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

df = pd.read_csv("xAPI-Edu-Data.csv")

df
```

```
    gender NationalITy PlaceofBirth       StageID GradeID SectionID  \
0        M          KW       KuwaIT    lowerlevel    G-04         A
1        M          KW       KuwaIT    lowerlevel    G-04         A
2        M          KW       KuwaIT    lowerlevel    G-04         A
3        M          KW       KuwaIT    lowerlevel    G-04         A
4        M          KW       KuwaIT    lowerlevel    G-04         A
..     ...         ...          ...           ...     ...       ...
475      F      Jordan       Jordan  MiddleSchool    G-08         A
476      F      Jordan       Jordan  MiddleSchool    G-08         A
477      F      Jordan       Jordan  MiddleSchool    G-08         A
478      F      Jordan       Jordan  MiddleSchool    G-08         A
479      F      Jordan       Jordan  MiddleSchool    G-08         A

         Topic Semester Relation  raisedhands  VisITedResources  \
0           IT        F   Father           15                16
1           IT        F   Father           20                20
2           IT        F   Father           10                 7
3           IT        F   Father           30                25
4           IT        F   Father           40                50
..         ...      ...      ...          ...               ...
475  Chemistry        S   Father            5                 4
476    Geology        F   Father           50                77
477    Geology        S   Father           55                74
478    History        F   Father           30                17
479    History        S   Father           35                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
0                    2          20                   Yes
1                    3          25                   Yes
2                    0          30                    No
3                    5          35                    No
4                   12          50                    No
..                 ...         ...                   ...
475                  5           8                    No
476                 14          28                    No
477                 25          29                    No
478                 14          57                    No
479                 23          62                    No

    ParentschoolSatisfaction StudentAbsenceDays Class
```

```
0                       Good            Under-7     M
1                       Good            Under-7     M
2                        Bad            Above-7     L
3                        Bad            Above-7     L
4                        Bad            Above-7     M
..                       ...                ...   ...
475                      Bad            Above-7     L
476                      Bad            Under-7     M
477                      Bad            Under-7     M
478                      Bad            Above-7     L
479                      Bad            Above-7     L

[480 rows x 17 columns]
```

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

```
The first five rows are as follows:

  gender NationalITy PlaceofBirth     StageID GradeID SectionID Topic
\
0      M          KW       KuwaIT  lowerlevel    G-04         A    IT

1      M          KW       KuwaIT  lowerlevel    G-04         A    IT

2      M          KW       KuwaIT  lowerlevel    G-04         A    IT

3      M          KW       KuwaIT  lowerlevel    G-04         A    IT

4      M          KW       KuwaIT  lowerlevel    G-04         A    IT


  Semester Relation  raisedhands  VisITedResources  AnnouncementsView
\
0        F   Father           15                16                  2

1        F   Father           20                20                  3

2        F   Father           10                 7                  0

3        F   Father           30                25                  5

4        F   Father           40                50                 12


   Discussion ParentAnsweringSurvey ParentschoolSatisfaction  \
0          20                   Yes                     Good
1          25                   Yes                     Good
2          30                    No                      Bad
3          35                    No                      Bad
4          50                    No                      Bad
```

```
  StudentAbsenceDays Class
0            Under-7     M
1            Under-7     M
2            Above-7     L
3            Above-7     L
4            Above-7     M
```

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

```
The last five rows are as follows:
```

```
    gender NationalITy PlaceofBirth       StageID GradeID SectionID  \
475      F      Jordan       Jordan  MiddleSchool    G-08         A
476      F      Jordan       Jordan  MiddleSchool    G-08         A
477      F      Jordan       Jordan  MiddleSchool    G-08         A
478      F      Jordan       Jordan  MiddleSchool    G-08         A
479      F      Jordan       Jordan  MiddleSchool    G-08         A

         Topic Semester Relation  raisedhands  VisITedResources  \
475  Chemistry        S   Father            5                 4
476    Geology        F   Father           50                77
477    Geology        S   Father           55                74
478    History        F   Father           30                17
479    History        S   Father           35                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
475                  5           8                    No
476                 14          28                    No
477                 25          29                    No
478                 14          57                    No
479                 23          62                    No

    ParentschoolSatisfaction StudentAbsenceDays Class
475                      Bad            Above-7     L
476                      Bad            Under-7     M
477                      Bad            Under-7     M
478                      Bad            Above-7     L
479                      Bad            Above-7     L
```

```
df.describe()
```

```
       raisedhands  VisITedResources  AnnouncementsView  Discussion
count   480.000000        480.000000         480.000000  480.000000
mean     47.097917         54.797917          37.918750   43.283333
std      40.299256         33.080007          26.611244   27.637735
min    -300.000000          0.000000           0.000000    1.000000
25%      15.750000         20.000000          14.000000   20.000000
50%      50.000000         65.000000          33.000000   39.000000
```

```
75%      75.000000          84.000000          58.000000   70.000000
max     500.000000          99.000000          98.000000   99.000000
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 480 entries, 0 to 479
Data columns (total 17 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   gender                    480 non-null    object
 1   NationalITy               480 non-null    object
 2   PlaceofBirth              480 non-null    object
 3   StageID                   480 non-null    object
 4   GradeID                   480 non-null    object
 5   SectionID                 480 non-null    object
 6   Topic                     480 non-null    object
 7   Semester                  480 non-null    object
 8   Relation                  480 non-null    object
 9   raisedhands               480 non-null    int64
 10  VisITedResources          480 non-null    int64
 11  AnnouncementsView         480 non-null    int64
 12  Discussion                480 non-null    int64
 13  ParentAnsweringSurvey     480 non-null    object
 14  ParentschoolSatisfaction  480 non-null    object
 15  StudentAbsenceDays        480 non-null    object
 16  Class                     480 non-null    object
dtypes: int64(4), object(13)
memory usage: 63.9+ KB
```

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

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

```
Index(['gender', 'NationalITy', 'PlaceofBirth', 'StageID', 'GradeID',
       'SectionID', 'Topic', 'Semester', 'Relation', 'raisedhands',
       'VisITedResources', 'AnnouncementsView', 'Discussion',
       'ParentAnsweringSurvey', 'ParentschoolSatisfaction',
       'StudentAbsenceDays', 'Class'],
      dtype='object')
```

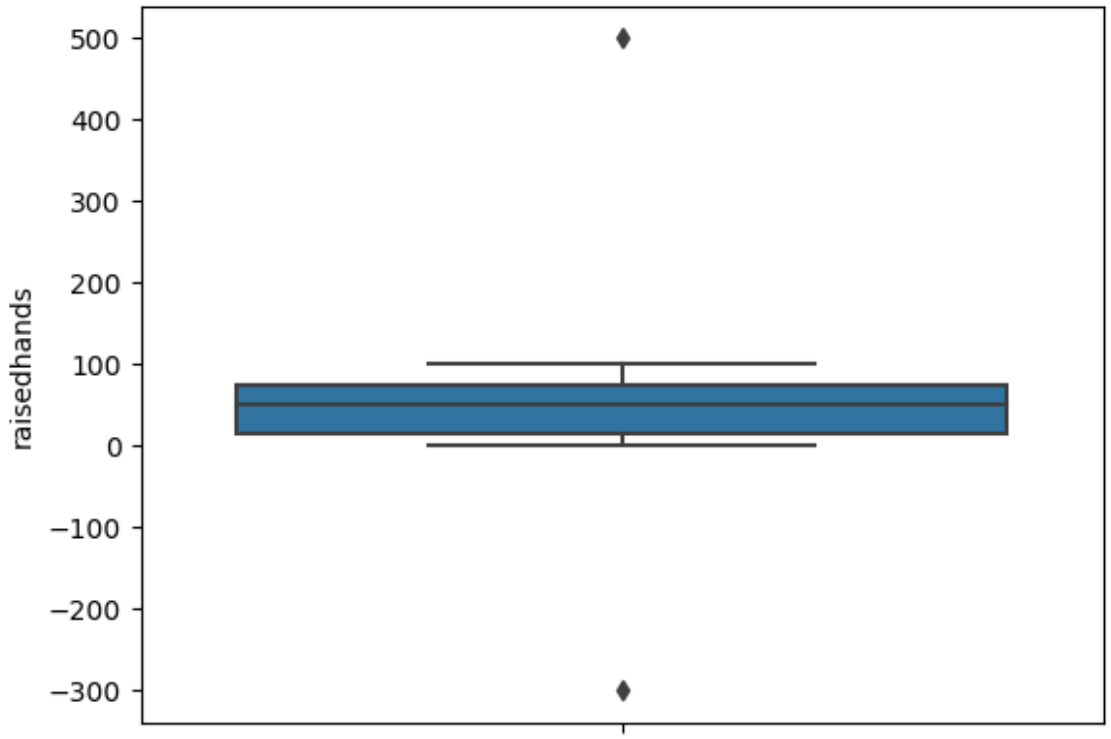
```
df.isnull().sum()
```

```
gender                      0
NationalITy                 0
PlaceofBirth                0
StageID                     0
GradeID                     0
SectionID                   0
Topic                       0
```

```
Semester                    0
Relation                    0
raisedhands                 0
VisITedResources            0
AnnouncementsView           0
Discussion                  0
ParentAnsweringSurvey       0
ParentschoolSatisfaction    0
StudentAbsenceDays          0
Class                       0
dtype: int64
```

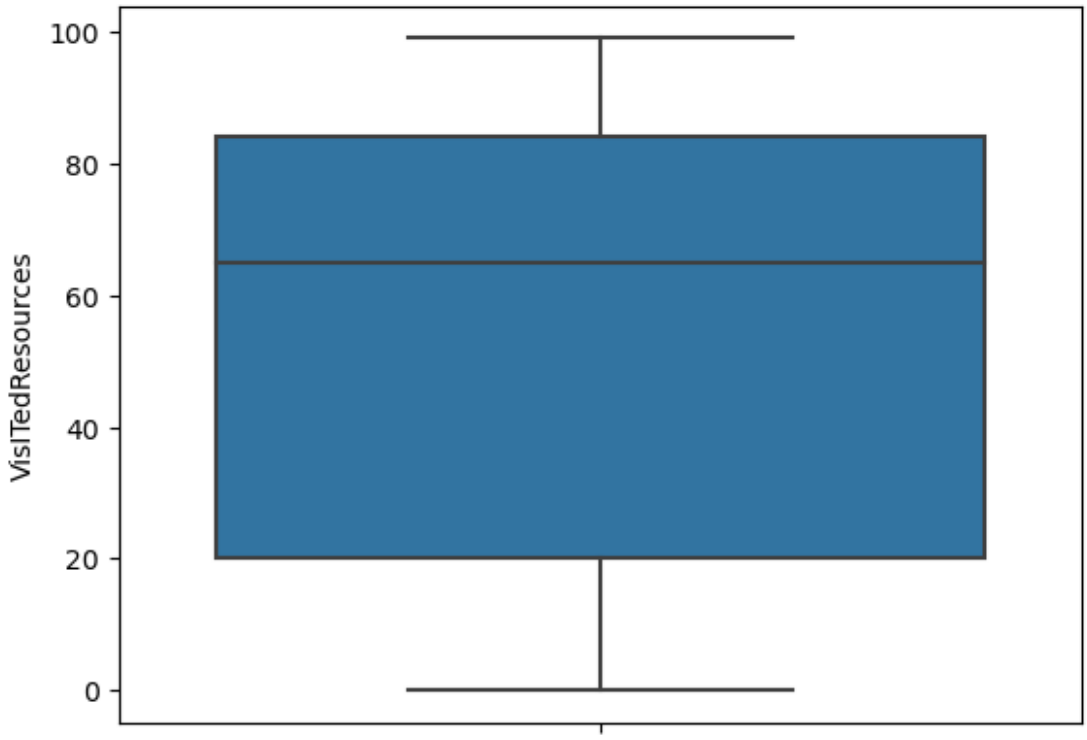
```
sns.boxplot(y=df['raisedhands'])
```

```
<Axes: ylabel='raisedhands'>
```



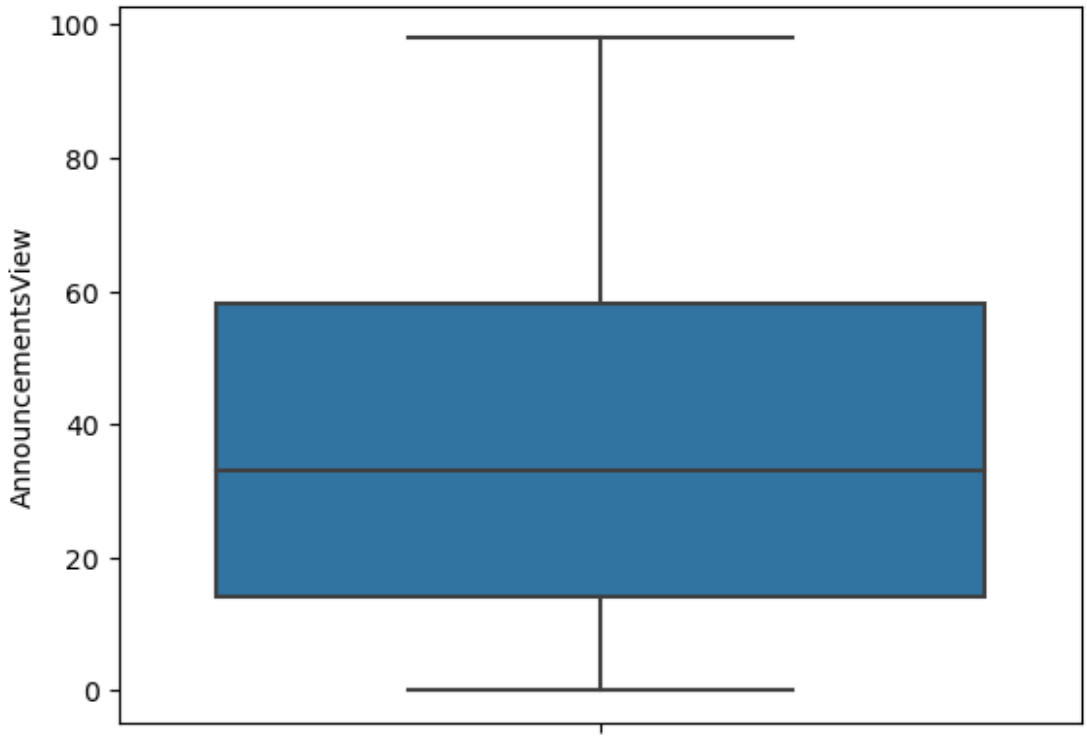

```
sns.boxplot(y=df['VisITedResources'])
```

```
<Axes: ylabel='VisITedResources'>
```



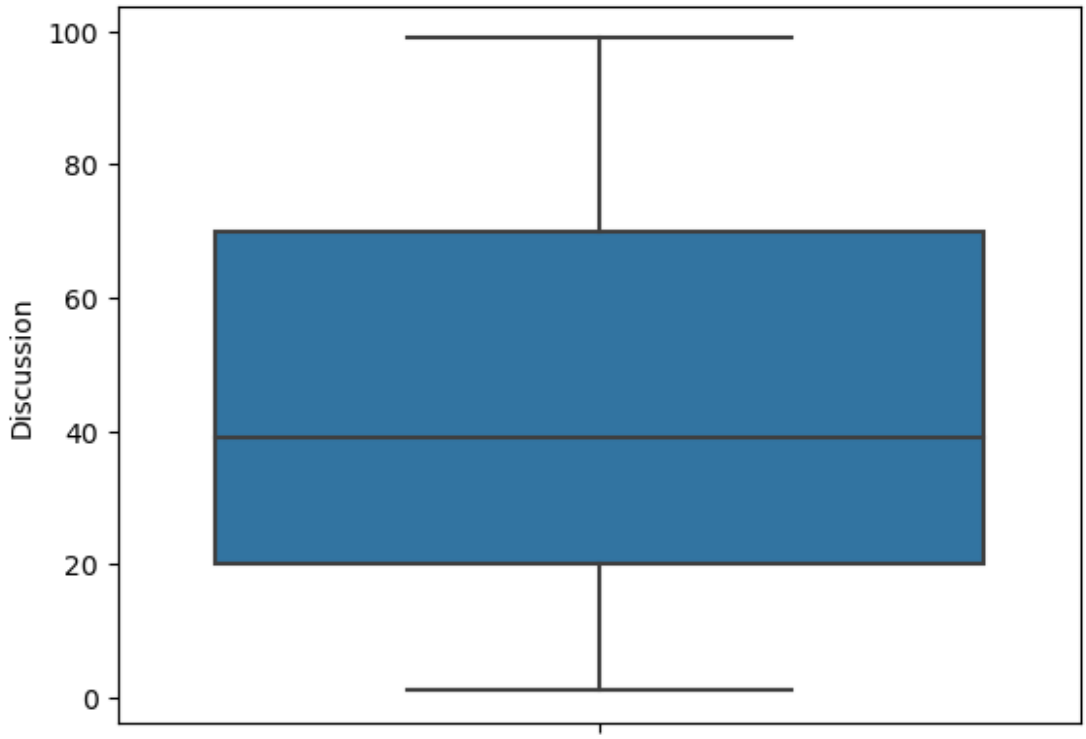

```
sns.boxplot(y=df["AnnouncementsView"])
```

```
<Axes: ylabel='AnnouncementsView'>
```

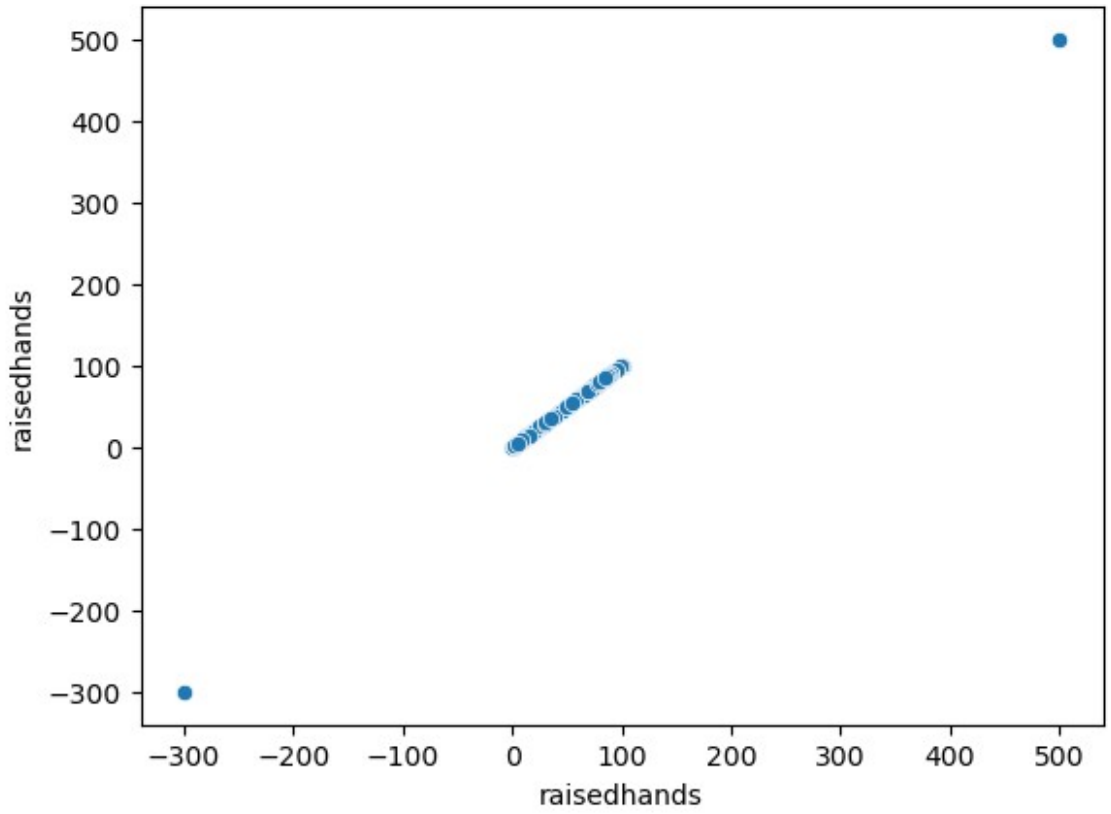


```
sns.boxplot(y=df["Discussion"])
```

```
<Axes: ylabel='Discussion'>
```




```
sns.scatterplot(x=df['raisedhands'], y=df['raisedhands'])
plt.show()
```




```
z = np.abs(stats.zscore(df['raisedhands']))

print(z)
```

```
0      0.797320
1      0.673119
2      0.921521
3      0.424716
4      0.176314
         ...
475    1.045722
476    0.072088
477    0.196290
478    0.424716
479    0.300515
Name: raisedhands, Length: 480, dtype: float64
```

```
df = df[z < 2.0]

df
```

```
    gender NationalITy PlaceofBirth       StageID GradeID SectionID  \
0        M          KW       KuwaIT    lowerlevel    G-04         A
1        M          KW       KuwaIT    lowerlevel    G-04         A
2        M          KW       KuwaIT    lowerlevel    G-04         A
```

```
3        M          KW       KuwaIT    lowerlevel    G-04          A
4        M          KW       KuwaIT    lowerlevel    G-04          A
..     ...         ...          ...           ...     ...        ...
475      F      Jordan       Jordan  MiddleSchool    G-08          A
476      F      Jordan       Jordan  MiddleSchool    G-08          A
477      F      Jordan       Jordan  MiddleSchool    G-08          A
478      F      Jordan       Jordan  MiddleSchool    G-08          A
479      F      Jordan       Jordan  MiddleSchool    G-08          A

          Topic Semester Relation  raisedhands  VisITedResources  \
0            IT        F   Father           15                16
1            IT        F   Father           20                20
2            IT        F   Father           10                 7
3            IT        F   Father           30                25
4            IT        F   Father           40                50
..          ...      ...      ...          ...               ...
475  Chemistry        S   Father            5                 4
476    Geology        F   Father           50                77
477    Geology        S   Father           55                74
478    History        F   Father           30                17
479    History        S   Father           35                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
0                    2          20                   Yes
1                    3          25                   Yes
2                    0          30                    No
3                    5          35                    No
4                   12          50                    No
..                 ...         ...                   ...
475                  5           8                    No
476                 14          28                    No
477                 25          29                    No
478                 14          57                    No
479                 23          62                    No

    ParentschoolSatisfaction StudentAbsenceDays Class
0                       Good            Under-7     M
1                       Good            Under-7     M
2                        Bad            Above-7     L
3                        Bad            Above-7     L
4                        Bad            Above-7     M
..                       ...                ...   ...
475                      Bad            Above-7     L
476                      Bad            Under-7     M
477                      Bad            Under-7     M
478                      Bad            Above-7     L
479                      Bad            Above-7     L

[478 rows x 17 columns]
```

```
Q1 = df['raisedhands'].quantile(0.25)
Q3 = df['raisedhands'].quantile(0.75)

IQR = Q3 - Q1

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

df = df[(df['raisedhands'] >= lower_quartile) & (df['raisedhands'] <=
upper_quartile)]

df
```

```
    gender NationalITy PlaceofBirth       StageID GradeID SectionID  \
0        M          KW       KuwaIT    lowerlevel    G-04         A
1        M          KW       KuwaIT    lowerlevel    G-04         A
2        M          KW       KuwaIT    lowerlevel    G-04         A
3        M          KW       KuwaIT    lowerlevel    G-04         A
4        M          KW       KuwaIT    lowerlevel    G-04         A
..     ...         ...          ...           ...     ...       ...
475      F      Jordan       Jordan  MiddleSchool    G-08         A
476      F      Jordan       Jordan  MiddleSchool    G-08         A
477      F      Jordan       Jordan  MiddleSchool    G-08         A
478      F      Jordan       Jordan  MiddleSchool    G-08         A
479      F      Jordan       Jordan  MiddleSchool    G-08         A

         Topic Semester Relation  raisedhands  VisITedResources  \
0           IT        F   Father           15                16
1           IT        F   Father           20                20
2           IT        F   Father           10                 7
3           IT        F   Father           30                25
4           IT        F   Father           40                50
..         ...      ...      ...          ...               ...
475  Chemistry        S   Father            5                 4
476    Geology        F   Father           50                77
477    Geology        S   Father           55                74
478    History        F   Father           30                17
479    History        S   Father           35                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
0                    2          20                   Yes
1                    3          25                   Yes
2                    0          30                    No
3                    5          35                    No
4                   12          50                    No
..                 ...         ...                   ...
475                  5           8                    No
476                 14          28                    No
477                 25          29                    No
478                 14          57                    No
479                 23          62                    No
```

```
    ParentschoolSatisfaction StudentAbsenceDays Class
0                       Good            Under-7     M
1                       Good            Under-7     M
2                        Bad            Above-7     L
3                        Bad            Above-7     L
4                        Bad            Above-7     M
..                       ...                ...   ...
475                      Bad            Above-7     L
476                      Bad            Under-7     M
477                      Bad            Under-7     M
478                      Bad            Above-7     L
479                      Bad            Above-7     L

[478 rows x 17 columns]
```

```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

df.loc[:, 'raisedhands'] = scaler.fit_transform(df[['raisedhands']])

df
```

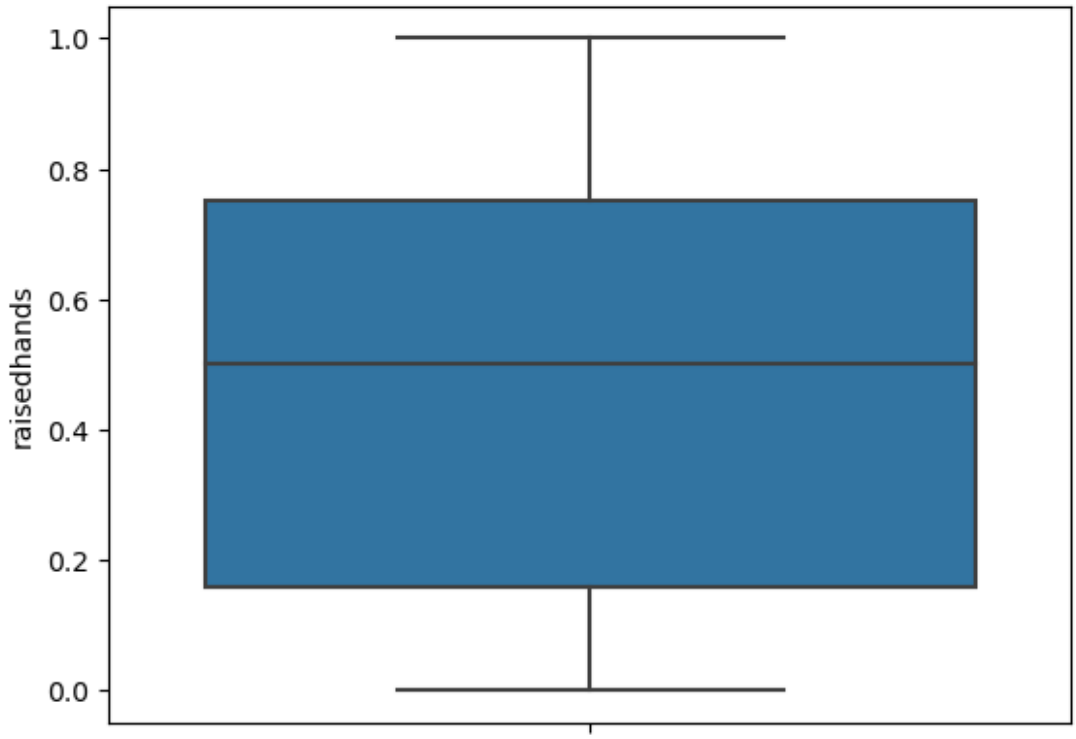
```
    gender NationalITy PlaceofBirth       StageID GradeID SectionID  \
0        M          KW       KuwaIT    lowerlevel    G-04         A
1        M          KW       KuwaIT    lowerlevel    G-04         A
2        M          KW       KuwaIT    lowerlevel    G-04         A
3        M          KW       KuwaIT    lowerlevel    G-04         A
4        M          KW       KuwaIT    lowerlevel    G-04         A
..     ...         ...          ...           ...     ...       ...
475      F      Jordan       Jordan  MiddleSchool    G-08         A
476      F      Jordan       Jordan  MiddleSchool    G-08         A
477      F      Jordan       Jordan  MiddleSchool    G-08         A
478      F      Jordan       Jordan  MiddleSchool    G-08         A
479      F      Jordan       Jordan  MiddleSchool    G-08         A

         Topic Semester Relation  raisedhands  VisITedResources  \
0           IT        F   Father         0.15                16
1           IT        F   Father         0.20                20
2           IT        F   Father         0.10                 7
3           IT        F   Father         0.30                25
4           IT        F   Father         0.40                50
..         ...      ...      ...          ...               ...
475  Chemistry        S   Father         0.05                 4
476    Geology        F   Father         0.50                77
477    Geology        S   Father         0.55                74
478    History        F   Father         0.30                17
479    History        S   Father         0.35                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
```

```
0                    2          20                   Yes
1                    3          25                   Yes
2                    0          30                    No
3                    5          35                    No
4                   12          50                    No
..                 ...         ...                  ...
475                  5           8                    No
476                 14          28                    No
477                 25          29                    No
478                 14          57                    No
479                 23          62                    No
```

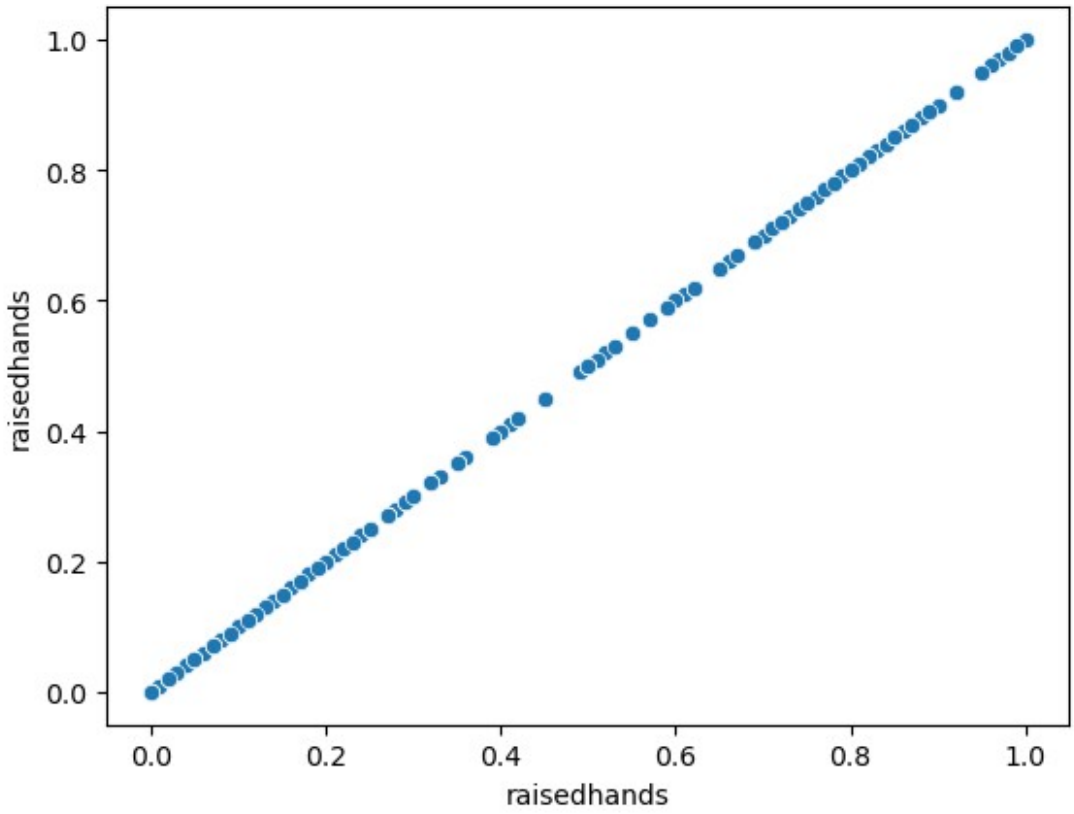
```
    ParentschoolSatisfaction StudentAbsenceDays Class
0                       Good            Under-7     M
1                       Good            Under-7     M
2                        Bad            Above-7     L
3                        Bad            Above-7     L
4                        Bad            Above-7     M
..                       ...                ...   ...
475                      Bad            Above-7     L
476                      Bad            Under-7     M
477                      Bad            Under-7     M
478                      Bad            Above-7     L
479                      Bad            Above-7     L

[478 rows x 17 columns]
```

```
sns.boxplot(y=df['raisedhands'])
```

```
<Axes: ylabel='raisedhands'>
```




```
sns.scatterplot(x=df['raisedhands'], y=df['raisedhands'])
plt.show()
```


1.0
0.8
0.6
0.4
0.2
0.0
raisedhands
0.0
0.2
0.4
0.6
0.8
1.0
raisedhands