```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import math
from scipy import stats

data = pd.read_csv("academic_performance.csv")

data
```

```
      1st   2nd   3rd   4th   5th  College Code  Gender     Roll  Roll
no.  \
0    8.11  7.68  7.11  7.43  8.18           115  Female      NaN
17020.0
1    6.48  5.90  4.15  4.29  4.96           115    Male      NaN
17021.0
2    8.41  8.24  7.52  8.25  7.75           115  Female      NaN
17022.0
3    7.33  6.83  6.33  6.79  6.89           115    Male      NaN
17023.0
4    7.89  7.34  7.22  7.32  7.46           115    Male      NaN
17024.0
..    ...   ...   ...   ...   ...           ...     ...      ...
...
173  7.48  7.55  7.67  7.39  8.65           241       F  17048.0
NaN
174  7.30  6.41  6.59  7.11  7.38           241       M  17049.0
NaN
175  6.30  6.28  5.89  5.71  6.50           241       M  17050.0
NaN
176  7.04  7.10  6.81  7.00  6.92           241       M  17051.0
NaN
177  6.70  6.81  6.52  5.39  7.00           241       M  17052.0
NaN

     Subject Code
0              16
1              16
2              16
3              16
4              16
..            ...
173            28
174            28
175            28
176            28
177            28

[178 rows x 10 columns]
```

```python
print("The first five rows are as follows: ")
data.head()
```

The first five rows are as follows:

|   | 1st | 2nd | 3rd | 4th | 5th | College Code | Gender | Roll | Roll no. |
|---|-----|-----|-----|-----|-----|--------------|--------|------|----------|
| 0 | 8.11 | 7.68 | 7.11 | 7.43 | 8.18 | 115 | Female | NaN | 17020.0 |
| 1 | 6.48 | 5.90 | 4.15 | 4.29 | 4.96 | 115 | Male | NaN | 17021.0 |
| 2 | 8.41 | 8.24 | 7.52 | 8.25 | 7.75 | 115 | Female | NaN | 17022.0 |
| 3 | 7.33 | 6.83 | 6.33 | 6.79 | 6.89 | 115 | Male | NaN | 17023.0 |
| 4 | 7.89 | 7.34 | 7.22 | 7.32 | 7.46 | 115 | Male | NaN | 17024.0 |

|   | Subject Code |
|---|--------------|
| 0 | 16 |
| 1 | 16 |
| 2 | 16 |
| 3 | 16 |
| 4 | 16 |

```python
print("The last five rows are as follows: ")
data.tail()
```

The last five rows are as follows:

|   | 1st | 2nd | 3rd | 4th | 5th | College Code | Gender | Roll | Roll no. |
|---|-----|-----|-----|-----|-----|--------------|--------|------|----------|
| 173 | 7.48 | 7.55 | 7.67 | 7.39 | 8.65 | 241 | F | 17048.0 | NaN |
| 174 | 7.30 | 6.41 | 6.59 | 7.11 | 7.38 | 241 | M | 17049.0 | NaN |
| 175 | 6.30 | 6.28 | 5.89 | 5.71 | 6.50 | 241 | M | 17050.0 | NaN |
| 176 | 7.04 | 7.10 | 6.81 | 7.00 | 6.92 | 241 | M | 17051.0 | NaN |
| 177 | 6.70 | 6.81 | 6.52 | 5.39 | 7.00 | 241 | M | 17052.0 | NaN |

|   | Subject Code |
|---|--------------|
| 173 | 28 |
| 174 | 28 |
| 175 | 28 |
| 176 | 28 |
| 177 | 28 |

```python
data.describe()
```

```
            1st          2nd         3rd          4th          5th  \
count  176.000000  174.000000  176.00000  173.000000  172.000000
mean     7.038864    6.943391    6.62250    7.027746    7.432558
std      0.859521    0.881088    0.91926    1.035675    1.068866
min      3.850000    3.900000    3.96000    4.290000    4.000000
25%      6.440000    6.380000    6.14000    6.330000    6.750000
50%      7.070000    6.800000    6.68500    7.070000    7.590000
75%      7.590000    7.612500    7.26000    7.860000    8.210000
max      9.150000    9.210000    9.59000    9.310000    9.460000

       College Code          Roll      Roll no.   Subject Code
count    178.000000   132.000000     46.000000     178.000000
mean     152.983146  17040.666667  17042.500000    18.359551
std       48.357599     13.622874     13.422618     4.782849
min      115.000000  17017.000000  17020.000000    16.000000
25%      115.000000  17029.000000  17031.250000    16.000000
50%      116.000000  17040.000000  17042.500000    16.000000
75%      165.000000  17051.000000  17053.750000    16.000000
max      241.000000  17067.000000  17065.000000    28.000000
```

```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   1st           176 non-null    float64
 1   2nd           174 non-null    float64
 2   3rd           176 non-null    float64
 3   4th           173 non-null    float64
 4   5th           172 non-null    float64
 5   College Code  178 non-null    int64
 6   Gender        177 non-null    object
 7   Roll          132 non-null    float64
 8   Roll no.      46 non-null     float64
 9   Subject Code  178 non-null    int64
dtypes: float64(7), int64(2), object(1)
memory usage: 14.0+ KB
```

```python
print("The column names of the dataset are as follows: ")
data.columns
```

```
The column names of the dataset are as follows:

Index(['1st', '2nd', '3rd', '4th', '5th', 'College Code', 'Gender',
'Roll',
       'Roll no.', 'Subject Code'],
      dtype='object')
```

```python
data.isnull().sum()
```

```
1st                 2
2nd                 4
3rd                 2
4th                 5
5th                 6
College Code        0
Gender              1
Roll               46
Roll no.          132
Subject Code        0
dtype: int64
```

```python
data.shape
```

```
(178, 10)
```

```python
m_v=data['1st'].mean()
data['1st']=data['1st'].fillna(value=m_v)

m_v=data['2nd'].mean()
data['2nd']=data['2nd'].fillna(value=m_v)

m_v=data['3rd'].mean()
data['3rd']=data['3rd'].fillna(value=m_v)

m_v=data['4th'].mean()
data['4th']=data['4th'].fillna(value=m_v)

m_v=data['5th'].mean()
data['5th']=data['5th'].fillna(value=m_v)

m_v=data['Roll'].mean()
data['Roll']=data['Roll'].fillna(value=m_v)

m_v=data['Roll no.'].mean()
data['Roll no.']=data['Roll no.'].fillna(value=m_v)

data['Gender'] = data['Gender'].ffill()

data
```

```
     1st   2nd   3rd   4th   5th  College Code  Gender          Roll
\
0   8.11  7.68  7.11  7.43  8.18           115  Female  17040.666667

1   6.48  5.90  4.15  4.29  4.96           115    Male  17040.666667

2   8.41  8.24  7.52  8.25  7.75           115  Female  17040.666667

3   7.33  6.83  6.33  6.79  6.89           115    Male  17040.666667

4   7.89  7.34  7.22  7.32  7.46           115    Male  17040.666667
```

|     |      |      |      |      |      |     |     |     |              |
| --- | ---- | ---- | ---- | ---- | ---- | --- | --- | --- | ------------ |
| ..  | ...  | ...  | ...  | ...  | ...  |     | ... | ... |          ... |
| 173 | 7.48 | 7.55 | 7.67 | 7.39 | 8.65 |     | 241 | F   | 17048.000000 |
| 174 | 7.30 | 6.41 | 6.59 | 7.11 | 7.38 |     | 241 | M   | 17049.000000 |
| 175 | 6.30 | 6.28 | 5.89 | 5.71 | 6.50 |     | 241 | M   | 17050.000000 |
| 176 | 7.04 | 7.10 | 6.81 | 7.00 | 6.92 |     | 241 | M   | 17051.000000 |
| 177 | 6.70 | 6.81 | 6.52 | 5.39 | 7.00 |     | 241 | M   | 17052.000000 |

```
     Roll no.  Subject Code
0     17020.0            16
1     17021.0            16
2     17022.0            16
3     17023.0            16
4     17024.0            16
..        ...           ...
173   17042.5            28
174   17042.5            28
175   17042.5            28
176   17042.5            28
177   17042.5            28

[178 rows x 10 columns]
```

```
data.isnull().sum()
```

```
1st             0
2nd             0
3rd             0
4th             0
5th             0
College Code    0
Gender          0
Roll            0
Roll no.        0
Subject Code    0
dtype: int64
```
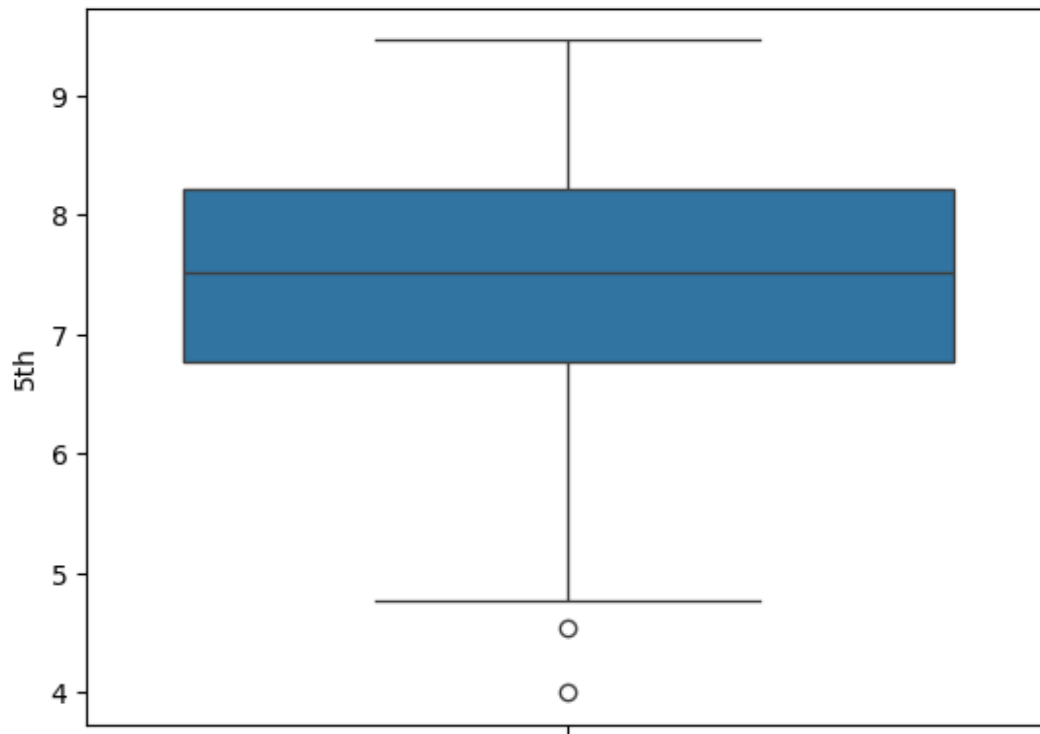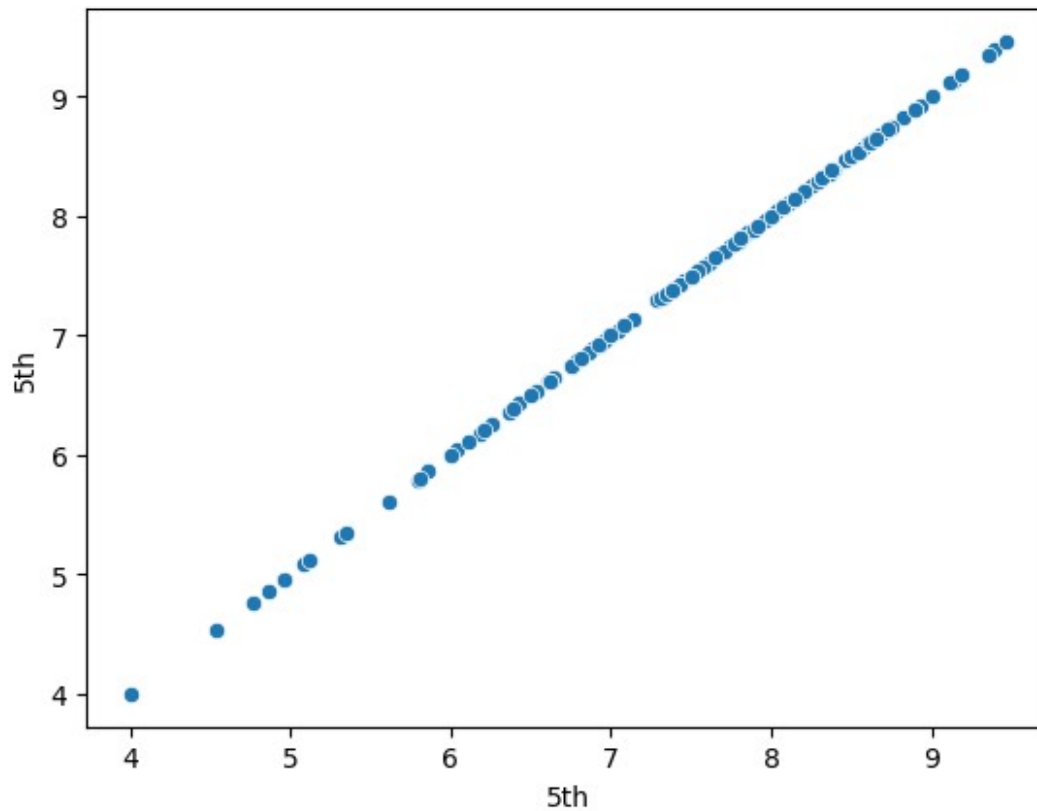
```
sns.boxplot(y=data['5th'])
```

```
<Axes: ylabel='5th'>
```

```
sns.scatterplot(x=data['5th'], y=data['5th'])
plt.show()
```

```
z = np.abs(stats.zscore(data['5th']))
print(z)

0        0.713454
1        2.360125
2        0.303007
3        0.517887
4        0.026194
            ...
173      1.162082
174      0.050168
175      0.890153
176      0.489251
177      0.412889
Name: 5th, Length: 178, dtype: float64

data = data[z < 2.0]

data
```

| | 1st | 2nd | 3rd | 4th | 5th | College Code | Gender | Roll |
|---|---|---|---|---|---|---|---|---|
| 0 | 8.11 | 7.68 | 7.11 | 7.43 | 8.18 | 115 | Female | 17040.666667 |
| 2 | 8.41 | 8.24 | 7.52 | 8.25 | 7.75 | 115 | Female | 17040.666667 |

| | 1st | 2nd | 3rd | 4th | 5th | College Code | Gender | Roll |
|---|---|---|---|---|---|---|---|---|
| 3 | 7.33 | 6.83 | 6.33 | 6.79 | 6.89 | 115 | Male | 17040.666667 |
| 4 | 7.89 | 7.34 | 7.22 | 7.32 | 7.46 | 115 | Male | 17040.666667 |
| 5 | 7.33 | 6.72 | 6.48 | 6.50 | 7.39 | 115 | Male | 17040.666667 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 173 | 7.48 | 7.55 | 7.67 | 7.39 | 8.65 | 241 | F | 17048.000000 |
| 174 | 7.30 | 6.41 | 6.59 | 7.11 | 7.38 | 241 | M | 17049.000000 |
| 175 | 6.30 | 6.28 | 5.89 | 5.71 | 6.50 | 241 | M | 17050.000000 |
| 176 | 7.04 | 7.10 | 6.81 | 7.00 | 6.92 | 241 | M | 17051.000000 |
| 177 | 6.70 | 6.81 | 6.52 | 5.39 | 7.00 | 241 | M | 17052.000000 |

```
     Roll no.  Subject Code
0     17020.0            16
2     17022.0            16
3     17023.0            16
4     17024.0            16
5     17025.0            16
..        ...           ...
173   17042.5            28
174   17042.5            28
175   17042.5            28
176   17042.5            28
177   17042.5            28

[169 rows x 10 columns]
```

```python
Q1 = data['5th'].quantile(0.25)
Q3 = data['5th'].quantile(0.75)

IQR = Q3 - Q1

lower_quartile = Q1 - 1.5 * IQR
upper_quartile = Q3 + 1.5 * IQR

data = data[(data['5th'] >= lower_quartile) & (data['5th'] <=
upper_quartile)]

data
```

| | 1st | 2nd | 3rd | 4th | 5th | College Code | Gender | Roll |
|---|---|---|---|---|---|---|---|---|
| 0 | 8.11 | 7.68 | 7.11 | 7.43 | 8.18 | 115 | Female | 17040.666667 |

```
2     8.41  8.24  7.52  8.25  7.75           115   Female  17040.666667

3     7.33  6.83  6.33  6.79  6.89           115     Male  17040.666667

4     7.89  7.34  7.22  7.32  7.46           115     Male  17040.666667

5     7.33  6.72  6.48  6.50  7.39           115     Male  17040.666667

..    ...   ...   ...   ...   ...            ...      ...           ...

173   7.48  7.55  7.67  7.39  8.65           241        F  17048.000000

174   7.30  6.41  6.59  7.11  7.38           241        M  17049.000000

175   6.30  6.28  5.89  5.71  6.50           241        M  17050.000000

176   7.04  7.10  6.81  7.00  6.92           241        M  17051.000000

177   6.70  6.81  6.52  5.39  7.00           241        M  17052.000000
```

```
      Roll no.  Subject Code
0      17020.0            16
2      17022.0            16
3      17023.0            16
4      17024.0            16
5      17025.0            16
..         ...           ...
173    17042.5            28
174    17042.5            28
175    17042.5            28
176    17042.5            28
177    17042.5            28
```

```
[169 rows x 10 columns]
```

```python
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

data.loc[:, '5th'] = scaler.fit_transform(data[['5th']])

data
```

```
       1st   2nd   3rd   4th        5th  College Code   Gender
Roll  \
0     8.11  7.68  7.11  7.43  0.688564           115   Female
17040.666667
2     8.41  8.24  7.52  8.25  0.583942           115   Female
17040.666667
3     7.33  6.83  6.33  6.79  0.374696           115     Male
```

```
     17040.666667
4    7.89  7.34  7.22  7.32  0.513382            115    Male
     17040.666667
5    7.33  6.72  6.48  6.50  0.496350            115    Male
     17040.666667
..   ...   ...   ...   ...        ...            ...     ...
     ...
173  7.48  7.55  7.67  7.39  0.802920            241       F
     17048.000000
174  7.30  6.41  6.59  7.11  0.493917            241       M
     17049.000000
175  6.30  6.28  5.89  5.71  0.279805            241       M
     17050.000000
176  7.04  7.10  6.81  7.00  0.381995            241       M
     17051.000000
177  6.70  6.81  6.52  5.39  0.401460            241       M
     17052.000000

     Roll no.  Subject Code
0     17020.0            16
2     17022.0            16
3     17023.0            16
4     17024.0            16
5     17025.0            16
..        ...           ...
173   17042.5            28
174   17042.5            28
175   17042.5            28
176   17042.5            28
177   17042.5            28

[169 rows x 10 columns]
```