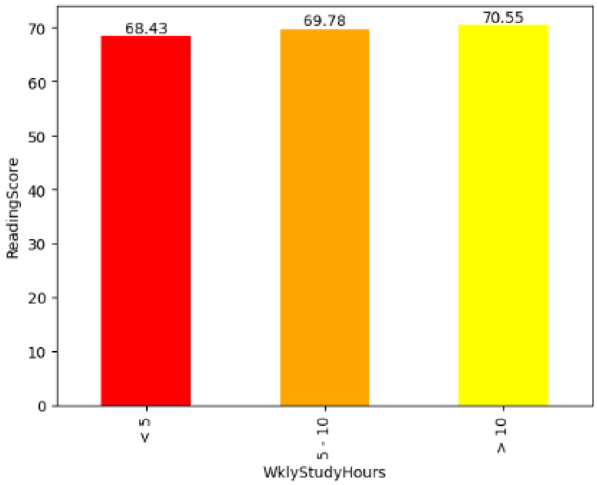




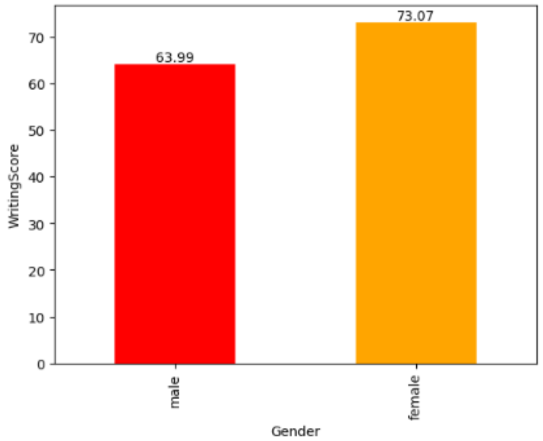


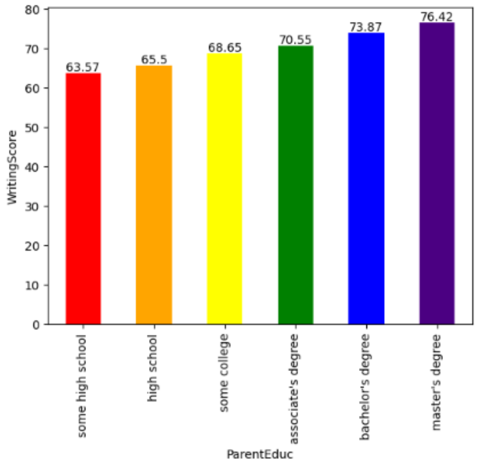

69.17
69.41
69.43
69.43
69.62
69.7
69.85
70.97
ReadingScore
0
10
20
30
40
50
60
70
6.0
1.0
2.0
4.0
3.0
5.0
0.0
7.0
NrSiblings

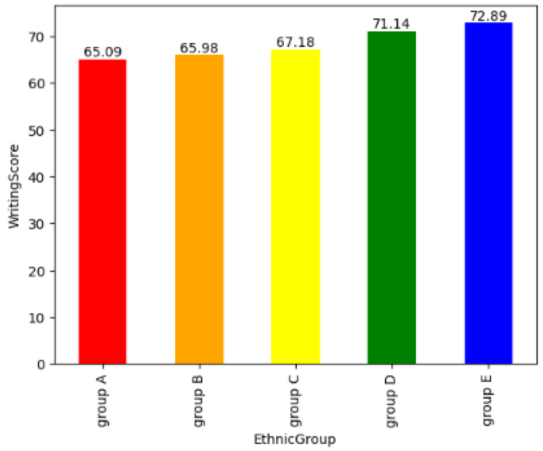

69.52
69.54
ReadingScore
0
10
20
30
40
50
60
70
private
school_bus
TransportMeans

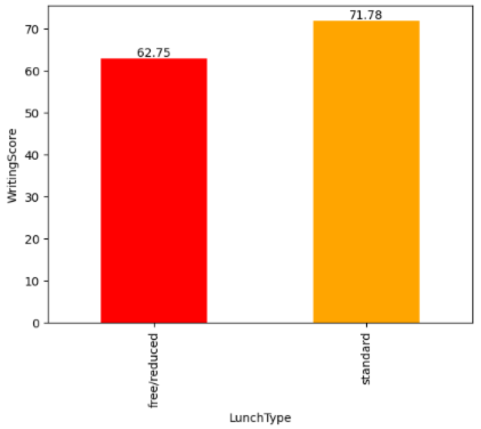

68.43
69.78
70.55
ReadingScore
0
10
20
30
40
50
60
70
< 5
5 - 10
> 10
WklyStudyHours

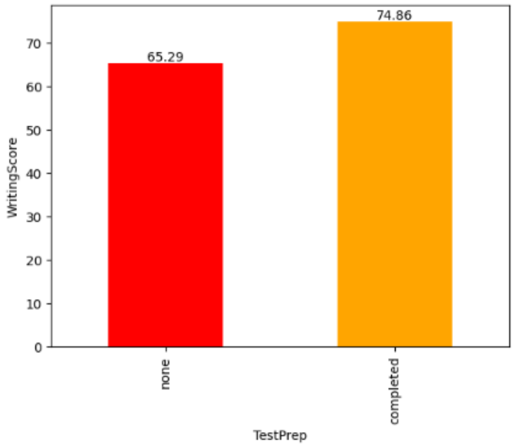
3. Output for Writing Score:

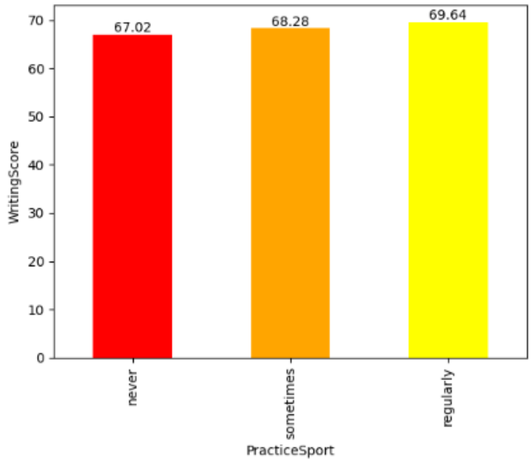


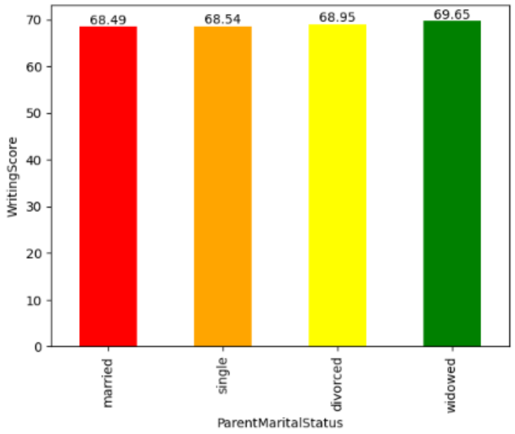


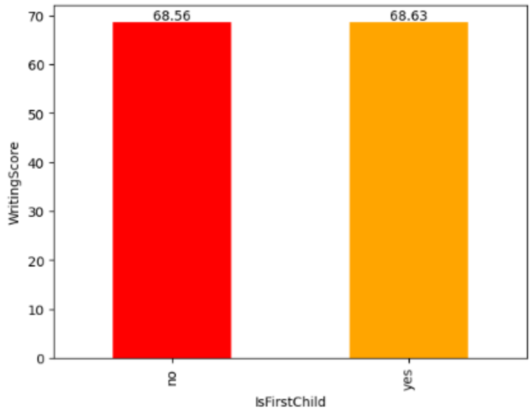


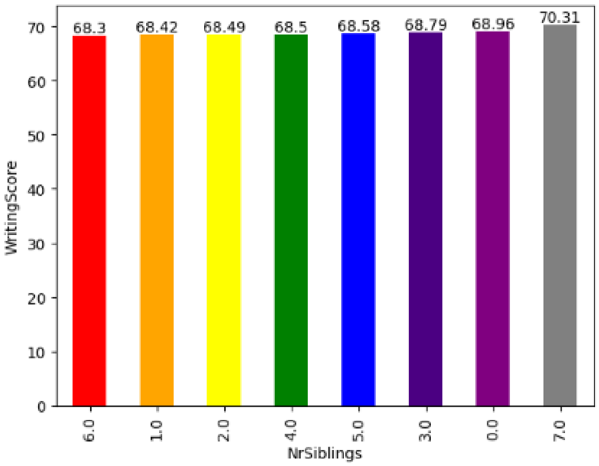


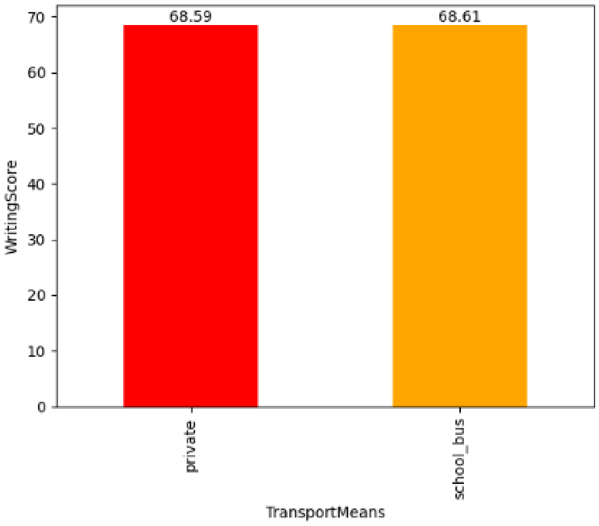


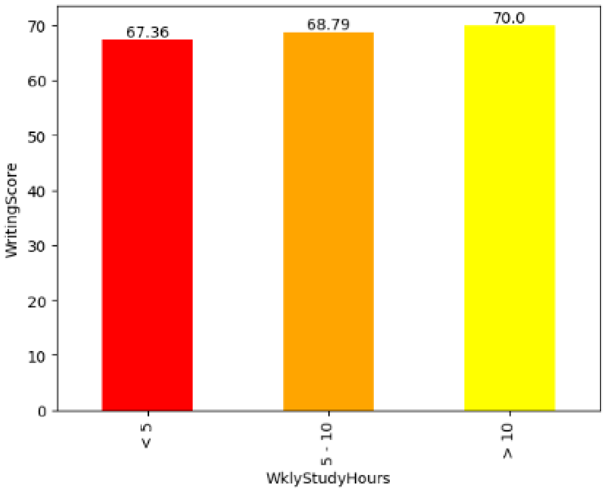




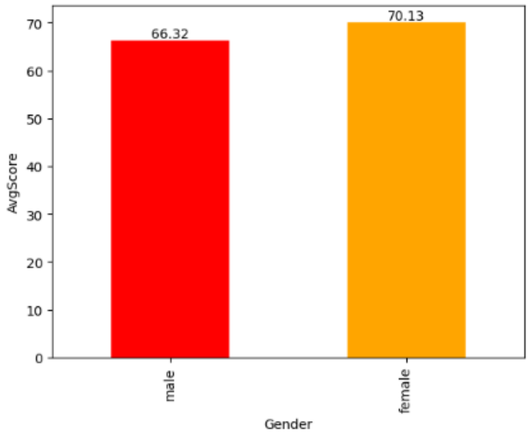


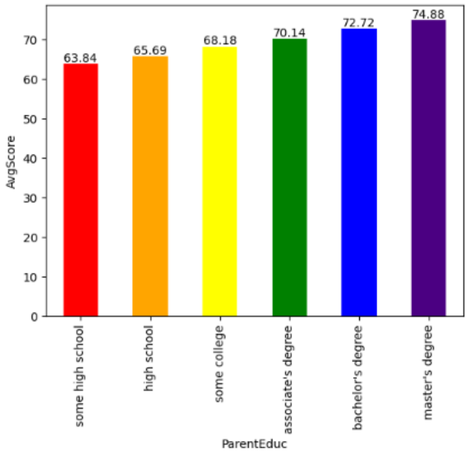

WritingScore
0
10
20
30
40
50
60
70
68.3
68.42
68.49
68.5
68.58
68.79
68.96
70.31
6.0
1.0
2.0
4.0
5.0
3.0
0.0
7.0
NrSiblings

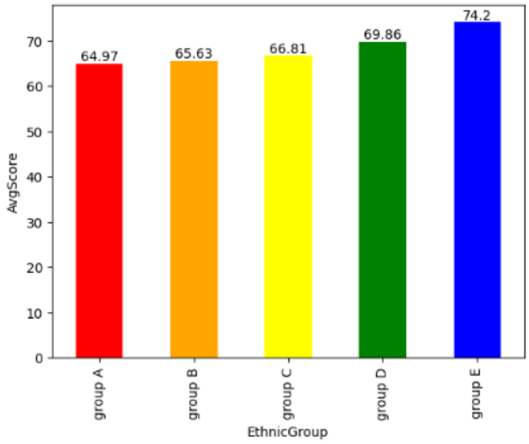

WritingScore
0
10
20
30
40
50
60
70
68.59
68.61
private
school_bus
TransportMeans

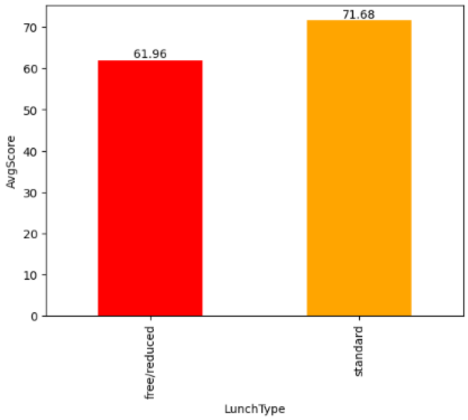

WritingScore
0
10
20
30
40
50
60
70
67.36
68.79
70.0
< 5
5 - 10
> 10
WklyStudyHours

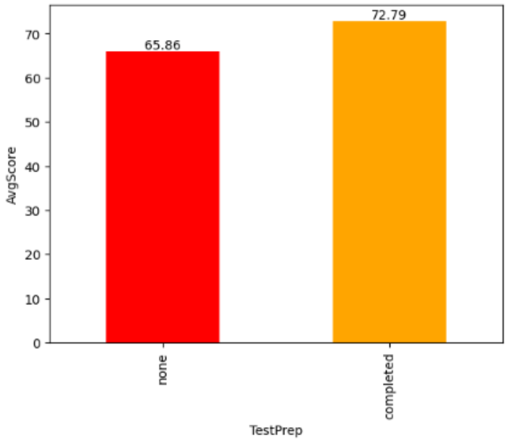
4. Output for Average Score:

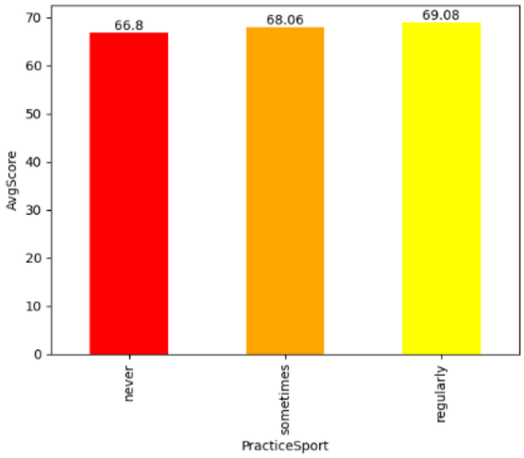


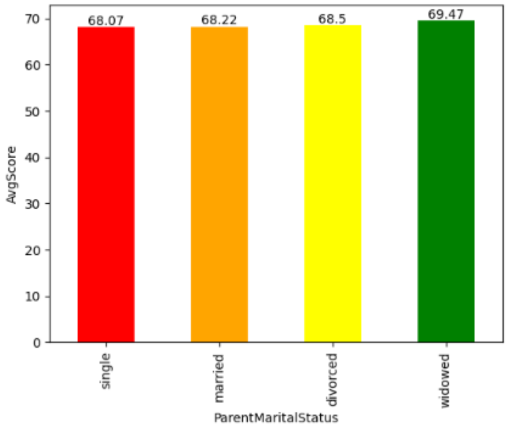


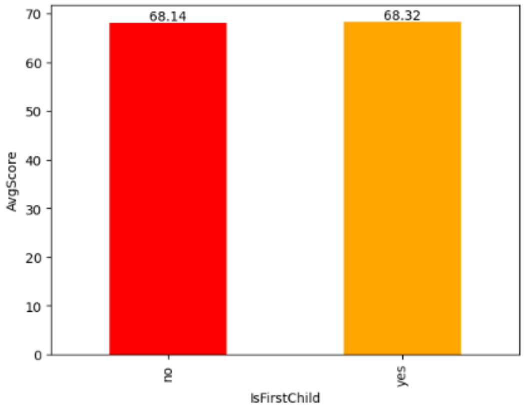


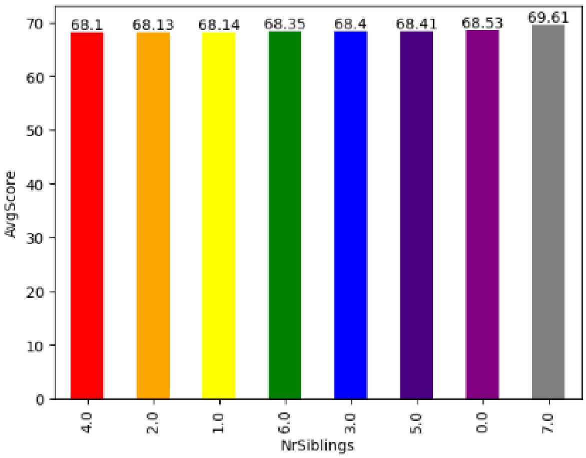


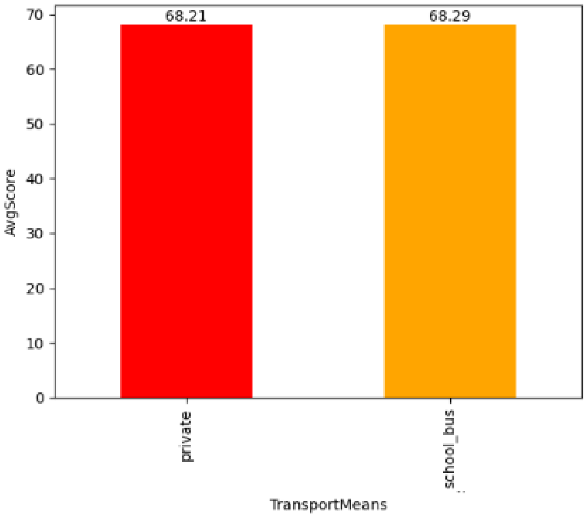


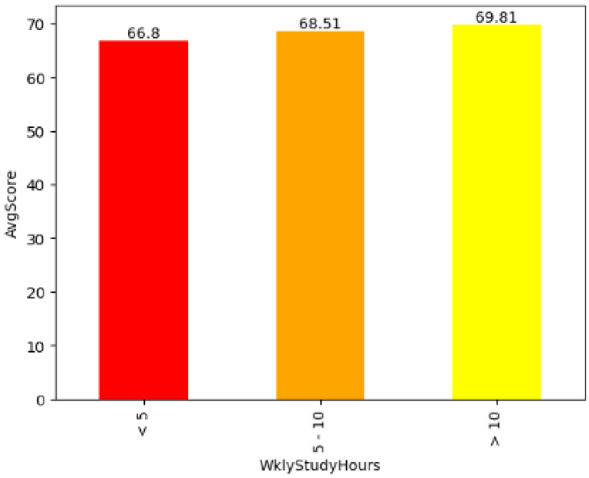







68.1
68.13
68.14
68.35
68.4
68.41
68.53
69.61
0
10
20
30
40
50
60
70
AvgScore
4.0
2.0
1.0
6.0
3.0
5.0
0.0
7.0
NrSiblings


68.21
68.29
0
10
20
30
40
50
60
70
AvgScore
private
school_bus
TransportMeans


66.8
68.51
69.81
0
10
20
30
40
50
60
70
AvgScore
< 5
5 - 10
> 10
WklyStudyHours

4.0 Dataset

Kaggle: Students Exam Scores: Extended Dataset
Link:

Students Exam Scores: Extended Dataset | Kaggle
This dataset includes scores from three test scores of students at a (fictional) public school and a variety of personal and socio-economic factors that may have interaction effects upon them.

The variables in the dataset include:

- Gender
- Ethnic Group
- Parent Education
- Lunch Type
- Test Preparation
- Parent Marital Status
- Practice Sport
- Is First Child
- Number Siblings
- Transportation Mean
- Weekly Study Hours
- Math Score
- Reading Score
- Writing Score

5.0 Tools and Programming Language

Tool(s)	Google Colab
Programming Language	Python