In [1]:
```
import pandas as pd;
import numpy as ny;
import seaborn as sns;
```

In [2]:
```
data = pd.read_csv('xAPI-Edu-Data.csv')
```

In [3]:
```
data.head()
```

Out[3]:

	gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation
0	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father
1	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father
2	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father
3	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father
4	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father

In [4]:
```
data.isnull().sum()
```

Out[4]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               7
VisITedResources          8
AnnouncementsView         8
Discussion                9
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [5]:
```
mv = data['raisedhands'].mean
mv
```

Out[5]:
```
<bound method NDFrame._add_numeric_operations.<locals>.mean of 0
15.0
1      20.0
2      10.0
3      30.0
4      40.0
       ...
475     5.0
476    50.0
477    55.0
478     NaN
479    35.0
Name: raisedhands, Length: 480, dtype: float64>
```

In [6]:
```
data['raisedhands']=data['raisedhands'].fillna(mv);
```

In [7]:
```
data['raisedhands']
```

Out[7]:
```
0                                                    15.0
1                                                    20.0
2                                                    10.0
3                                                    30.0
4                                                    40.0
                                ...
475                                                   5.0
476                                                  50.0
477                                                  55.0
478    <bound method NDFrame._add_numeric_operations....
479                                                  35.0
Name: raisedhands, Length: 480, dtype: object
```

In [8]:
```
data.isnull().sum()
```

Out[8]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               0
VisITedResources          8
AnnouncementsView         8
Discussion                9
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [9]:
```
mv = data['VisITedResources'].mode;
```

In [10]:
```
data['VisITedResources'] = data['VisITedResources'].fillna(mv)
```

In [11]:
```
data.isnull().sum()
```

Out[11]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               0
VisITedResources          0
AnnouncementsView         8
Discussion                9
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [12]:
```
mv = data['AnnouncementsView'].mean;
print(mv)
```

```
<bound method NDFrame._add_numeric_operations.<locals>.mean of 0
2.0
1       3.0
2       0.0
3       5.0
4      12.0
       ...
475     5.0
476    14.0
477    25.0
478    14.0
479    23.0
Name: AnnouncementsView, Length: 480, dtype: float64>
```

In [13]:
```
data['AnnouncementsView'] = data['AnnouncementsView'].fillna(mv)
```

In [14]:
```
data.isnull().sum()
```

Out[14]:
```
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                 9
ParentAnsweringSurvey      0
ParentschoolSatisfaction   0
StudentAbsenceDays         0
Class                      0
dtype: int64
```

In [15]:
```
mv = data['Discussion'].median;
mv
```

Out[15]:
```
<bound method NDFrame._add_numeric_operations.<locals>.median of 0
20.0
1      25.0
2      30.0
3      35.0
4      50.0
       ...
475     8.0
476    28.0
477    29.0
478    57.0
479    62.0
Name: Discussion, Length: 480, dtype: float64>
```

In [16]:
```
data['Discussion'] = data['Discussion'].fillna(mv)
```

In [17]:
```
data.isnull().sum()
```

Out[17]:
```
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
```

In [18]:
```
data1 = pd.read_csv('xAPI-Edu-Data.csv')
```

In [19]:
```
data1.isnull().sum()
```

Out[19]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               7
VisITedResources          8
AnnouncementsView         8
Discussion                9
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [20]:
```
data1.shape
```

Out[20]:
```
(480, 17)
```

In [21]:
```
d = data1.dropna()
```

In [22]:
```
d.shape
```

Out[22]:
```
(448, 17)
```

In [23]:
```
d.isnull().sum()
```

Out[23]:
```
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
```

In [24]:
```
data1.isnull().sum()
```

Out[24]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               7
VisITedResources          8
AnnouncementsView         8
Discussion                9
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [25]:
```
d1 = data1.fillna(method ='bfill')
```

In [26]:
```
d1.isnull().sum()
```

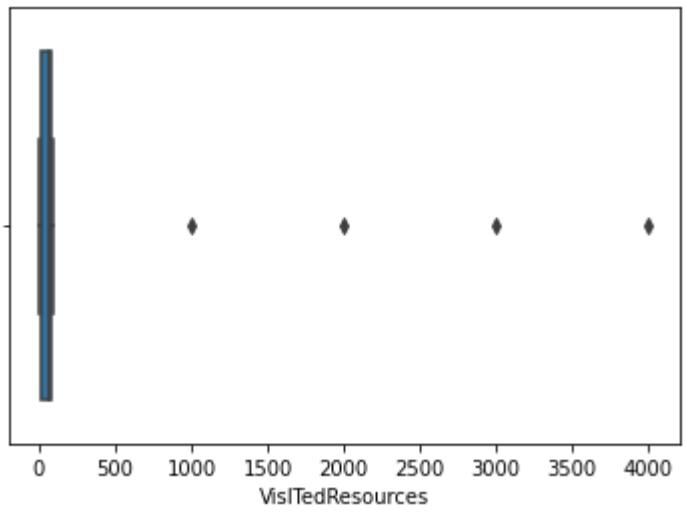
Out[26]:
```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               0
VisITedResources          0
AnnouncementsView         0
Discussion                0
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

In [27]:
```
d1.dtypes
```

Out[27]:
```
gender                     object
NationalITy                object
PlaceofBirth               object
StageID                    object
GradeID                    object
SectionID                  object
Topic                      object
Semester                   object
Relation                   object
raisedhands               float64
VisITedResources          float64
AnnouncementsView         float64
Discussion                float64
ParentAnsweringSurvey      object
ParentschoolSatisfaction   object
StudentAbsenceDays         object
Class                      object
dtype: object
```

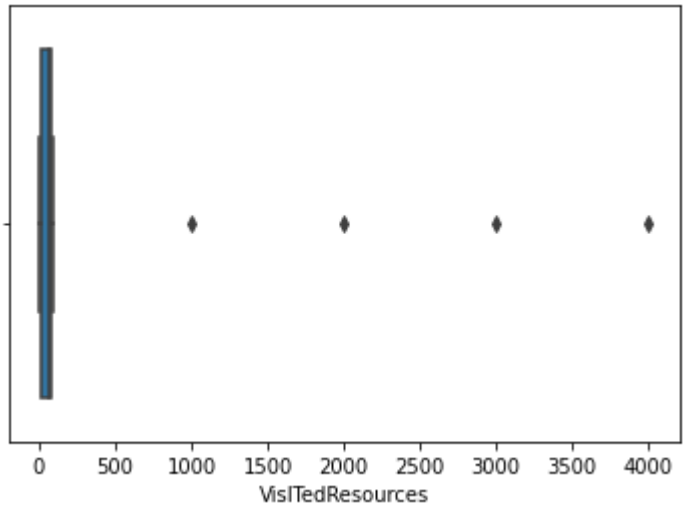
In [28]:
```
sns.boxplot(x=d1['VisITedResources'])
```

Out[28]:
```
<AxesSubplot:xlabel='VisITedResources'>
```



In [29]:
```
d2 = d1
sns.boxplot(x=d2['VisITedResources'])
d2['VisITedResources']
```

Out[29]:
```
0      1000.0
1        20.0
2         7.0
3        25.0
4        50.0
        ...
475       4.0
476      77.0
477      74.0
478      17.0
479      14.0
Name: VisITedResources, Length: 480, dtype: float64
```



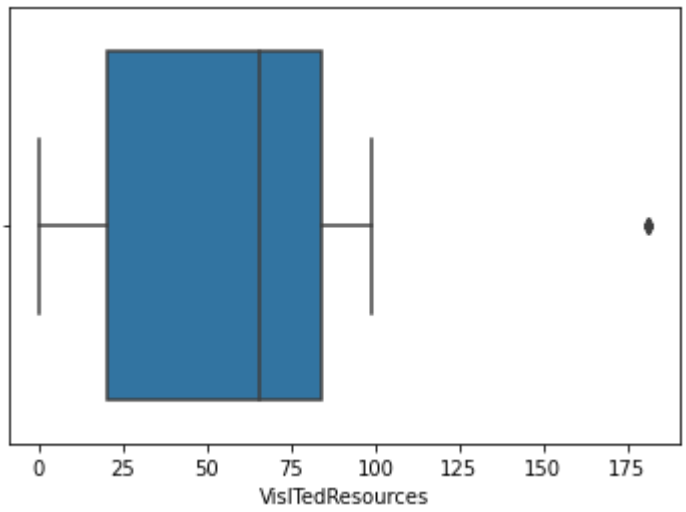
In [30]:
```
q1 = ny.percentile(d2['VisITedResources'],25,method='midpoint')
q3 = ny.percentile(d2['VisITedResources'],75,method='midpoint')
print(q1)
print(q3)
irp = q3-q1

upper = q3 + (1.5*irp)
lower = q1 - (1.5*irp)

outlierhandler=[]
for i in d2['VisITedResources']:

    if i < lower:
        outlierhandler.append(lower);
    elif i>upper:
        outlierhandler.append(upper);
    else :
        outlierhandler.append(i);

d2['VisITedResources'] = outlierhandler;
```

```
20.0
84.5
```

In [ ]:

In [31]:
```
sns.boxplot(x=d2['VisITedResources'])
```

Out[31]: <AxesSubplot:xlabel='VisITedResources'>



In [32]:
```
mn = ny.mean(d1['VisITedResources'])
stdd = ny.std(d1['VisITedResources'])
```

In [33]:
```
mn
```

Out[33]:
```
55.77916666666667
```

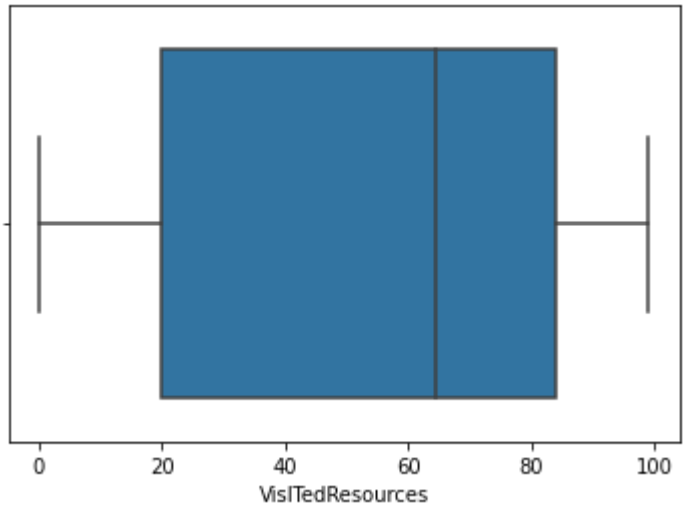
In [34]:
```
threshold = 3
outlierhandler = []
for i in d1['VisITedResources']:
    z = (i-mn)/stdd
    if z > threshold:
        outlierhandler.append(mn);
    else:
        outlierhandler.append(i);

d1['VisITedResources'] = outlierhandler;
```

In [35]:
```
d1['VisITedResources'].head()
```

Out[35]:
```
0    55.779167
1    20.000000
2     7.000000
3    25.000000
4    50.000000
Name: VisITedResources, dtype: float64
```

In [36]:
```
sns.boxplot(x=d1['VisITedResources'])
```

Out[36]:
```
<AxesSubplot:xlabel='VisITedResources'>
```



In [ ]:
```

```