In [1]:
```
import pandas as pd
```

In [2]:
```
import numpy as nm
```

In [3]:
```
df=pd.read_csv("xAPI-Edu-Data.csv")
```

In [4]:
```
print(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         15.0                16
1           IT        F   Father         20.0                20
2           IT        F   Father          NaN                 7
3           IT        F   Father         30.0                25
4           IT        F   Father         40.0                50
..         ...      ...      ...          ...               ...
475  Chemistry        S   Father          5.0                 4
476    Geology        F   Father         50.0                77
477    Geology        S   Father          NaN                74
478    History        F   Father         30.0                17
479    History        S   Father         35.0                14

     AnnouncementsView  Discussion ParentAnsweringSurvey  \
0                    2        20.0                   Yes
1                    3        25.0                   Yes
2                    0        30.0                    No
3                    5         NaN                    No
4                   12        50.0                    No
..                 ...         ...                   ...
475                  5         NaN                    No
476                 14        28.0                    No
477                 25        29.0                    No
478                 14         NaN                    No
479                 23        62.0                    No

    ParentschoolSatisfaction StudentAbsenceDays Class
0                       Good            Under-7     M
1                        NaN            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                      NaN            Under-7     M
478                      Bad            Above-7     L
479                      NaN            Above-7     L

[480 rows x 17 columns]
```

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
gender                      0
NationalITy                 0
PlaceofBirth                0
StageID                     0
GradeID                     0
SectionID                   0
Topic                       0
Semester                    0
Relation                    0
raisedhands                 3
VisITedResources            0
AnnouncementsView           0
Discussion                  6
ParentAnsweringSurvey       0
ParentschoolSatisfaction    6
StudentAbsenceDays          0
Class                       0
dtype: int64
```

In [6]:
```
df['ParentschoolSatisfaction']=df['ParentschoolSatisfaction'].replace
```

In [7]:
```
df['raisedhands']=df['raisedhands'].replace(nm.NaN,df['raisedhands'].
```

In [8]:
```
df['Discussion']=df['Discussion'].replace(nm.NaN,df['Discussion'].mea
```

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
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 [10]:
```
from scipy import stats
```

In [11]:
```
z=stats.zscore(df['Discussion'])
```

In [12]:
```
threshold=1
```

In [13]:
```
outliers=df[z>threshold]
```

In [15]:
```
print(outliers.index)
```

```
Index([ 10,  16,  17,  18,  19,  20,  21,  22,  37,  43,  44,  47,
48,  49,
        53,  62,  67,  82,  96, 100, 105, 111, 138, 151, 155, 159,
162, 180,
       200, 209, 218, 223, 228, 239, 240, 241, 244, 246, 247, 252,
258, 282,
       283, 286, 287, 289, 292, 293, 294, 295, 296, 297, 305, 306,
307, 308,
       309, 314, 315, 328, 329, 372, 373, 378, 379, 380, 381, 386,
387, 395,
       398, 399, 402, 403, 405, 413, 416, 417, 418, 419, 424, 432,
433, 442,
       446, 447, 448, 449, 454, 455, 458, 459, 460, 461, 462, 463,
464, 465,
       468, 469],
      dtype='int64')
```

In [16]:
```
Q1=df['Discussion'].quantile(0.25)
```

In [19]:
```
Q3=df['Discussion'].quantile(0.75)
```

In [20]:
```
IQR=Q3-Q1
```

In [49]:
```
threshold=0.5
```

In [50]:
```
outliers1=df['Discussion']<(Q1-threshold*IQR)
```

In [51]:
```
print(outliers1)
```

```
0      False
1      False
2      False
3      False
4      False
       ...
475    False
476    False
477    False
478    False
479    False
Name: Discussion, Length: 480, dtype: bool
```

In [52]:
```
outliers2=df['Discussion']>(Q3+threshold*IQR)
```

In [53]:
```
print(outliers2)
```

```
0      False
1      False
2      False
3      False
4      False
       ...
475    False
476    False
477    False
478    False
479    False
Name: Discussion, Length: 480, dtype: bool
```

In [54]:
```
print(Q1)
```

```
20.0
```

In [ ]:
```

```