

Displaying a Dataset in Streamlit

	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
5	F	KW	Kuwait	lowerlevel	G-04	A	IT	F	Father
6	M	KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father
7	M	KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father
8	F	KW	Kuwait	MiddleSchool	G-07	A	Math	F	Father
9	F	KW	Kuwait	MiddleSchool	G-07	B	IT	F	Father

dataset Shape:

(480, 17)

Show Dataset information:

```
<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   Father           480 non-null    object 
 8   Mother           480 non-null    object 
 9   Son              480 non-null    object 
 10  Daughter         480 non-null    object 
 11  Grandson        480 non-null    object 
 12  Granddaughter   480 non-null    object 
 13  Brother          480 non-null    object 
 14  Sister           480 non-null    object 
 15  Husband          480 non-null    object 
 16  Wife             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

```

Describe Dataset:

	raisedhands	VisITEDResources	AnnouncementsView	Discussion
count	480	480	480	480
mean	46.775	54.7979	37.9188	43.2833
std	30.7792	33.08	26.6112	27.6377
min	0	0	0	1
25%	15.75	20	14	20
50%	50	65	33	39
75%	75	84	58	70
max	100	99	98	99

Show the column name:

1. gender
2. Nationality
3. PlaceofBirth
4. StageID
5. GradeID

6. SectionID
7. Topic
8. Semester
9. Relation
10. raisedhands
11. VisITedResources
12. AnnouncementsView
13. Discussion
14. ParentAnsweringSurvey
15. ParentschoolSatisfaction
16. StudentAbsenceDays
17. Class

the missing values for each column:

	0
gender	0
NationalITY	0
PlaceofBirth	0
StageID	0
GradeID	0
SectionID	0
Topic	0
Semester	0
Relation	0
raisedhands	0

nun values heatmap:

gender	-
Nationality	-
PlaceOfBirth	-
StageID	-
GradeID	-
SectionID	-
Topic	-
Semester	-
Relation	-
raisedhands	-
VisitedResources	-
AnnouncementsView	-
Discussion	-
ParentAnsweringSurvey	-
ParentschoolSatisfaction	-
StudentAbsenceDays	-
Class	-

check the duplicated data:

are duplicates: `True`

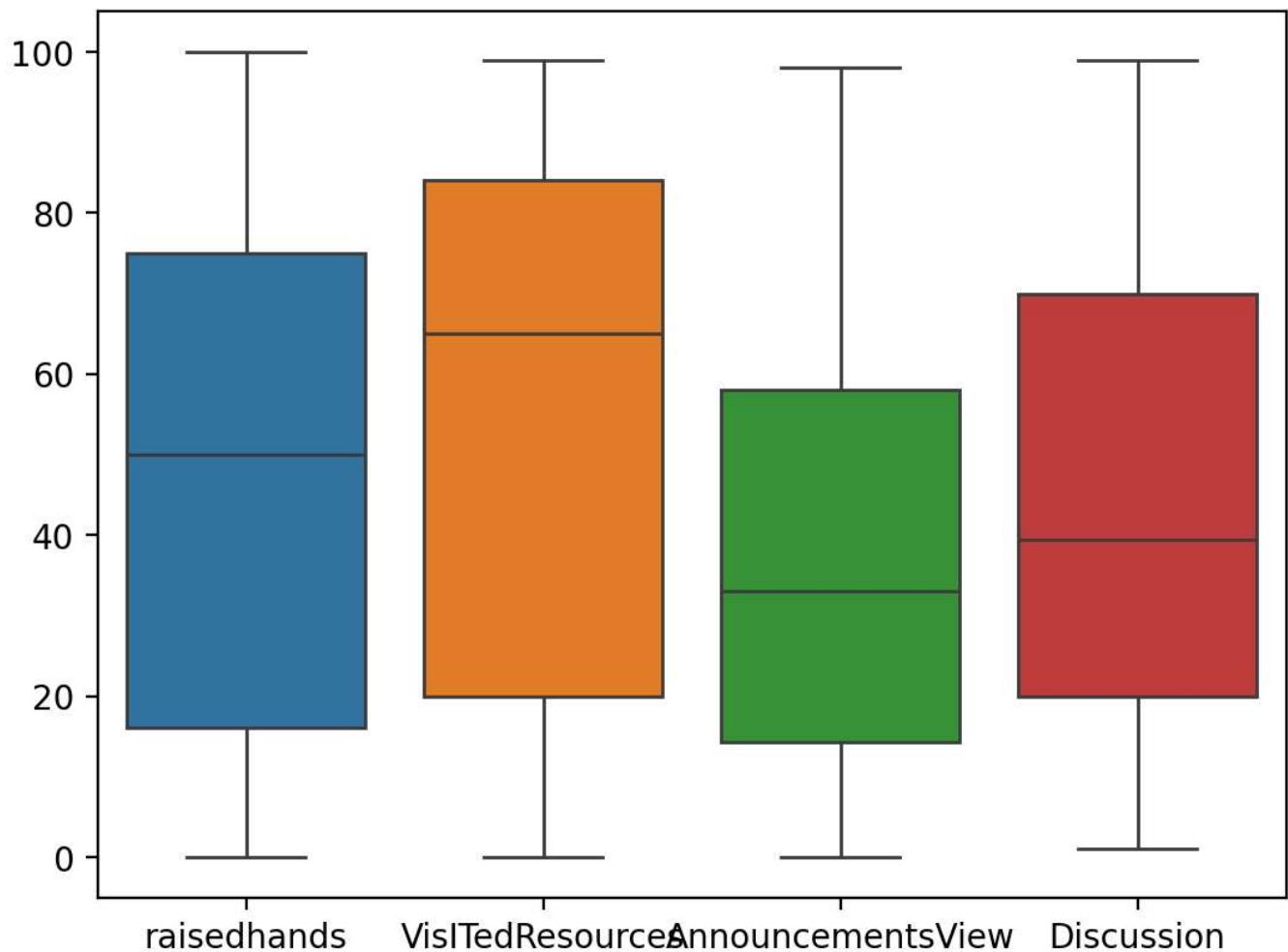
number of duplicates: `2`

check the duplicated data:

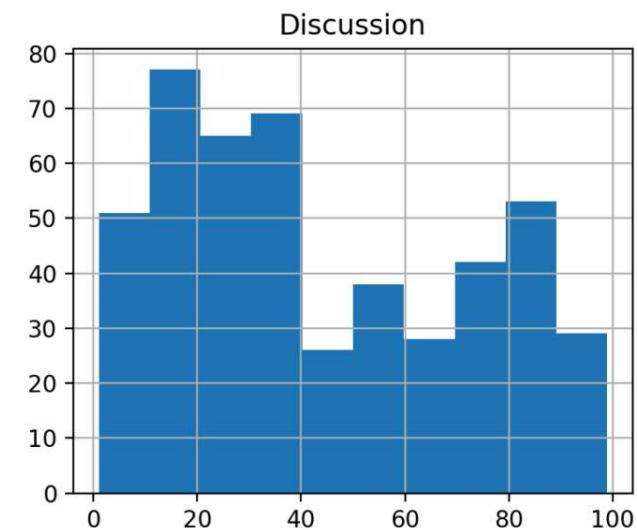
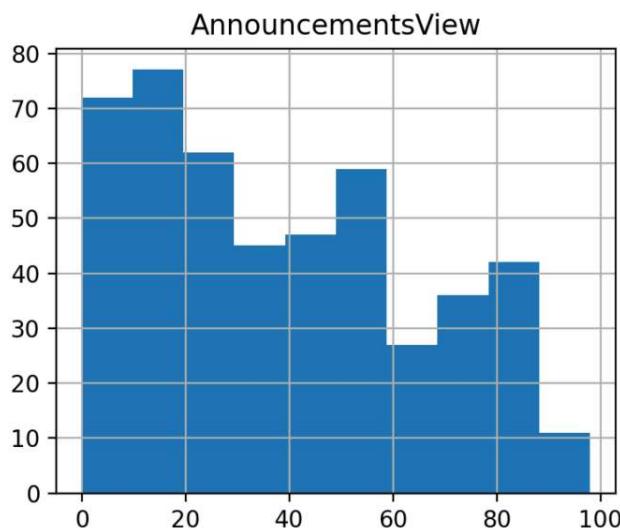
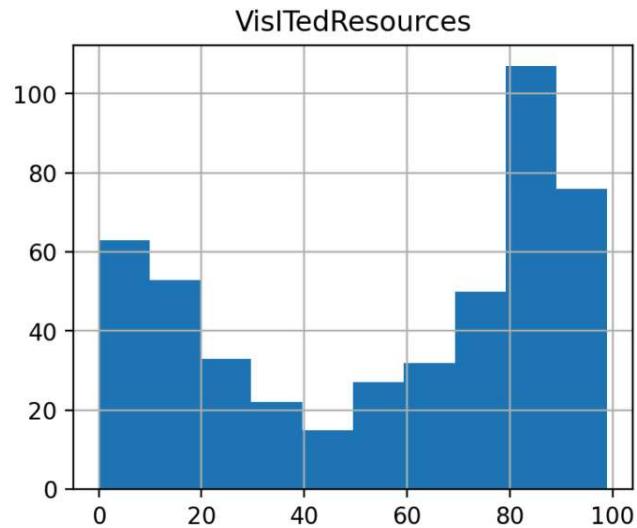
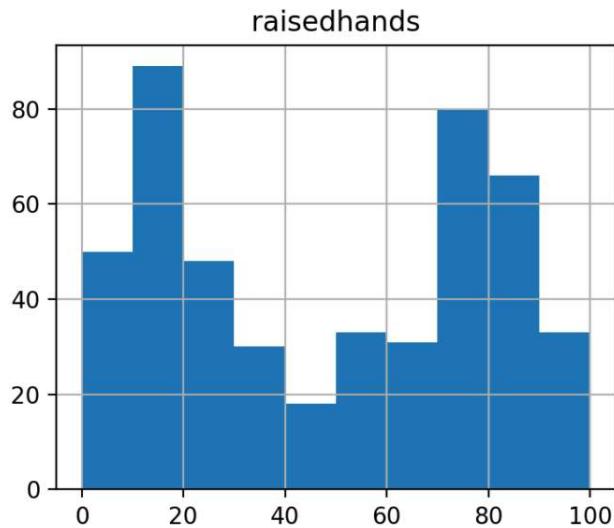
are duplicates: `False`

number of duplicates: `0`

visualize main columns:

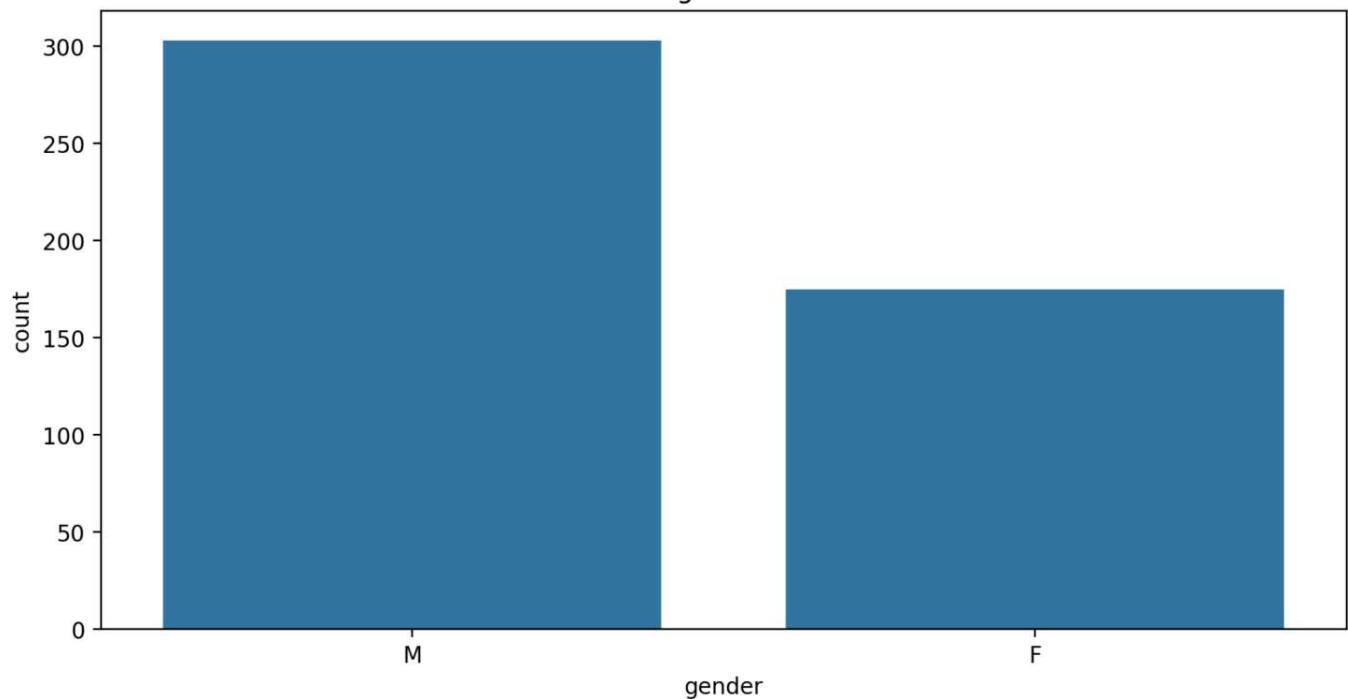


create histograms:

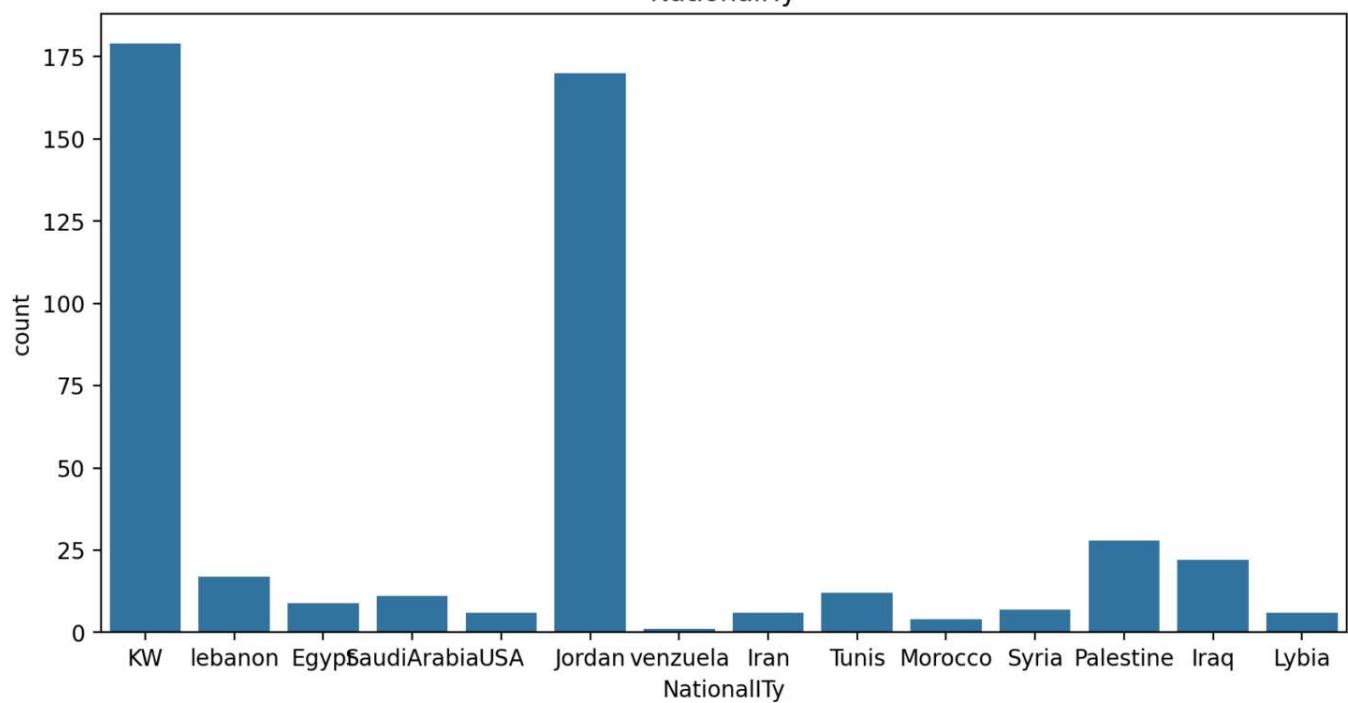


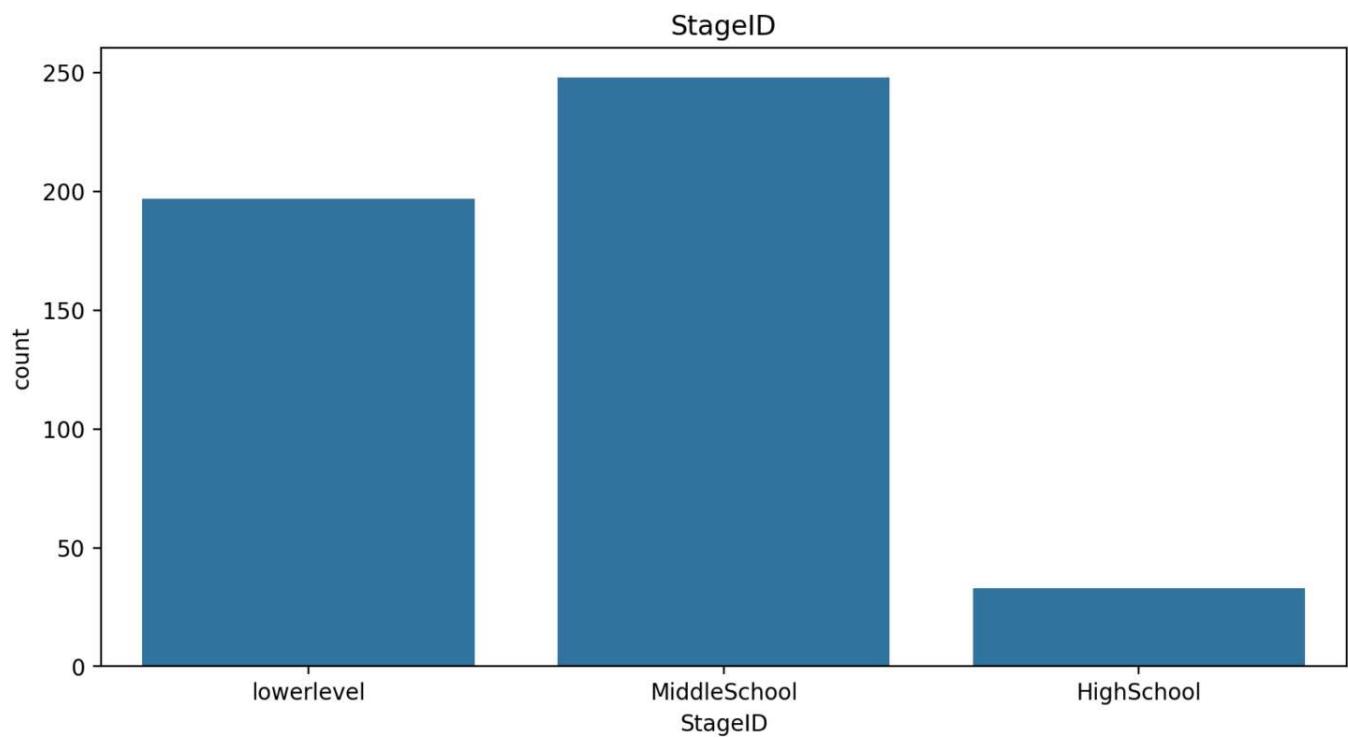
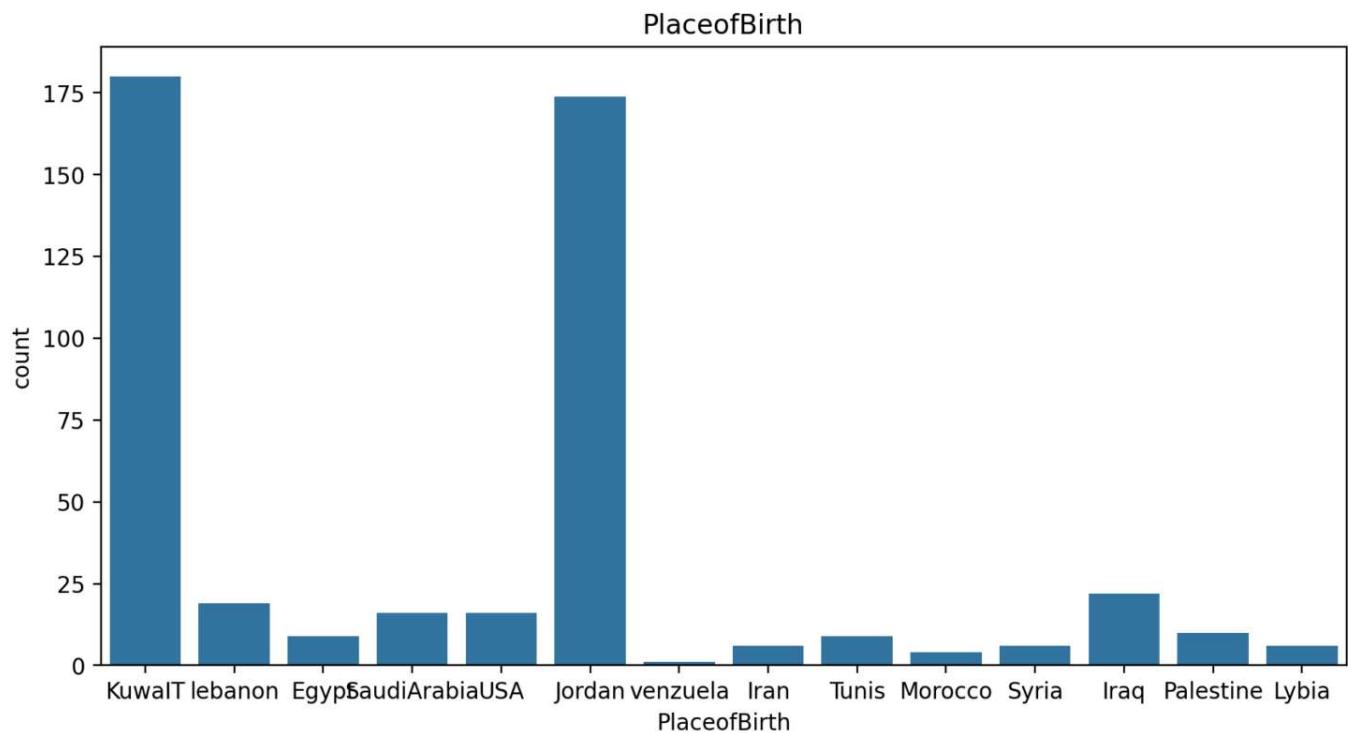
Bar plots for categorical data

gender

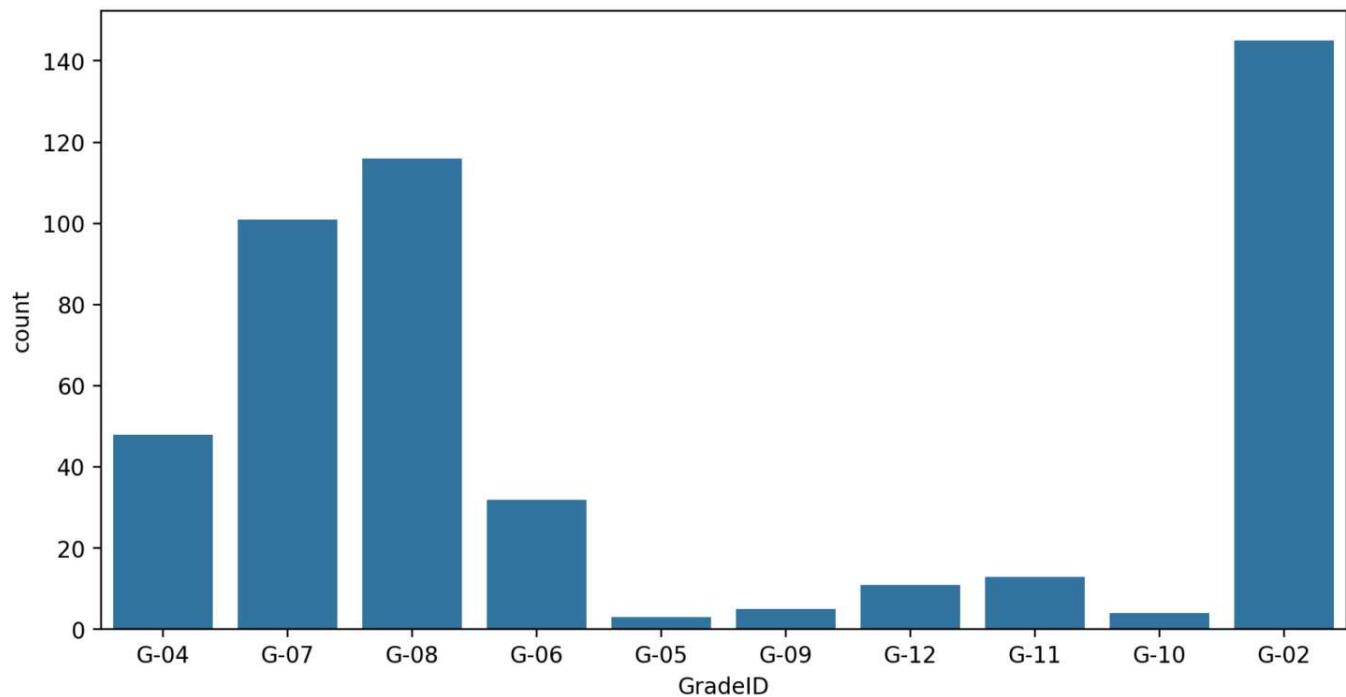


NationalITY

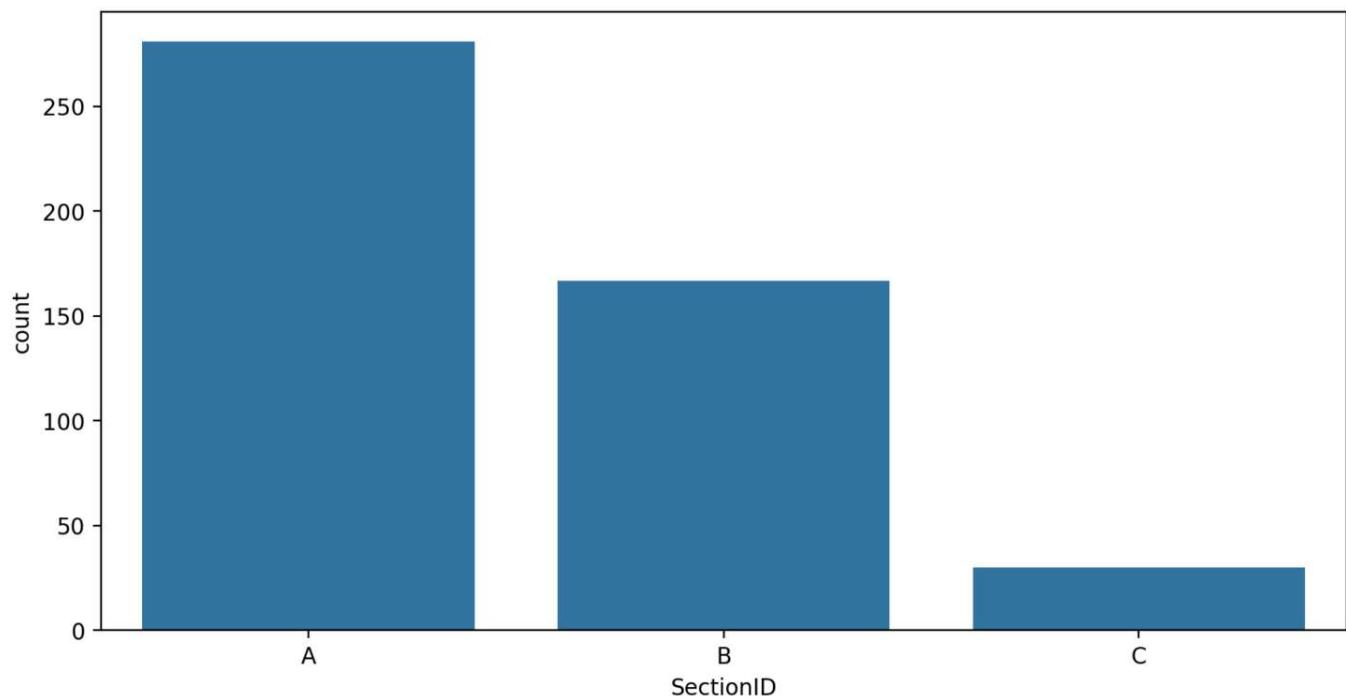




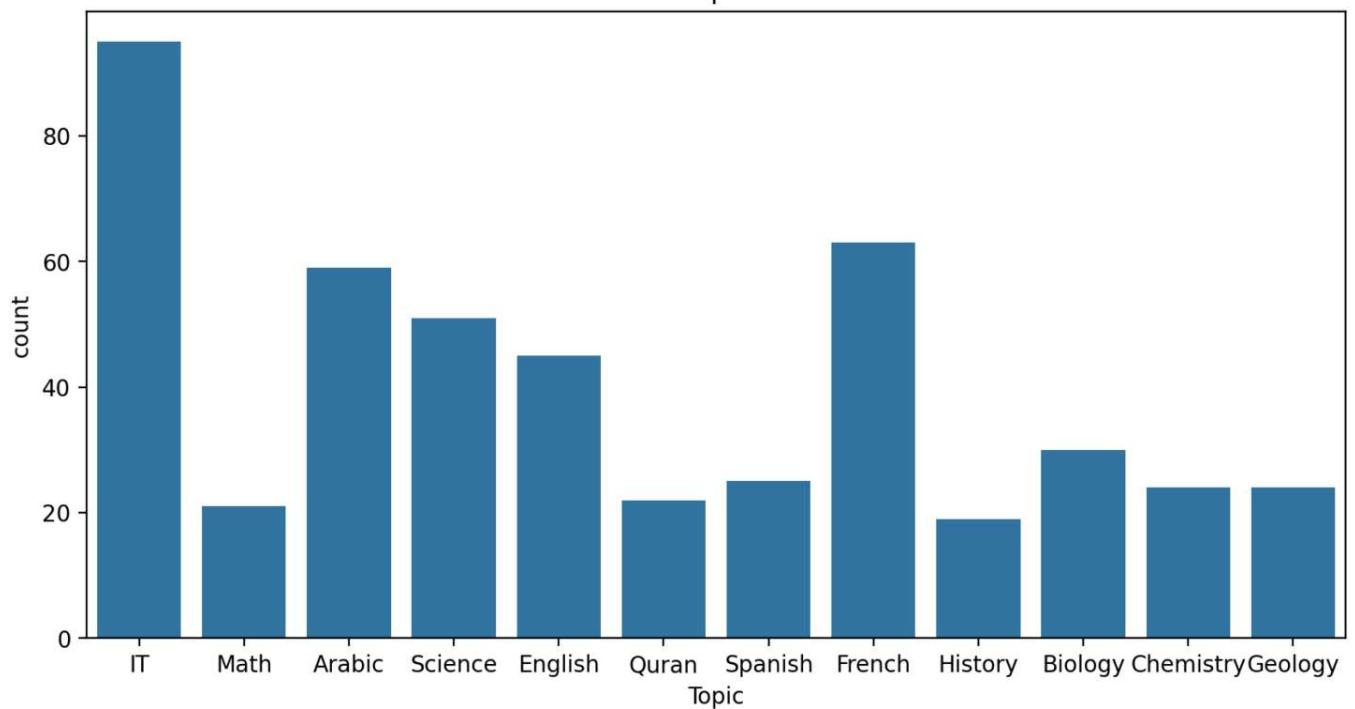
GradeID



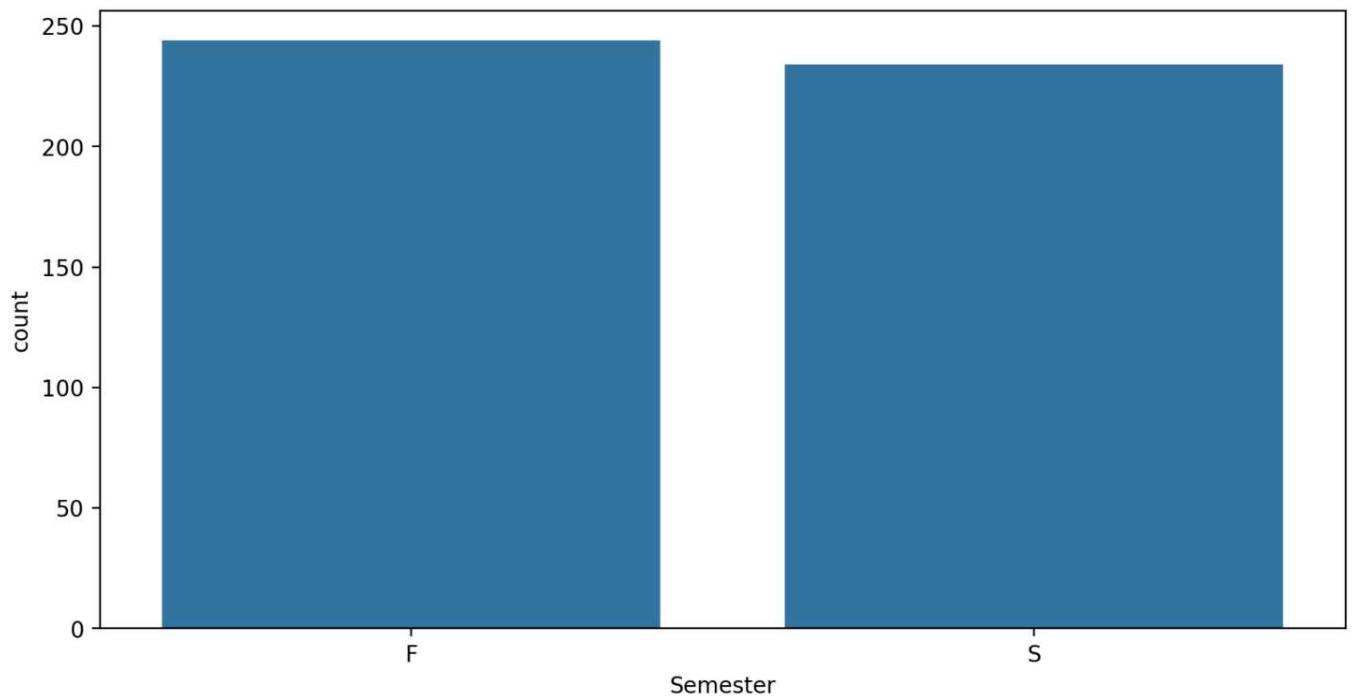
SectionID

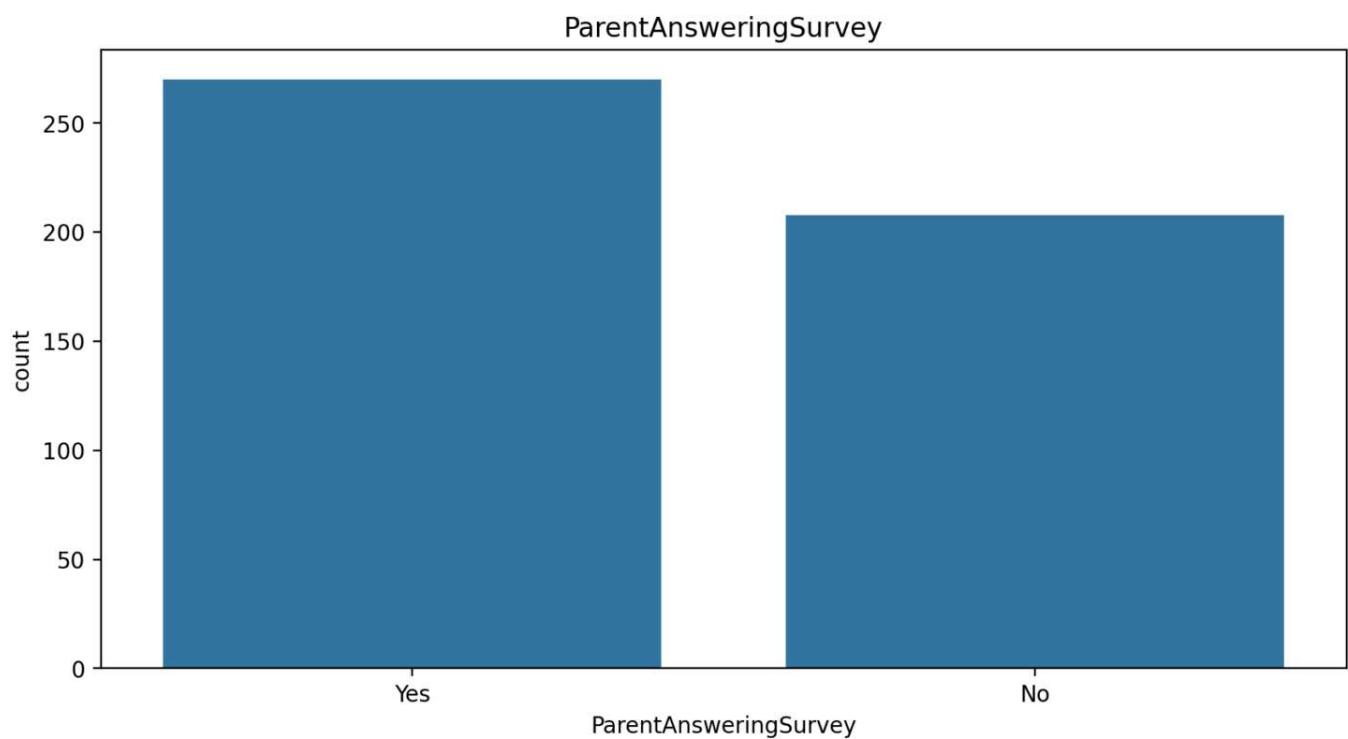
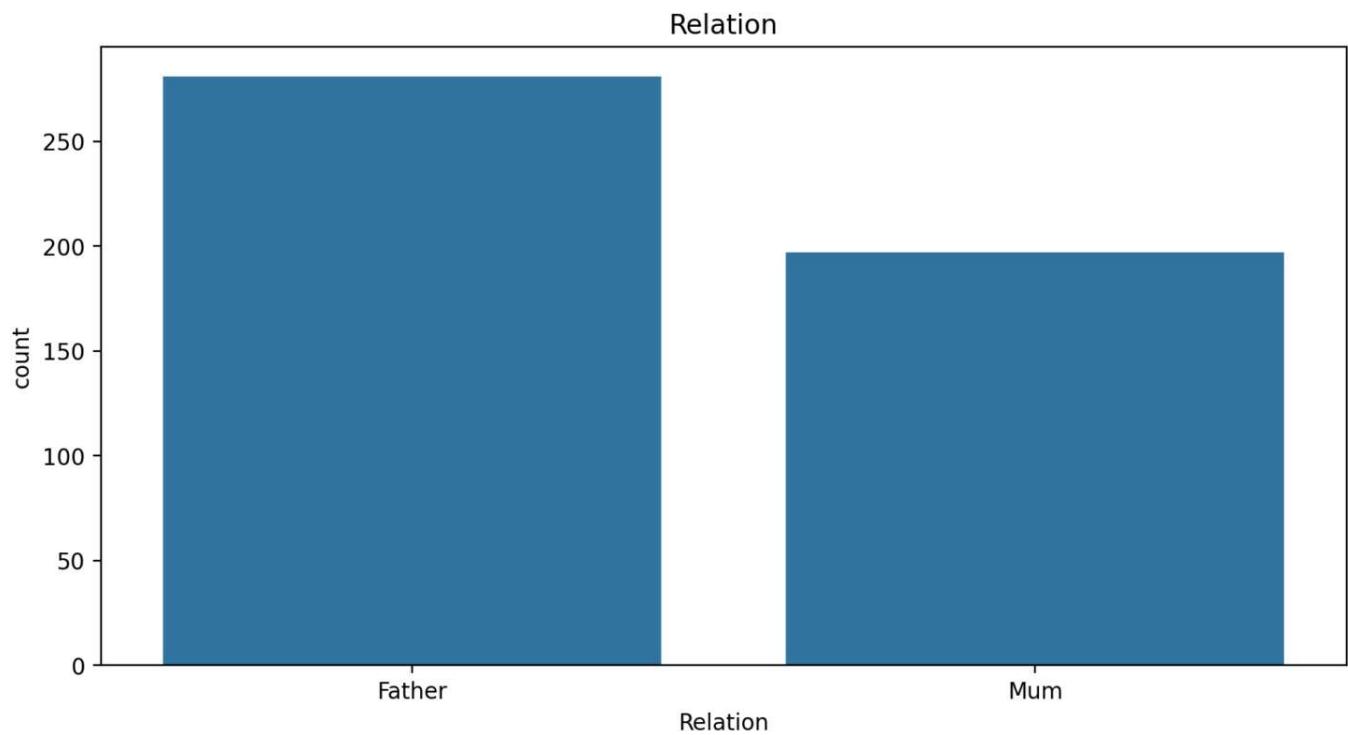


Topic

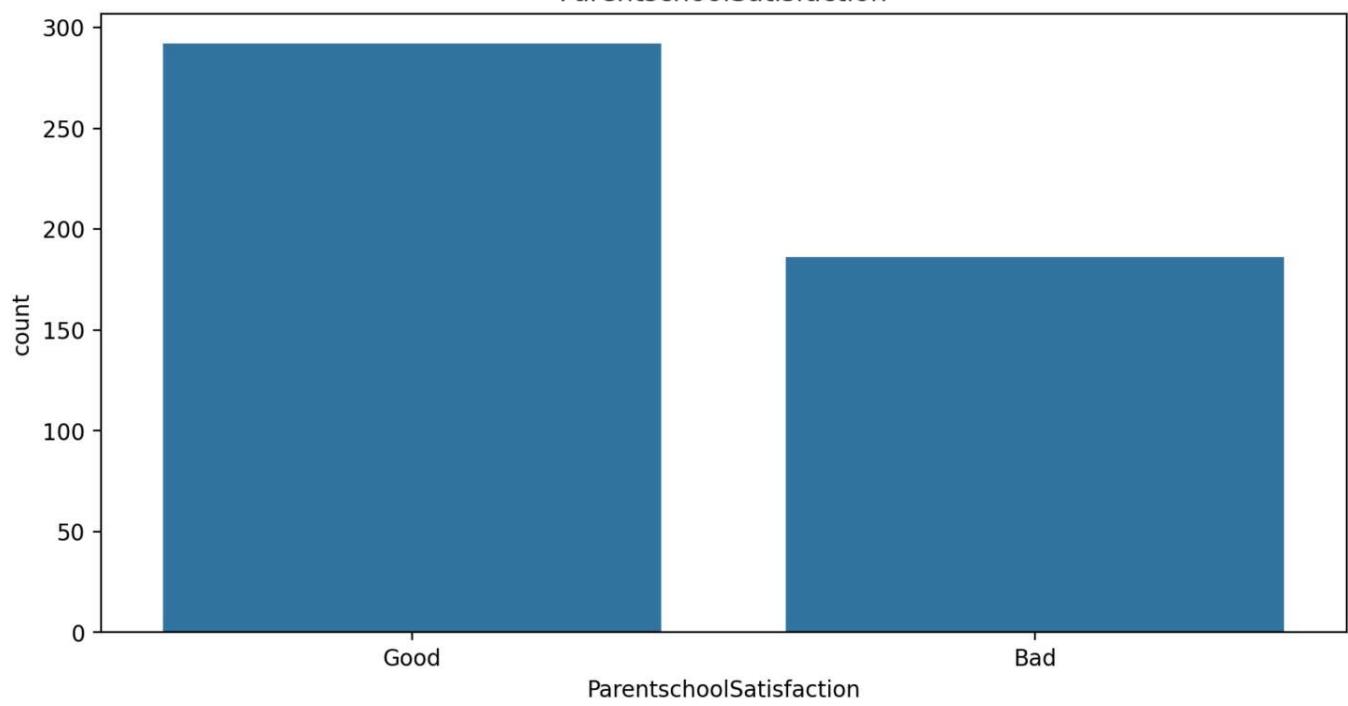


Semester

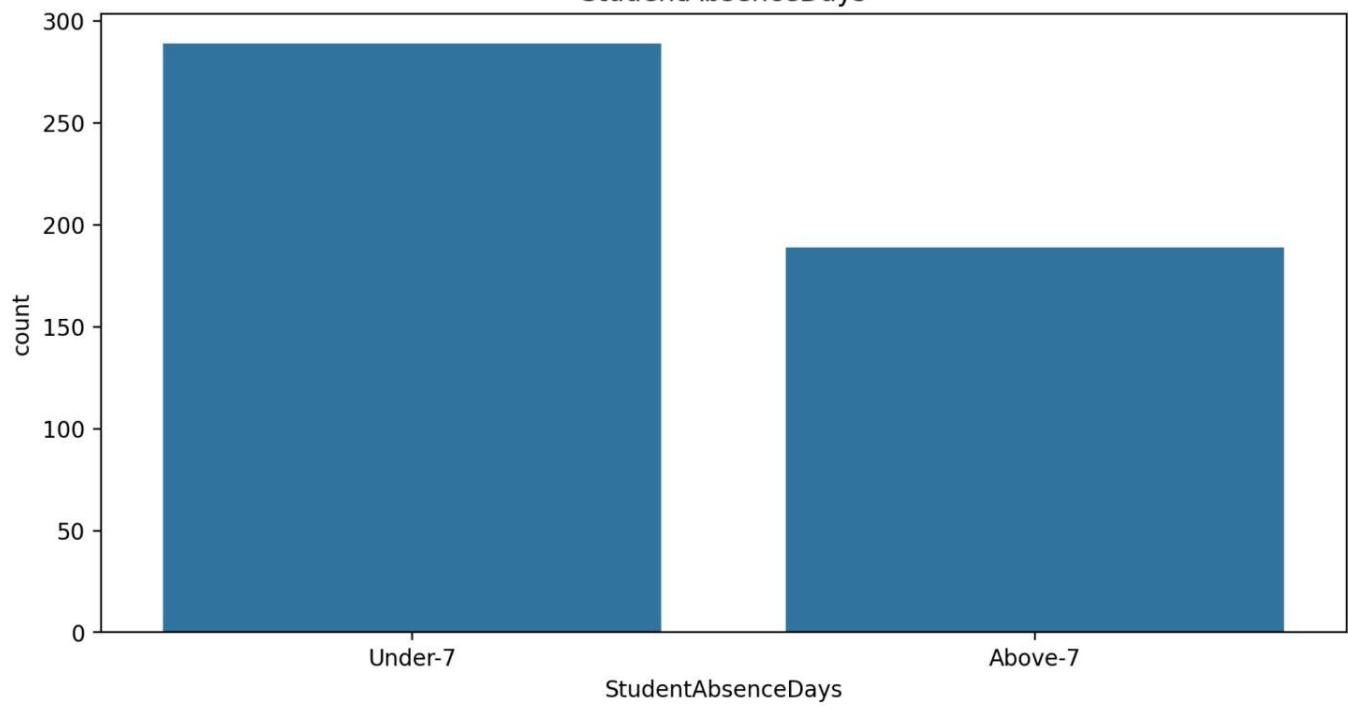




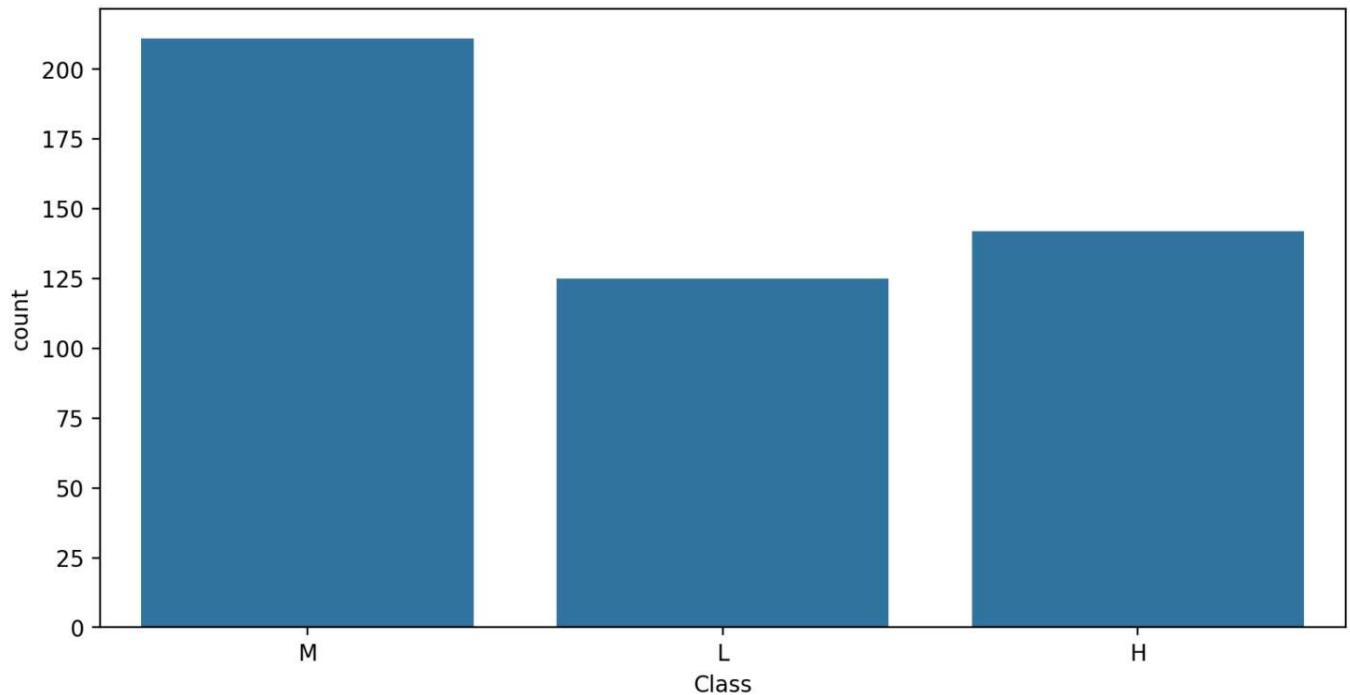
ParentschoolSatisfaction



StudentAbsenceDays



Class



Iterate over each object column and print the unique values:

Unique values for column 'gender':

M

F

Unique values for column 'NationalITY':

KW

lebanon

Egypt

SaudiArabia

USA

Jordan

venzuela

Iran

Tunis

Morocco

Syria

Palestine

Iraq

Lybia

Unique values for column 'PlaceofBirth':

Kuwait

lebanon

Egypt

SaudiArabia

USA

Jordan

venzuela

Iran

Tunis

Morocco

Syria

Iraq

Palestine

Lybia

Unique values for column 'StageID':

lowerlevel

MiddleSchool

HighSchool

Unique values for column 'GradeID':

G-04

G-07

G-08

G-06

G-05

G-09

G-12

G-11

G-10

G-02

Unique values for column 'SectionID':

A

B

C

Unique values for column 'Topic':

IT

Math

Arabic

Science

English

Quran

Spanish

French

History

Biology

Chemistry

Geology

Unique values for column 'Semester':

F

S

Unique values for column 'Relation':

Father

Mum

Unique values for column 'ParentAnsweringSurvey':

Yes

No

Unique values for column 'ParentschoolSatisfaction':

Good

Bad

Unique values for column 'StudentAbsenceDays':

Under-7

Above-7

Unique values for column 'Class':

M

L

H

Display the dictionary of unique values

```
▼ {  
  ▼ "gender" : [  
    0 : "M"  
    1 : "F"  
  ]  
  ▼ "NationalITY" : [  
    0 : "KW"  
    1 : "lebanon"  
    2 : "Egypt"  
    3 : "SaudiArabia"  
    4 : "USA"  
    5 : "Jordan"  
    6 : "venzuela"  
    7 : "Iran"  
    8 : "Tunis"  
    9 : "Morocco"  
    10 : "Syria"  
    11 : "Palestine"  
    12 : "Iraq"  
    13 : "Lybia"  
  ]  
  ▼ "PlaceofBirth" : [  
    0 : "KuwaIT"  
    1 : "lebanon"  
    2 : "Egypt"  
    3 : "SaudiArabia"  
    4 : "USA"  
    5 : "Jordan"  
    6 : "venzuela"  
    7 : "Iran"  
    8 : "Tunis"  
    9 : "Morocco"  
    10 : "Syria"
```

```
    11 : "Iraq"
    12 : "Palestine"
    13 : "Lybia"
]

▼ "StageID" : [
    0 : "lowerlevel"
    1 : "MiddleSchool"
    2 : "HighSchool"
]

▼ "GradeID" : [
    0 : "G-04"
    1 : "G-07"
    2 : "G-08"
    3 : "G-06"
    4 : "G-05"
    5 : "G-09"
    6 : "G-12"
    7 : "G-11"
    8 : "G-10"
    9 : "G-02"
]

▼ "SectionID" : [
    0 : "A"
    1 : "B"
    2 : "C"
]

▼ "Topic" : [
    0 : "IT"
    1 : "Math"
    2 : "Arabic"
    3 : "Science"
    4 : "English"
    5 : "Quran"
    6 : "Spanish"
    7 : "French"
    8 : "History"
]
```

```
    9 : "Biology"
    10 : "Chemistry"
    11 : "Geology"
]
▼ "Semester" : [
    0 : "F"
    1 : "S"
]
▼ "Relation" : [
    0 : "Father"
    1 : "Mum"
]
▼ "ParentAnsweringSurvey" : [
    0 : "Yes"
    1 : "No"
]
▼ "ParentschoolSatisfaction" : [
    0 : "Good"
    1 : "Bad"
]
▼ "StudentAbsenceDays" : [
    0 : "Under-7"
    1 : "Above-7"
]
▼ "Class" : [
    0 : "M"
    1 : "L"
    2 : "H"
]
}
```

Display the column names of the DataFrame:

	0
0	gender
1	NationalITy
2	PlaceofBirth
3	StageID
4	GradeID
5	SectionID
6	Topic
7	Semester
8	Relation
9	raisedhands

Calculate and print the normalized value counts of 'gender' grouped by 'Class'

Class	gender	proportion
H	F	0.5282
H	M	0.4718
L	M	0.808
L	F	0.192
M	M	0.6398
M	F	0.3602

Calculate and print the normalized value counts of 'NationalITy' grouped by 'Class'

Class	Nationality	proportion
H	Jordan	0.3732
H	KW	0.2535
H	Iraq	0.0986
H	Palestine	0.0845
H	lebanon	0.0634
H	SaudiArabia	0.0423
H	USA	0.0211
H	Tunis	0.0211
H	Egypt	0.0141
H	Syria	0.0141

Calculate and print the normalized value counts of 'PlaceofBirth' grouped by 'Class'

Class	PlaceofBirth	proportion
H	Jordan	0.3732
H	Kuwait	0.2465
H	Iraq	0.0986
H	lebanon	0.0775
H	SaudiArabia	0.0493
H	USA	0.0423
H	Palestine	0.0423
H	Egypt	0.0211
H	Tunis	0.0211
H	Syria	0.0141

Calculate and print the normalized value counts of 'StageID' grouped by 'Class'

Class	StageID	proportion
H	MiddleSchool	0.5352
H	lowerlevel	0.3873
H	HighSchool	0.0775
L	lowerlevel	0.504
L	MiddleSchool	0.432
L	HighSchool	0.064
M	MiddleSchool	0.5592
M	lowerlevel	0.3744
M	HighSchool	0.0664

Display the rows of the encoded DataFrame to verify the transformations:

	gender	StageID	GradeID	Semester	Relation	raisedhands	VisitedResources	AnnouncementsView
0	1	0	1	0	1	15	16	2
1	1	0	1	0	1	20	20	2
2	1	0	1	0	1	10	7	0
3	1	0	1	0	1	30	25	2
4	1	0	1	0	1	40	50	12
5	0	0	1	0	1	42	30	13
6	1	1	4	0	1	35	12	0
7	1	1	4	0	1	50	10	11
8	0	1	4	0	1	12	21	16
9	0	1	4	0	1	70	80	21

Shape of encoded data:

(478, 56)

columns of encoded data:

1. gender
2. StageID
3. GradeID
4. Semester
5. Relation
6. raisedhands
7. VisitedResources
8. AnnouncementsView
9. Discussion

10. ParentAnsweringSurvey

11. ParentschoolSatisfaction

12. StudentAbsenceDays

13. Class

14. N_Egypt

15. N_Iran

16. N_Iraq

17. N_Jordan

18. N_KW

19. N_Lybia

20. N_Morocco

21. N_Palestine

22. N_SaudiArabia

23. N_Syria

24. N_Tunis

25. N_USA

26. N_lebanon

27. N_venzuela

28. B_Egypt

29. B_Iran

30. B_Iraq

31. B_Jordan

32. B_Kuwait

33. B_Lybia

34. B_Morocco

35. B_Palestine

36. B_SaudiArabia

37. B_Syria

38. B_Tunis

39. B_USA

40. B_lebanon

41. B_venzuela

42. S_A

43. S_B

44. S_C

45. T_Arabic

46. T_Biology

47. T_Chemistry

48. T_English

49. T_French

50. T_Geology

51. T_History

52. T_IT

53. T_Math

54. T_Quran

55. T_Science

56. T_Spanish

check the missing values of encoded data

Missing Values:

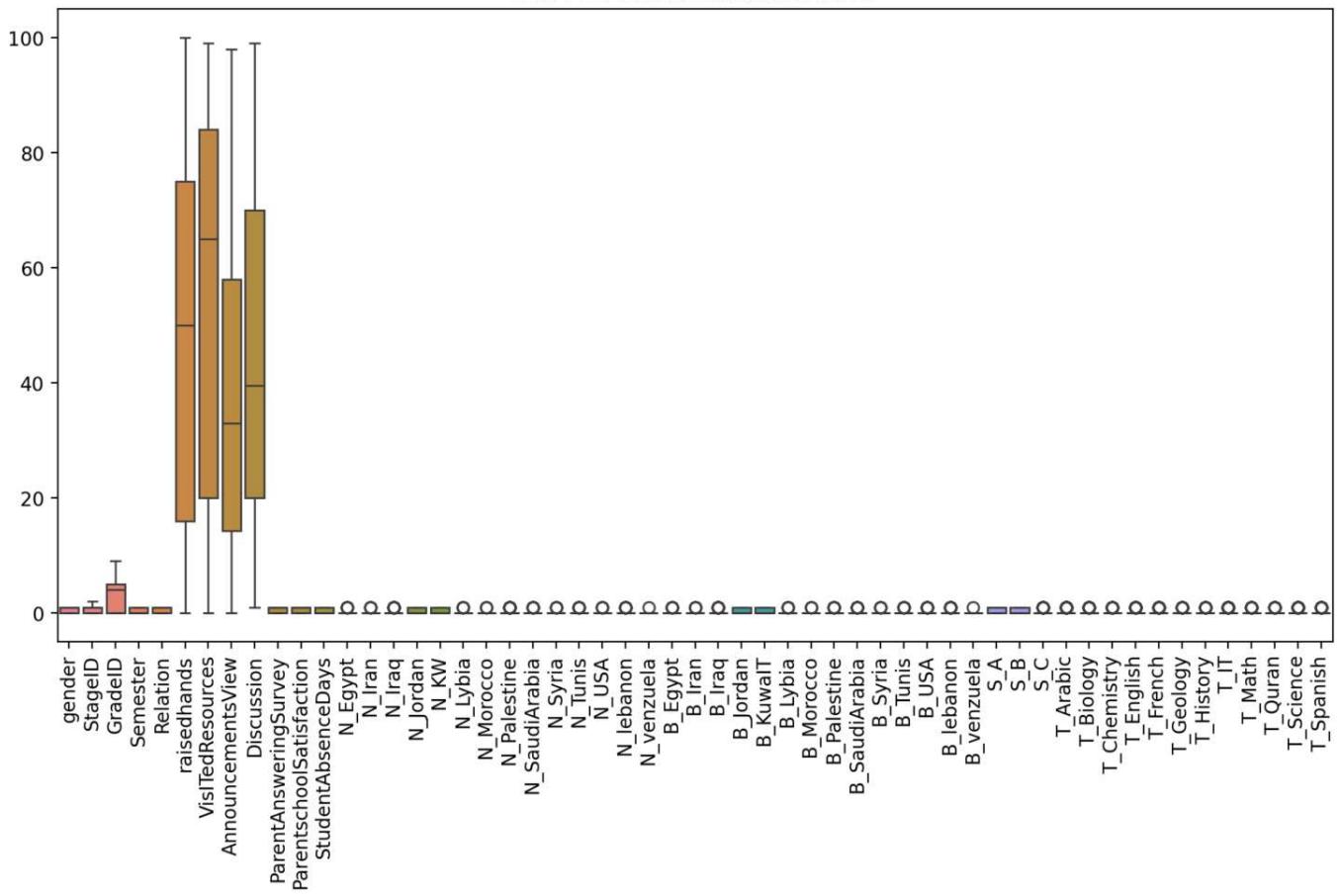
	0
gender	0
StageID	0
GradeID	0
Semester	0
Relation	0
raisedhands	0
VisitedResources	0
AnnouncementsView	0
Discussion	0
ParentAnsweringSurvey	0

display the encoded data:

	gender	StageID	GradeID	Semester	Relation	raisedhands	VisitedResources	AnnouncementsView
0	1	0	1	0	1	15	16	2
1	1	0	1	0	1	20	20	2
2	1	0	1	0	1	10	7	0
3	1	0	1	0	1	30	25	2
4	1	0	1	0	1	40	50	12
5	0	0	1	0	1	42	30	13
6	1	1	4	0	1	35	12	0
7	1	1	4	0	1	50	10	11
8	0	1	4	0	1	12	21	16
9	0	1	4	0	1	70	80	21

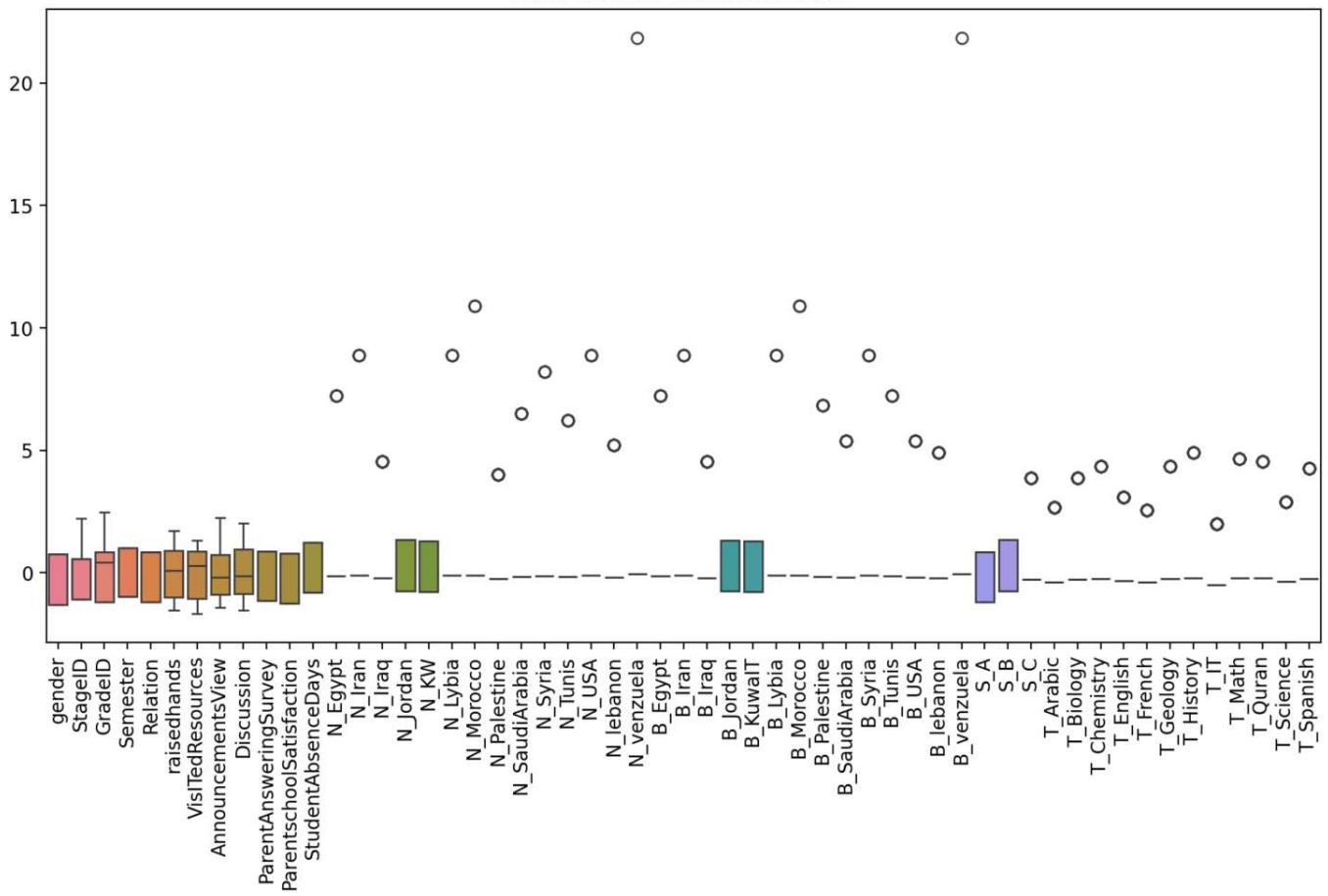
Plot a box plot to visualize the distribution of features before standardization:

Box Plot Before Standardization



Plot a box plot to visualize the distribution of features after standardization:

Box Plot After Standardization



Plot data after capping outliers:

After Capping Outliers

	gender	StageID	GradeID	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView
0	0.76	-1.0896	-0.78	-0.9793	0.8373	-1.0368	-1.181	-1.354
1	0.76	-1.0896	-0.78	-0.9793	0.8373	-0.8742	-1.0598	-1.316
2	0.76	-1.0896	-0.78	-0.9793	0.8373	-1.1994	-1.4538	-1.429
3	0.76	-1.0896	-0.78	-0.9793	0.8373	-0.5491	-0.9083	-1.241
4	0.76	-1.0896	-0.78	-0.9793	0.8373	-0.2239	-0.1508	-0.978
5	-1.3158	-1.0896	-0.78	-0.9793	0.8373	-0.1589	-0.7568	-0.940
6	0.76	0.5691	0.4397	-0.9793	0.8373	-0.3865	-1.3022	-1.429
7	0.76	0.5691	0.4397	-0.9793	0.8373	0.1012	-1.3629	-0.865
8	-1.3158	0.5691	0.4397	-0.9793	0.8373	-1.1343	-1.0295	-0.827
9	-1.3158	0.5691	0.4397	-0.9793	0.8373	0.7515	0.7583	-0.489

Display final VIF values:

	feature	VIF
0	gender	1.1411
1	StageID	1.2805
2	Semester	1.164
3	Relation	1.364
4	raisedhands	2.5875
5	VisITedResources	2.646
6	AnnouncementsView	2.4065
7	Discussion	1.3425
8	ParentAnsweringSurvey	1.6681
9	ParentschoolSatisfaction	1.641

Check for missing values in the VIF DataFrame:

Missing Values:

	0
feature	0
VIF	0

Display the rows of the DataFrame after dropping specified columns:

	gender	StageID	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discuss
0	0.76	-1.0896	-0.9793	0.8373	-1.0368	-1.181	-1.3542	-0.8
1	0.76	-1.0896	-0.9793	0.8373	-0.8742	-1.0598	-1.3166	-0.6
2	0.76	-1.0896	-0.9793	0.8373	-1.1994	-1.4538	-1.4294	-0.
3	0.76	-1.0896	-0.9793	0.8373	-0.5491	-0.9083	-1.2414	-0.3
4	0.76	-1.0896	-0.9793	0.8373	-0.2239	-0.1508	-0.9782	0.2
5	-1.3158	-1.0896	-0.9793	0.8373	-0.1589	-0.7568	-0.9406	0.9
6	0.76	0.5691	-0.9793	0.8373	-0.3865	-1.3022	-1.4294	-0.9
7	0.76	0.5691	-0.9793	0.8373	0.1012	-1.3629	-0.8654	-0.7
8	-1.3158	0.5691	-0.9793	0.8373	-1.1343	-1.0295	-0.8278	0.2
9	-1.3158	0.5691	-0.9793	0.8373	0.7515	0.7583	-0.4894	0.9

show the heatmap:

Correlation Matrix



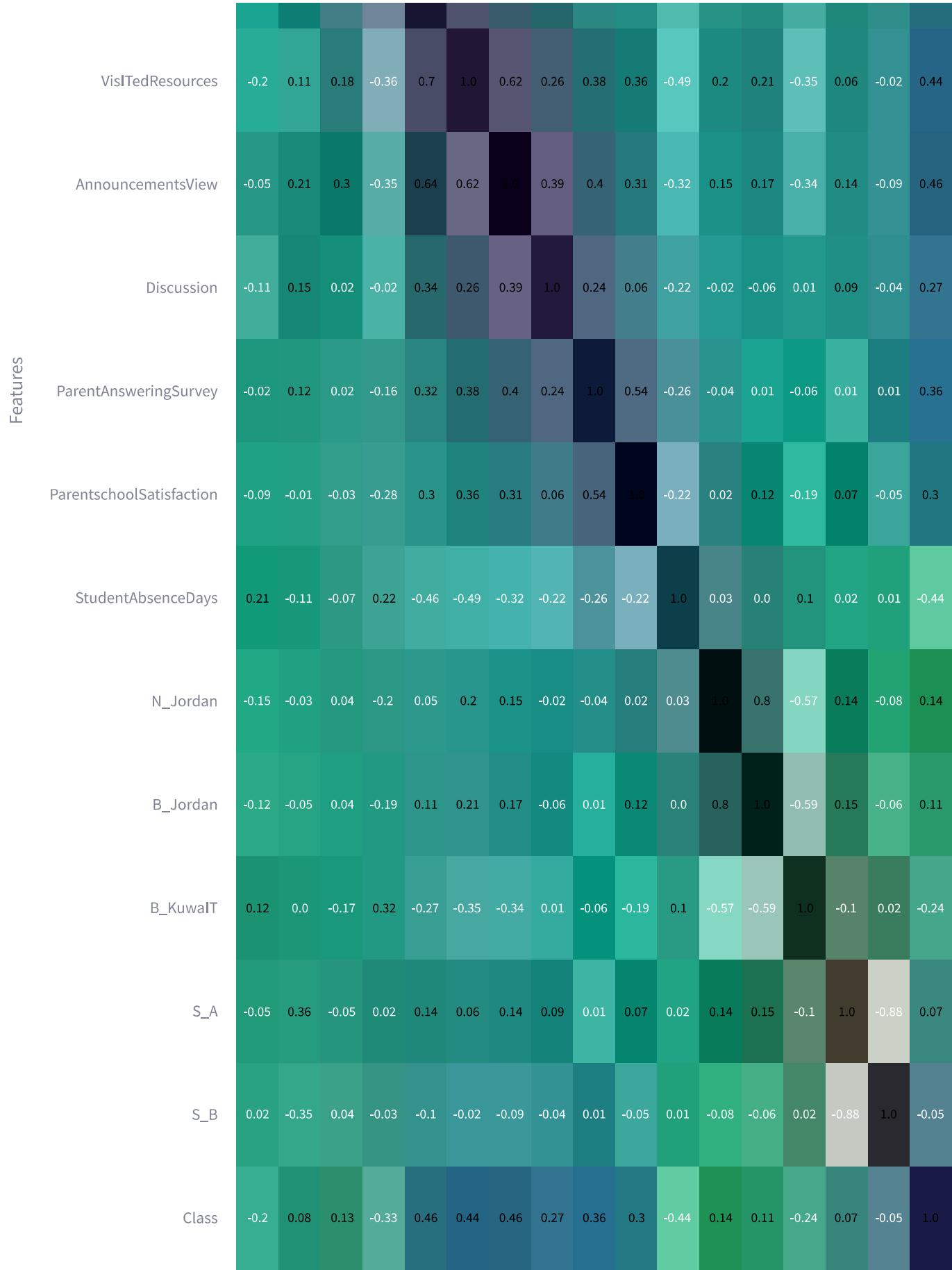
	gender	StageID	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentschoolSatisfaction	StudentAbsenceDays	N_Jordan	B_Jordan	B_Kuwait	app	C_A	C_B	Cross
Features	1.0 0.05 0.05 0.19 -0.14 -0.2 -0.05 -0.11 -0.02 -0.09 0.21 -0.15 -0.12 0.12 nan -0.05 0.02 -0.2	0.05 1.0 0.09 0.02 0.22 0.11 0.21 0.15 0.12 -0.01 -0.11 -0.03 -0.05 0.0 nan 0.36 -0.35 0.08	0.05 0.09 1.0 -0.15 0.19 0.18 0.3 0.02 0.02 -0.03 -0.07 0.04 0.04 -0.17 nan 	0.19 0.02 -0.15 1.0 -0.36 -0.36 -0.35 -0.02 -0.16 -0.28 0.22 -0.2 -0.19 0.32 nan 0.02 -0.03 -0.33	-0.14 0.22 0.19 -0.36 1.0 0.7 0.64 0.34 0.32 0.3 -0.46 0.05 0.11 -0.27 nan 0.14 -0.1 0.46	-0.2 0.11 0.18 -0.36 0.7 1.0 0.62 0.26 0.38 0.36 -0.49 0.2 0.21 -0.35 nan 0.06 -0.02 0.44	-0.05 0.21 0.3 -0.35 0.64 0.62 1.0 0.39 0.4 0.31 -0.32 0.15 0.17 -0.34 nan 0.14 -0.09 0.46	-0.11 0.15 0.02 -0.02 0.34 0.26 0.39 1.0 0.24 0.06 -0.22 -0.02 -0.06 0.01 nan 0.09 -0.04 0.27	-0.02 0.12 0.02 -0.16 0.32 0.38 0.4 0.24 1.0 0.54 -0.26 -0.04 0.01 -0.06 nan 0.01 0.01 0.36	-0.09 -0.01 -0.03 -0.28 0.3 0.36 0.31 0.06 0.54 1.0 -0.22 0.02 0.12 -0.19 nan 0.07 -0.05 0.3	0.21 -0.11 -0.07 0.22 -0.46 -0.49 -0.32 -0.22 -0.26 -0.22 1.0 0.03 0.0 0.1 nan 0.02 0.01 -0.44	-0.15 -0.03 0.04 -0.2 0.05 0.2 0.15 -0.02 -0.04 0.02 0.03 1.0 0.8 -0.57 nan 0.14 -0.08 0.14	-0.12 -0.05 0.04 -0.19 0.11 0.21 0.17 -0.06 0.01 0.12 0.0 0.8 1.0 -0.59 nan 0.15 -0.06 0.11	0.12 0.0 -0.17 0.32 -0.27 -0.35 -0.34 0.01 -0.06 -0.19 0.1 -0.57 -0.59 1.0 nan -0.1 0.02 -0.24	app C_A C_B Cross			



B_Lybia	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
S_A	-0.05	0.36	-0.05	0.02	0.14	0.06	0.14	0.09	0.01	0.07	0.02	0.14	0.15	-0.1	nan	1.0	-0.88	0.07	
S_B	0.02	-0.35	0.04	-0.03	-0.1	-0.02	-0.09	-0.04	0.01	-0.05	0.01	-0.08	-0.06	0.02	nan	-0.88	1.0	-0.05	
Class	-0.2	0.08	0.13	-0.33	0.46	0.44	0.46	0.27	0.36	0.3	-0.44	0.14	0.11	-0.24	nan	0.07	-0.05	1.0	

Update the layout of the heatmap for better visualization:

Correlation Matrix		Features																
		gender	StageID	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	ParentDiscussion	SchoolSatisfaction	StudentAbsenceDays	N_Jordan	B_Jordan	B_Kuwait	S_A	S_B	Class	
gender	1.0	0.05	0.05	0.19	-0.14	-0.2	-0.05	-0.11	-0.02	-0.09	0.21	-0.15	-0.12	0.12	-0.05	0.02	-0.2	
StageID	0.05	1.0	0.09	0.02	0.22	0.11	0.21	0.15	0.12	-0.01	-0.11	-0.03	-0.05	0.0	0.36	-0.35	0.08	
Semester	0.05	0.09	1.0	-0.15	0.19	0.18	0.3	0.02	0.02	-0.03	-0.07	0.04	0.04	-0.17	-0.05	0.04	0.13	
Relation	0.19	0.02	-0.15	1.0	-0.36	-0.36	-0.35	-0.02	-0.16	-0.28	0.22	-0.2	-0.19	0.32	0.02	-0.03	-0.33	
raisedhands	-0.14	0.22	0.19	-0.36	1.0	0.7	0.64	0.34	0.32	0.3	-0.46	0.05	0.11	-0.27	0.14	-0.1	0.46	



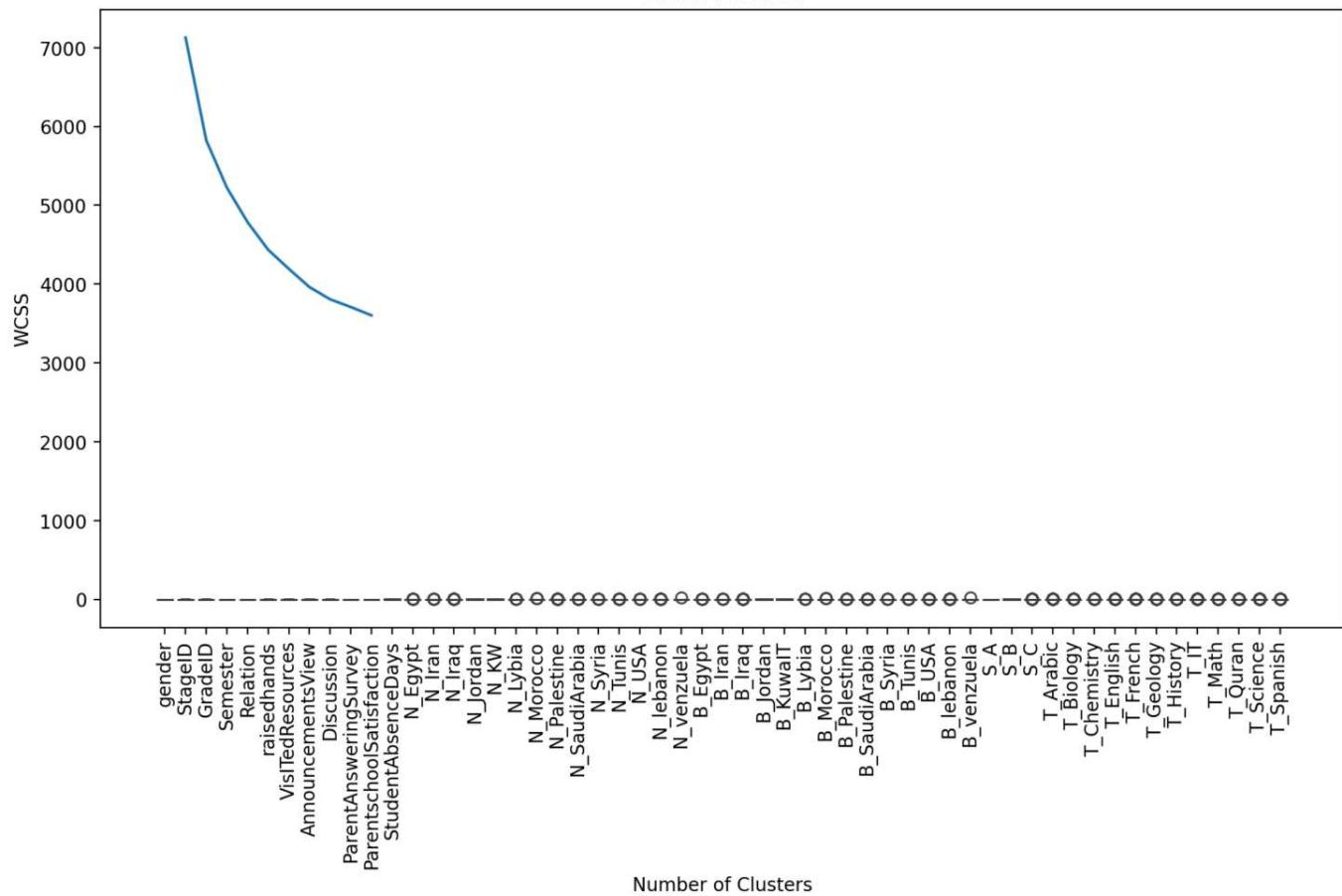
Display the column names of the DataFrame to verify the current set of features:

Current set of features:

```
▼ [ 0 : "gender" 1 : "StageID" 2 : "Semester" 3 : "Relation" 4 : "raisedhands" 5 : "VisITEDResources" 6 : "AnnouncementsView" 7 : "Discussion" 8 : "ParentAnsweringSurvey" 9 : "ParentschoolSatisfaction" 10 : "StudentAbsenceDays" 11 : "N_Jordan" 12 : "B_Jordan" 13 : "B_KuwaIT" 14 : "S_A" 15 : "S_B" 16 : "Class" ]
```

Plot the WCSS values to use the Elbow Method:

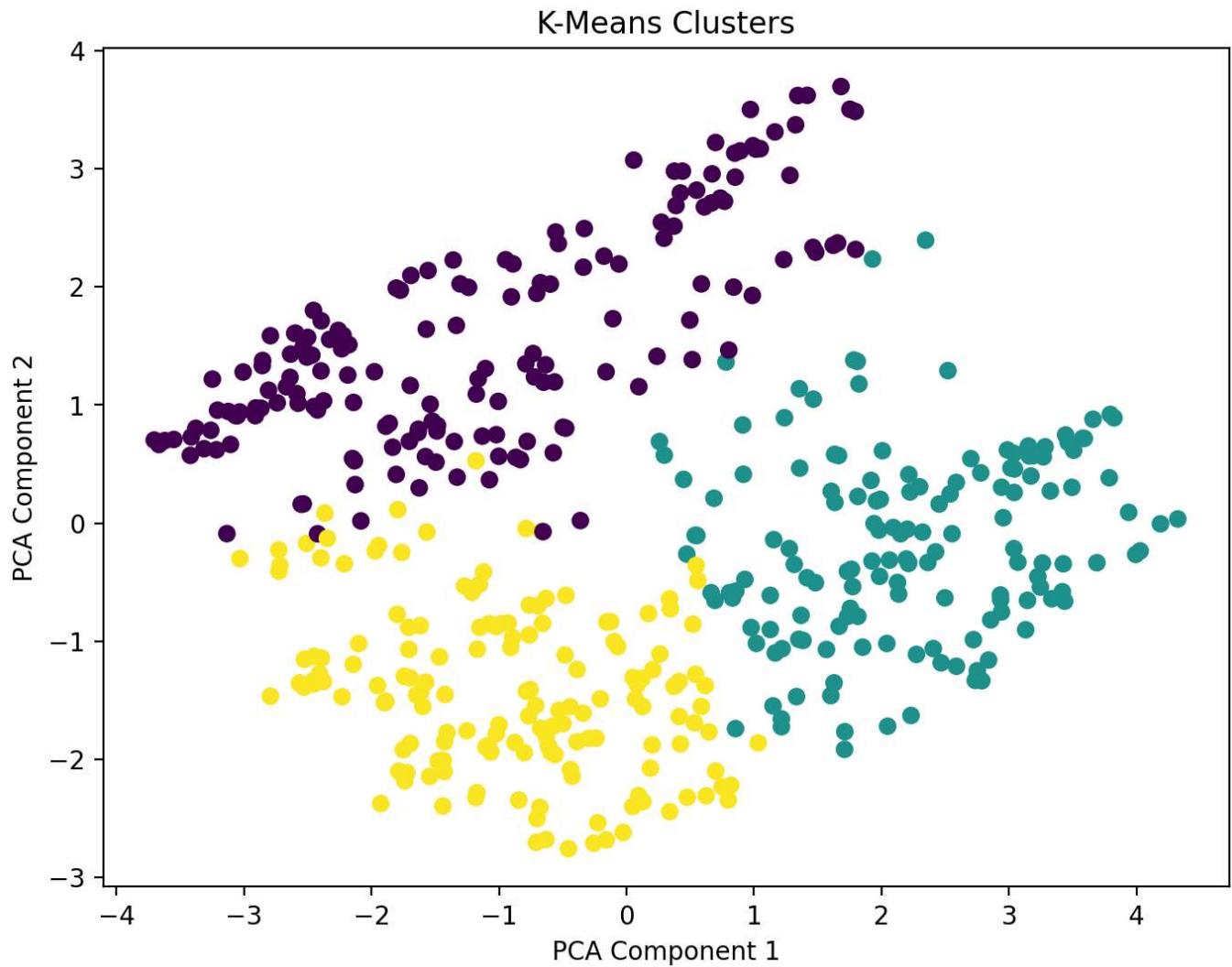
Elbow Method



Display the DataFrame with cluster labels:

	gender	StageID	Semester	Relation	raisedhands	VisiTedResources	AnnouncementsView	Discuss
0	0.76	-1.0896	-0.9793	0.8373	-1.0368	-1.181	-1.3542	-0.8
1	0.76	-1.0896	-0.9793	0.8373	-0.8742	-1.0598	-1.3166	-0.6
2	0.76	-1.0896	-0.9793	0.8373	-1.1994	-1.4538	-1.4294	-0.
3	0.76	-1.0896	-0.9793	0.8373	-0.5491	-0.9083	-1.2414	-0.3
4	0.76	-1.0896	-0.9793	0.8373	-0.2239	-0.1508	-0.9782	0.2
5	-1.3158	-1.0896	-0.9793	0.8373	-0.1589	-0.7568	-0.9406	0.9
6	0.76	0.5691	-0.9793	0.8373	-0.3865	-1.3022	-1.4294	-0.9
7	0.76	0.5691	-0.9793	0.8373	0.1012	-1.3629	-0.8654	-0.7
8	-1.3158	0.5691	-0.9793	0.8373	-1.1343	-1.0295	-0.8278	0.2
9	-1.3158	0.5691	-0.9793	0.8373	0.7515	0.7583	-0.4894	0.9

Plot the clusters:



Show the contribution of each feature to each principal component:

Principal Component 1:

gender: 0.1485

StagID: -0.0829

Semester: -0.1268

Relation: 0.2888

raisedhands: -0.3675

VisITedResources: -0.4083

AnnouncementsView: -0.3192

Discussion: -0.1436

ParentAnsweringSurvey: -0.2569

ParentschoolSatisfaction: -0.2738

StudentAbsenceDays: 0.2704

N_Jordan: -0.2267

B_Jordan: -0.2464

B_KuwaIT: 0.3096

S_A: -0.1280

S_B: 0.0942

Principal Component 2:

gender: -0.0333

StageID: -0.0300

Semester: -0.0324

Relation: 0.0073

raisedhands: -0.1917

VisITedResources: -0.1480

AnnouncementsView: -0.1164

Discussion: -0.1692

ParentAnsweringSurvey: -0.2922

ParentschoolSatisfaction: -0.1693

StudentAbsenceDays: 0.2687

N_Jordan: 0.5122

B_Jordan: 0.4906

B_Kuwait: -0.3404

S_A: 0.2315

S_B: -0.1929

Show the rows of the DataFrame to verify the contributions:

	Component	Feature	Value
0	Principal Component 1	gender	0.1485
1	Principal Component 1	StageID	-0.0829
2	Principal Component 1	Semester	-0.1268
3	Principal Component 1	Relation	0.2888
4	Principal Component 1	raisedhands	-0.3675
5	Principal Component 1	VisitedResources	-0.4083
6	Principal Component 1	AnnouncementsView	-0.3192
7	Principal Component 1	Discussion	-0.1436
8	Principal Component 1	ParentAnsweringSurvey	-0.2569
9	Principal Component 1	ParentschoolSatisfaction	-0.2738

Display the DataFrame to verify the contributions:

	Feature	Principal Component 1	Principal Component 2
0	AnnouncementsView	-0.3192	-0.1164
1	B_Jordan	-0.2464	0.4906
2	B_KuwaIT	0.3096	-0.3404
3	Discussion	-0.1436	-0.1692
4	N_Jordan	-0.2267	0.5122
5	ParentAnsweringSurvey	-0.2569	-0.2922
6	ParentschoolSatisfaction	-0.2738	-0.1693
7	Relation	0.2888	0.0073
8	S_A	-0.128	0.2315
9	S_B	0.0942	-0.1929

Sort and display the features by their contribution to each principal component:

Sorted features for Principal Component 1:

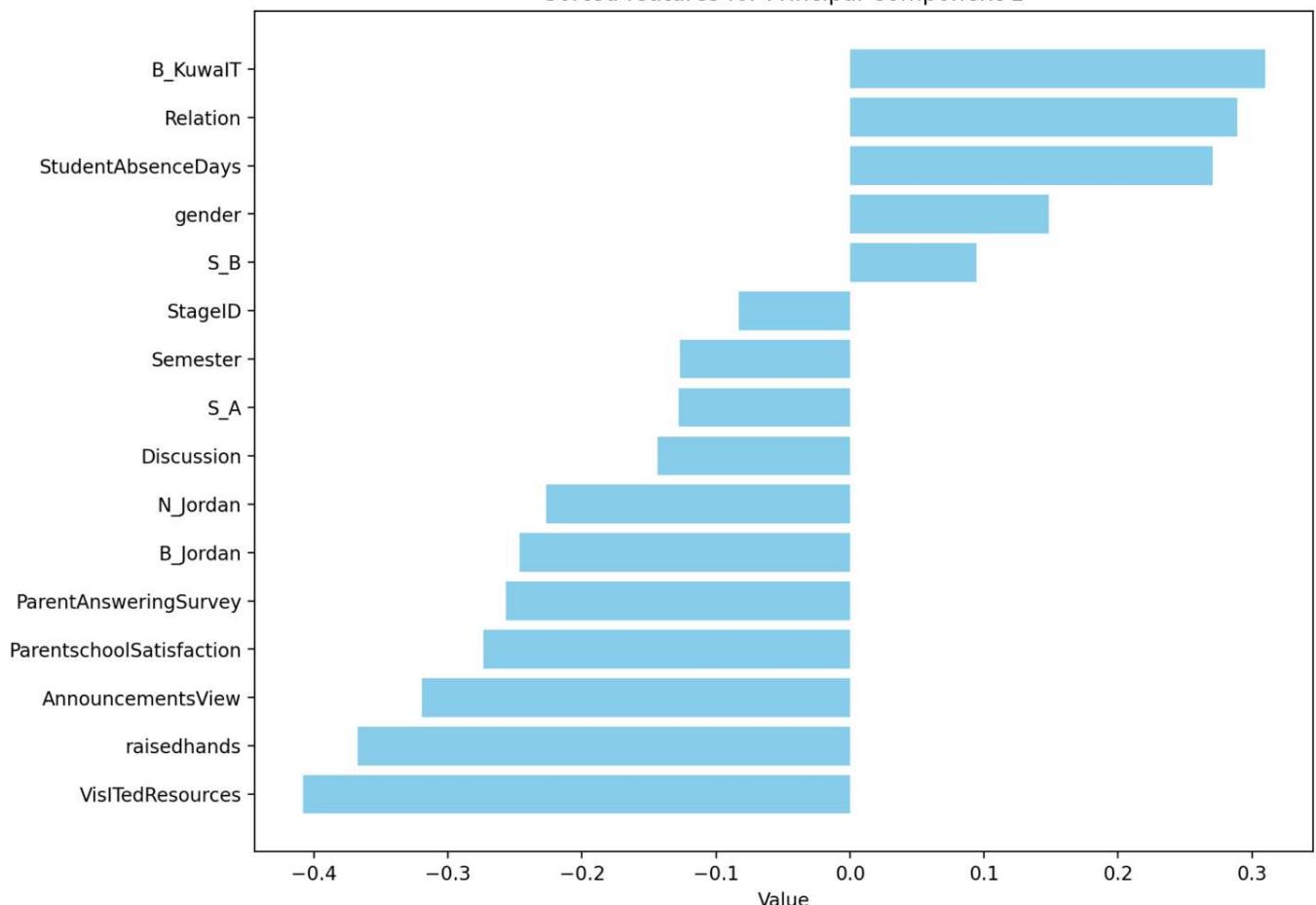
	Feature	Principal Component 1
2	B_KuwaIT	0.3096
7	Relation	0.2888
12	StudentAbsenceDays	0.2704
14	gender	0.1485
9	S_B	0.0942
11	StageID	-0.0829
10	Semester	-0.1268
8	S_A	-0.128
3	Discussion	-0.1436
4	N_Jordan	-0.2267

Sorted features for Principal Component 2:

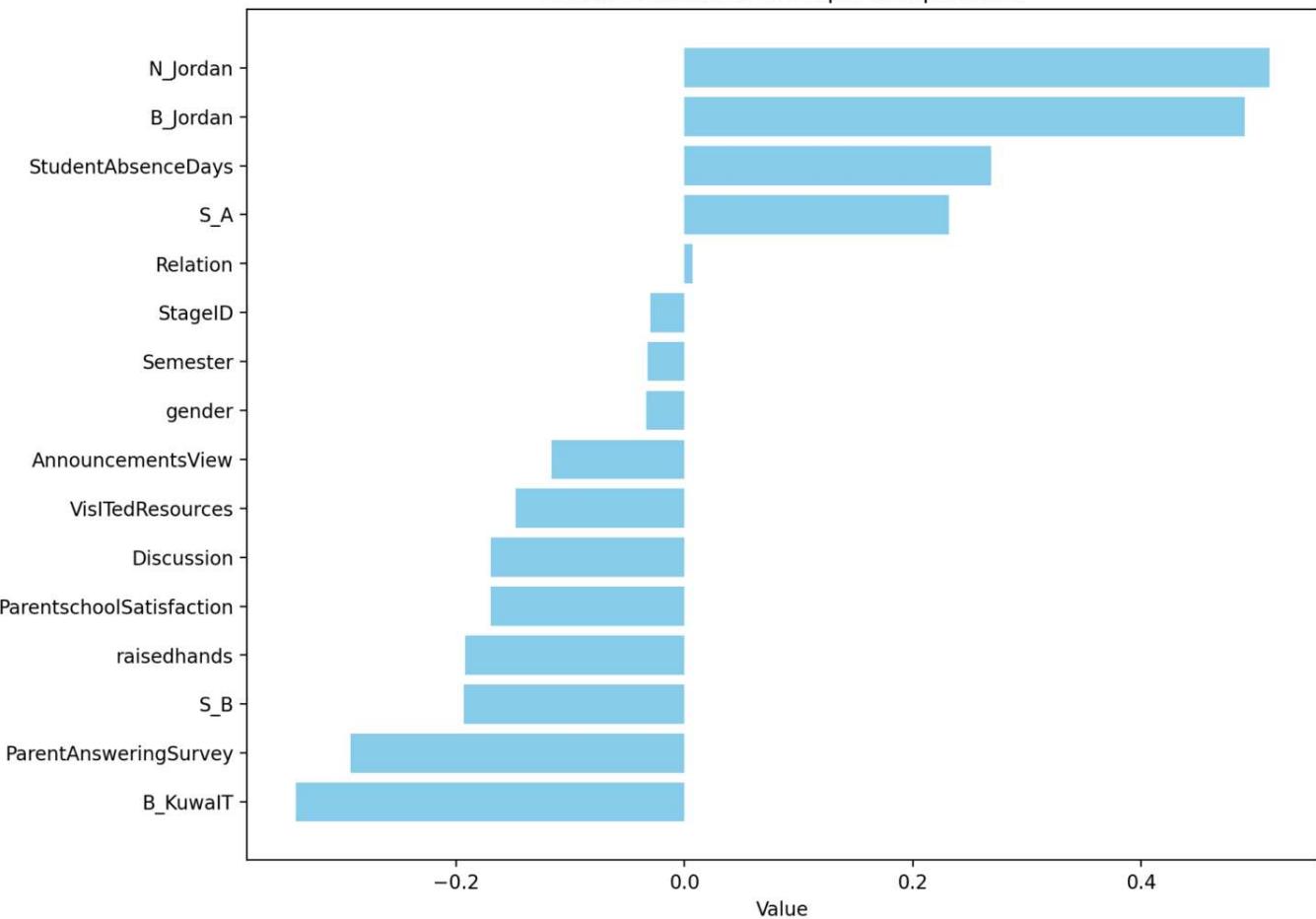
	Feature	Principal Component 2
4	N_Jordan	0.5122
1	B_Jordan	0.4906
12	StudentAbsenceDays	0.2687
8	S_A	0.2315
7	Relation	0.0073
11	StageID	-0.03
10	Semester	-0.0324
14	gender	-0.0333
0	AnnouncementsView	-0.1164
13	VisitedResources	-0.148

Plot the sorted feature contributions:

Sorted features for Principal Component 1



Sorted features for Principal Component 2

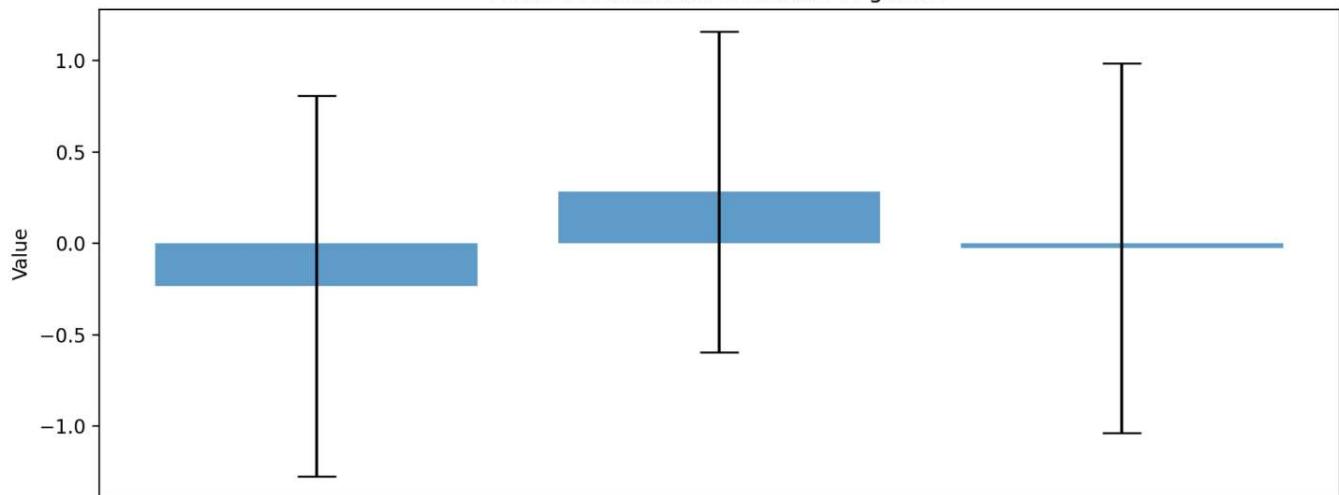


Display mean and standard deviation for each cluster:

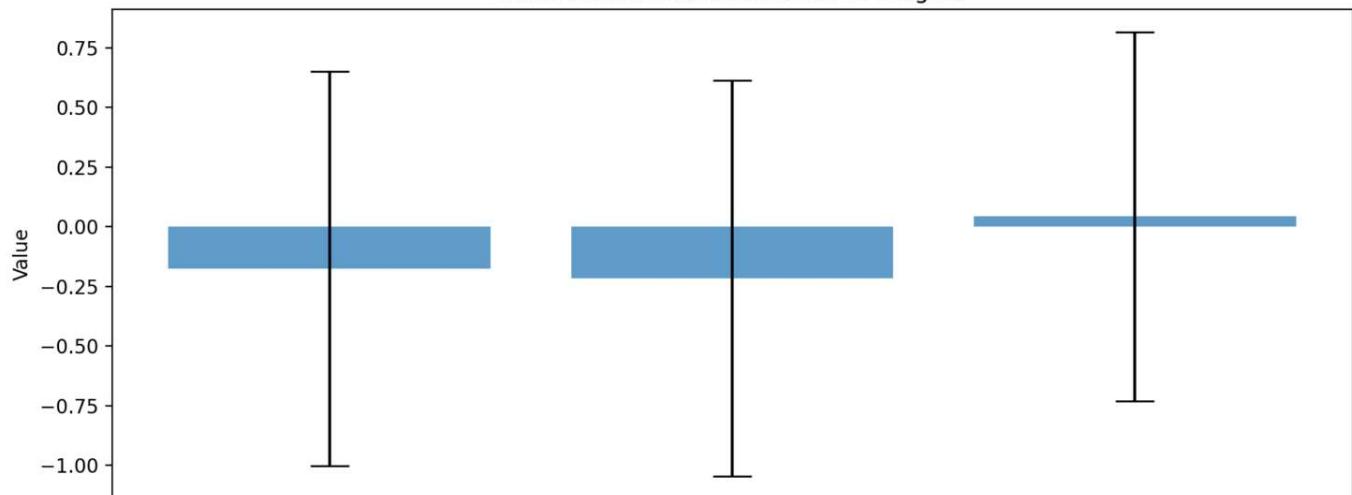
Cluster	gender	gender	StageID	StageID	Semester	Semester	Relation	Relation	raisedhands	raisedhands
None	mean	std	mean	std	mean	std	mean	std	mean	std
0	-0.2339	1.0401	-0.1748	0.8275	0.027	1.0032	-0.2709	1.0147	0.0567	1.0000
1	0.282	0.8768	-0.2166	0.8309	-0.3213	0.943	0.4898	0.7675	-0.8753	0.0000
2	-0.0265	1.0101	0.0437	0.7741	0.2756	0.9703	-0.1848	1.019	0.6319	0.0000

Mean and Standard Deviation values for each columns:

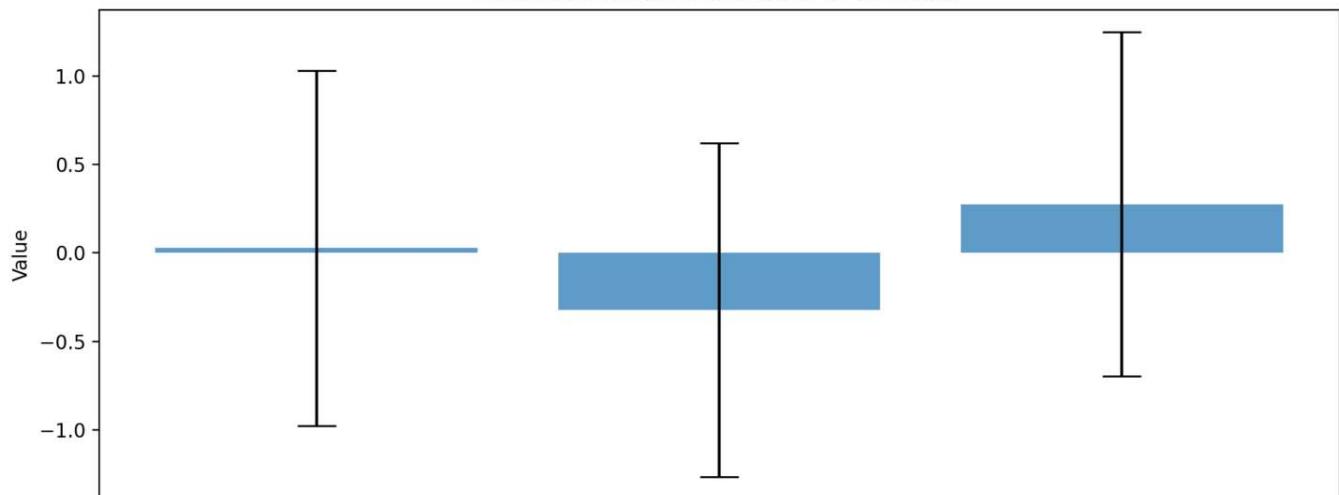
Mean and Standard Deviation of gender



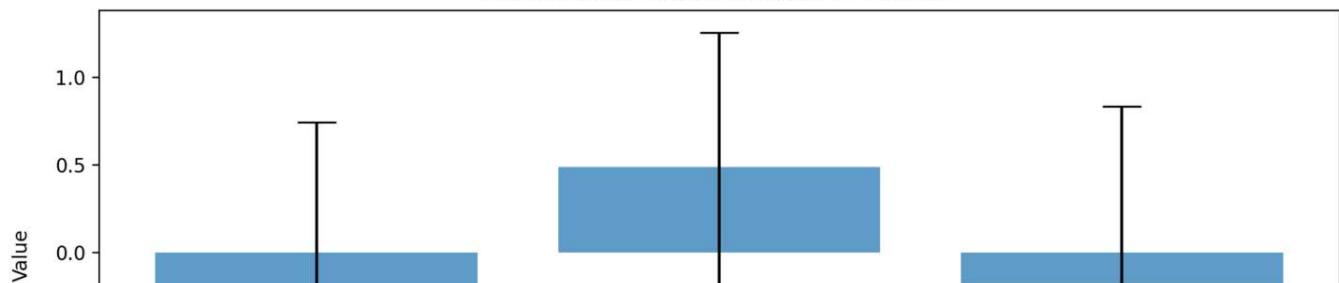
Mean and Standard Deviation of StageID

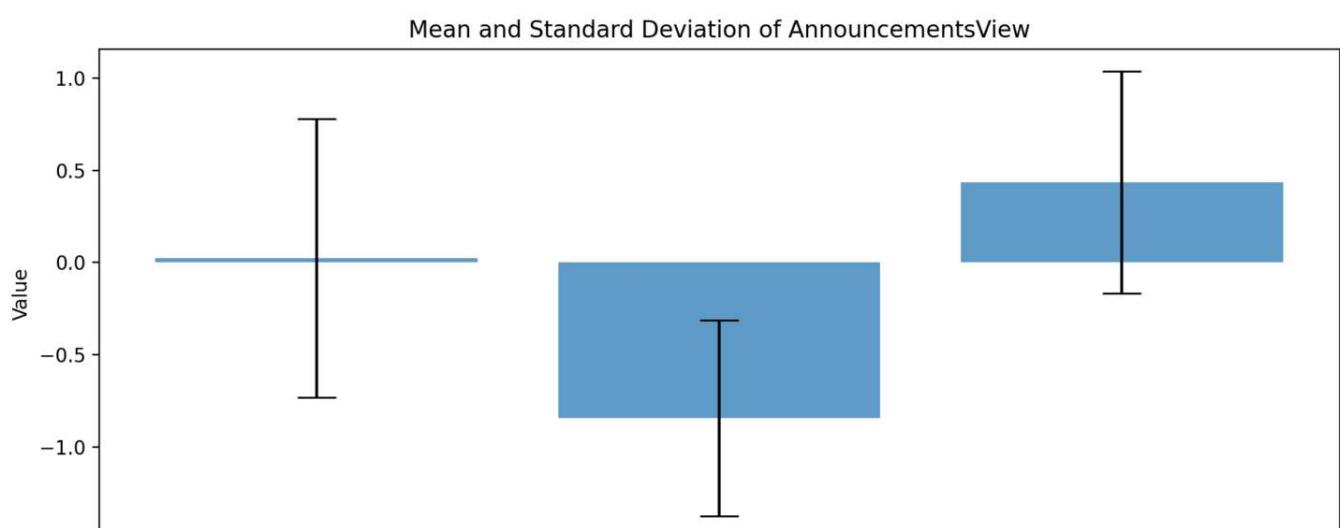
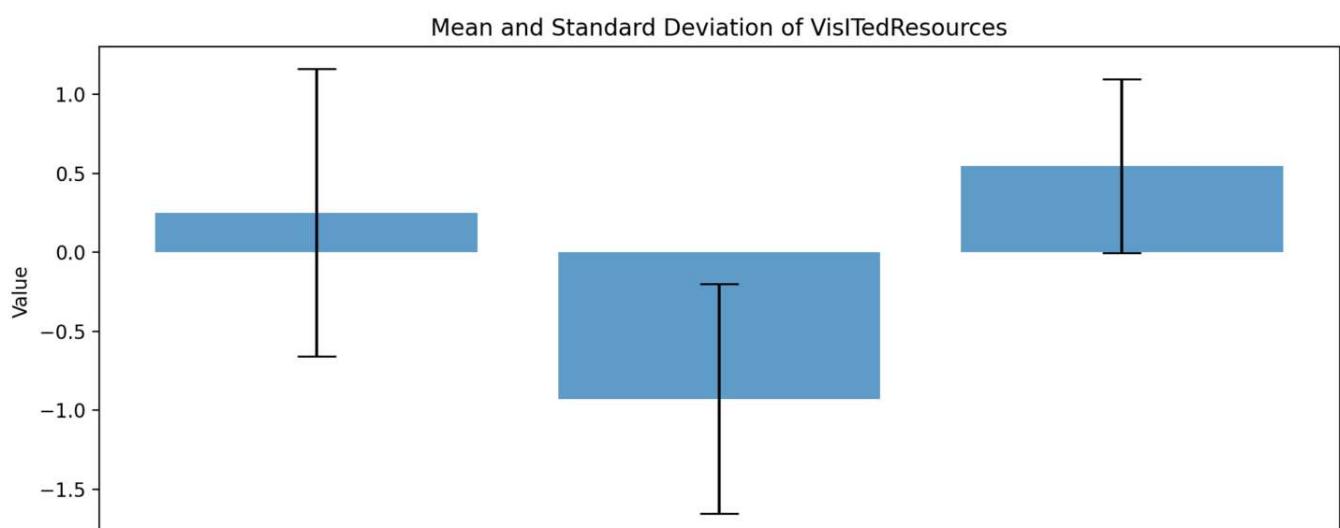
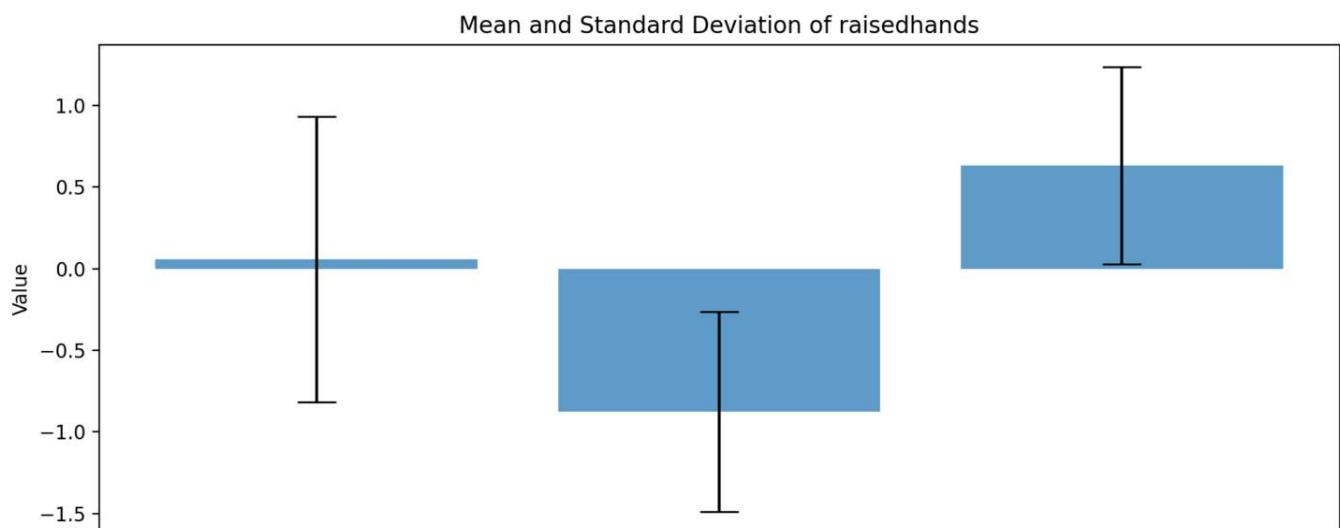
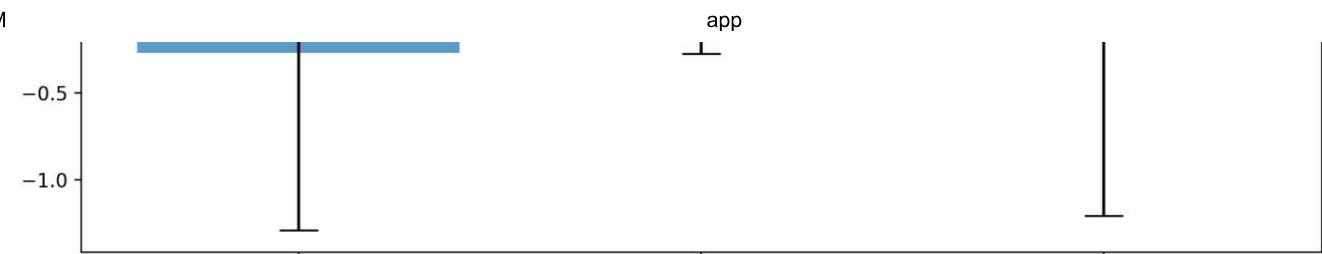


Mean and Standard Deviation of Semester



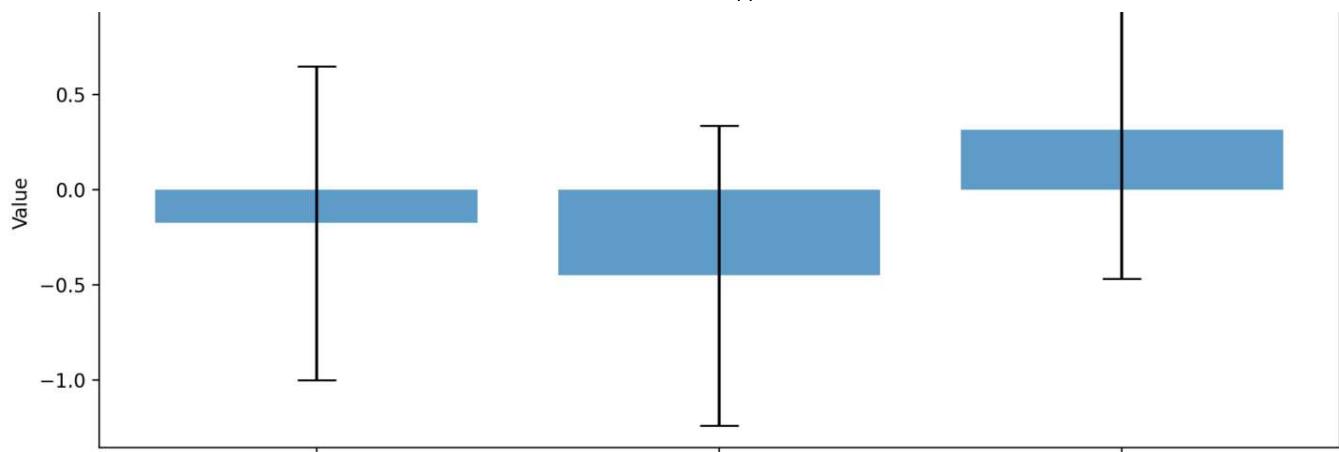
Mean and Standard Deviation of Relation



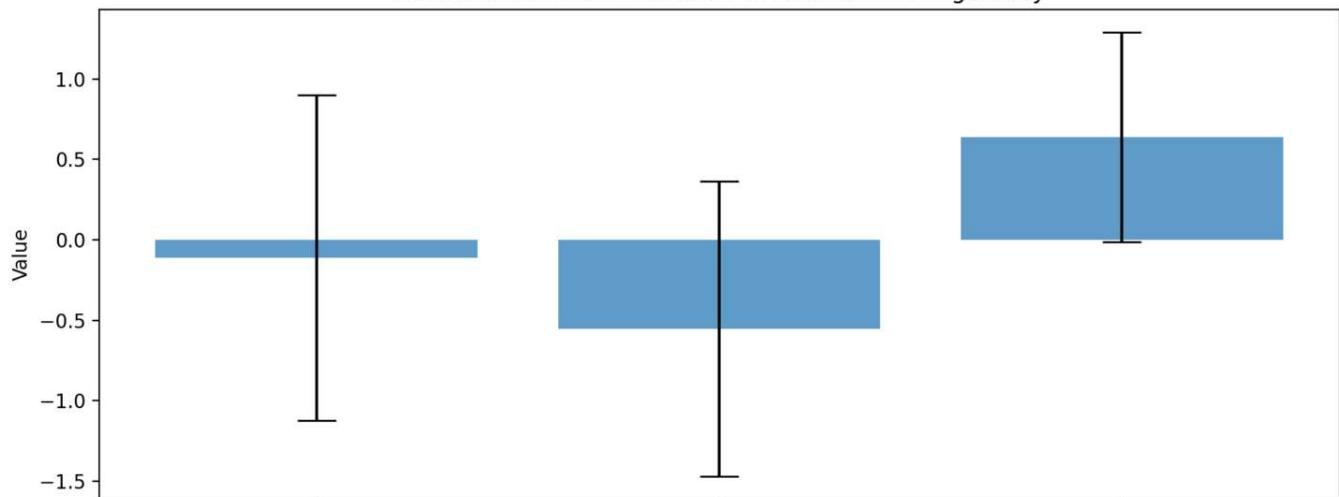


Mean and Standard Deviation of Discussion

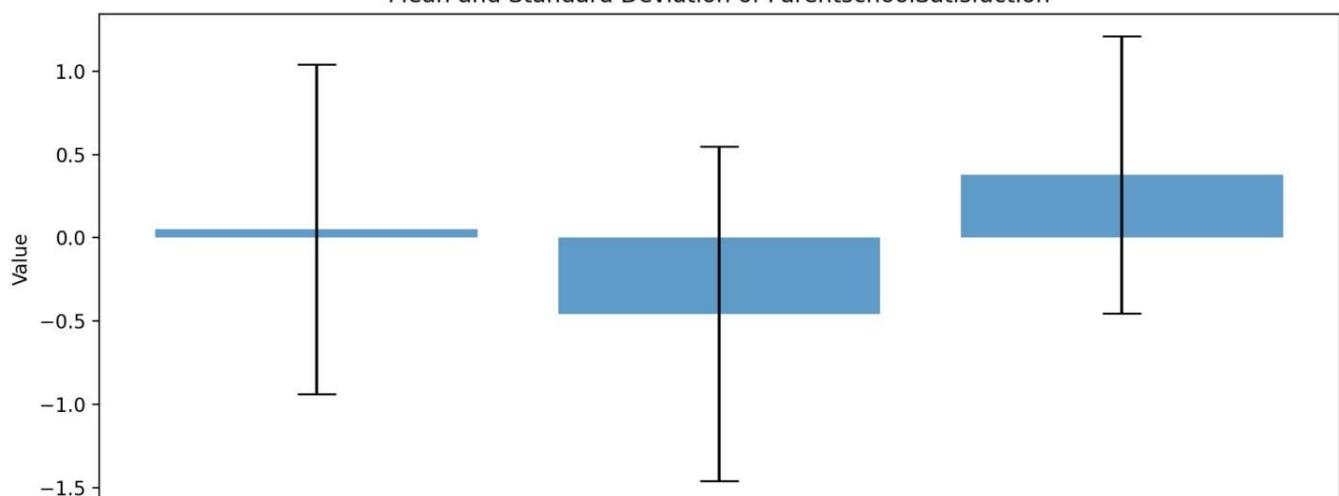




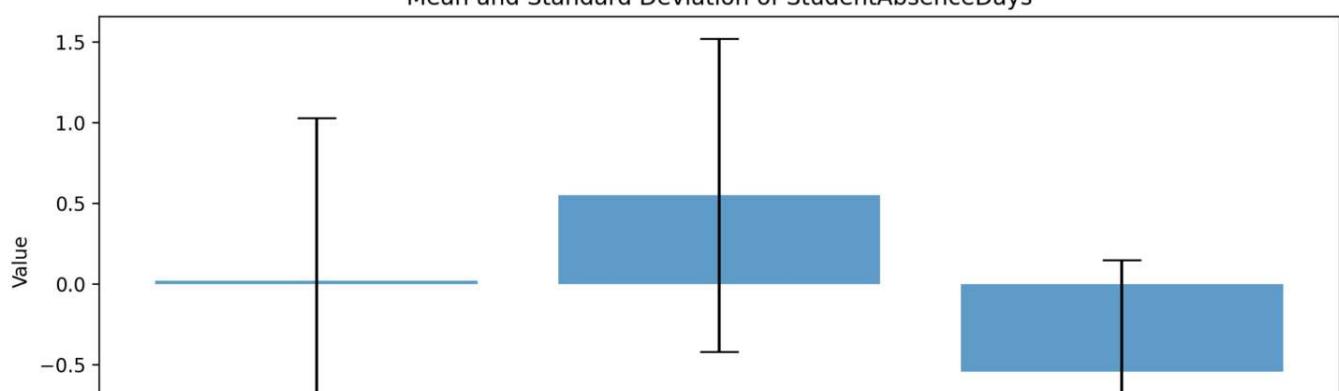
Mean and Standard Deviation of ParentAnsweringSurvey

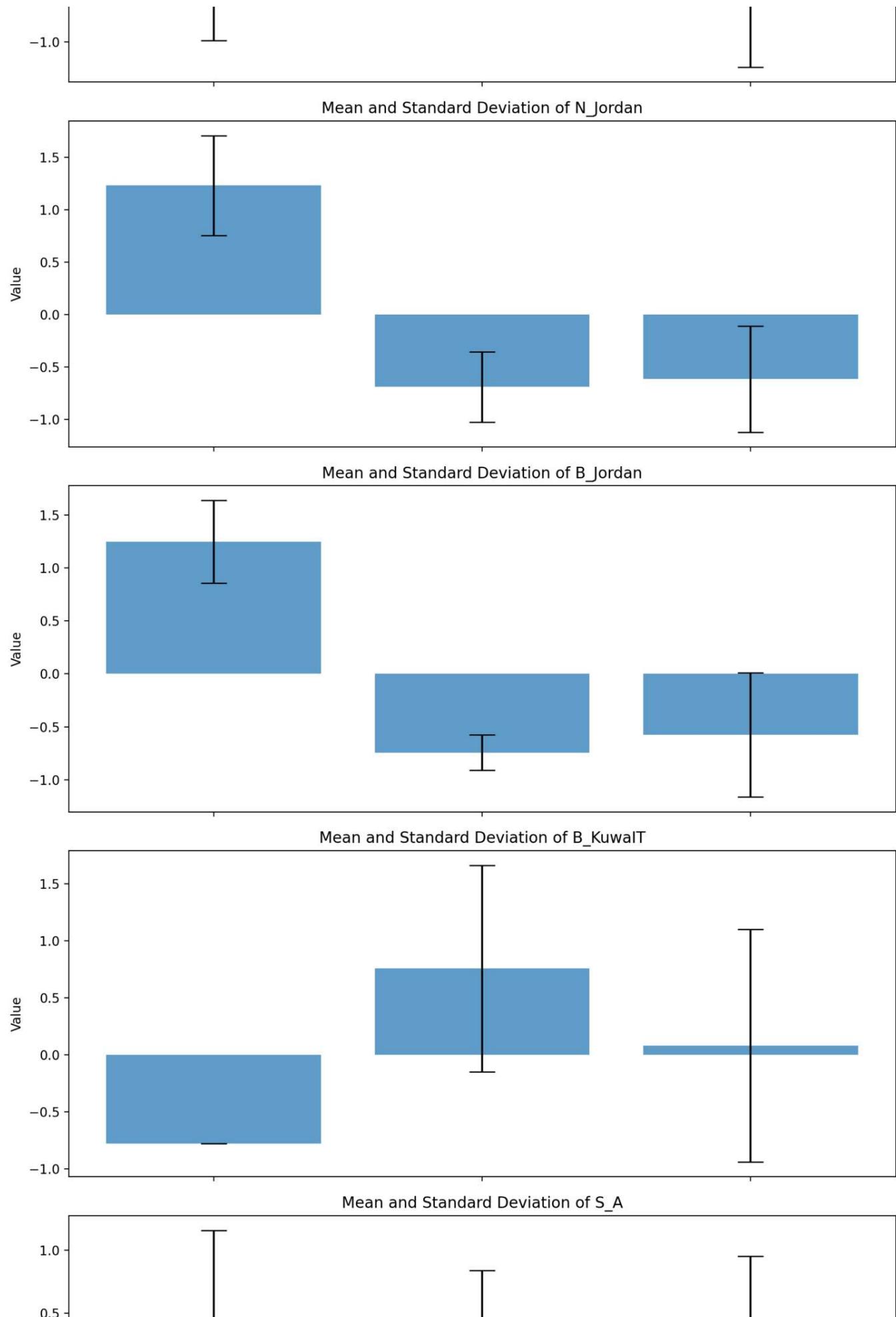


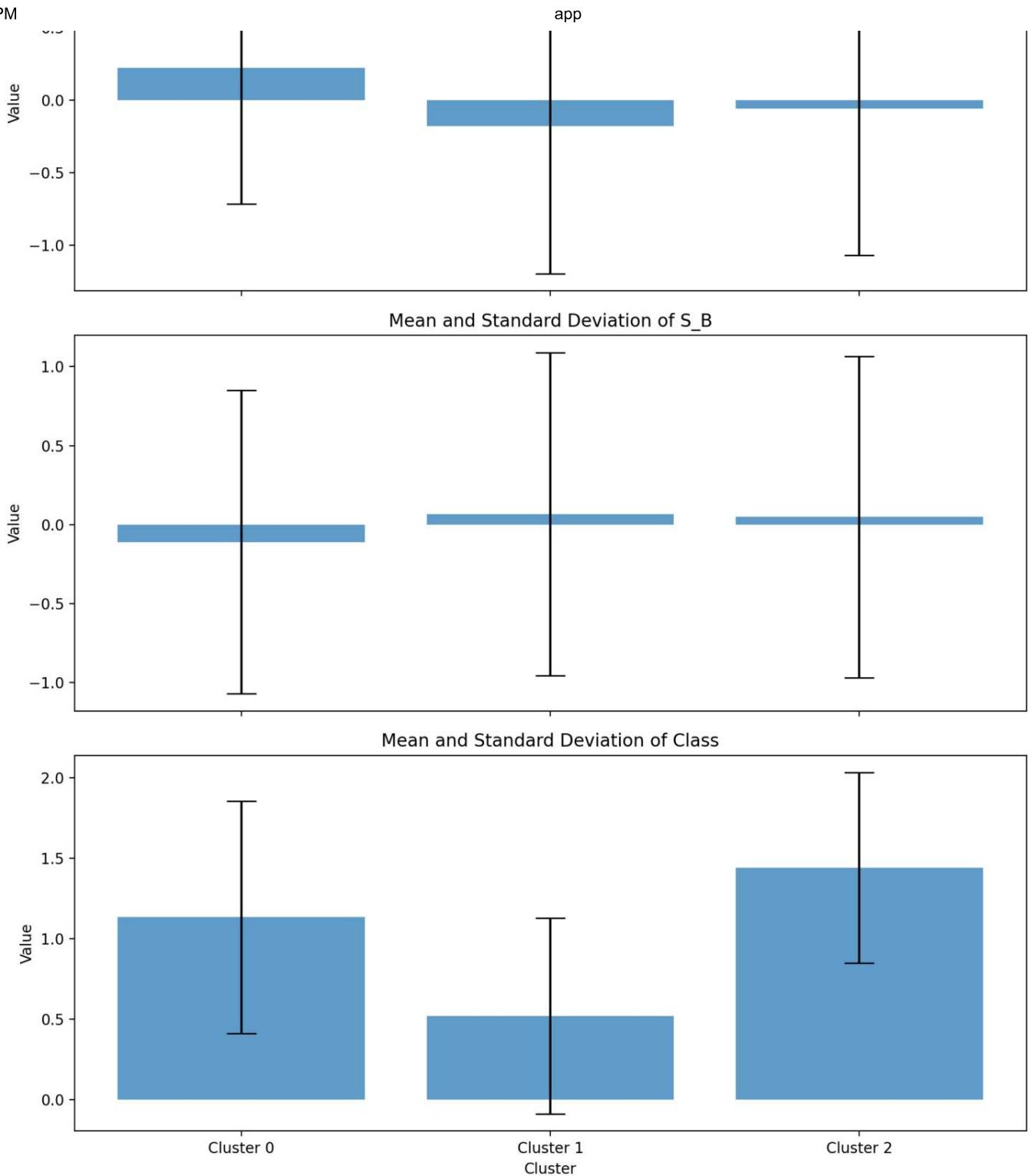
Mean and Standard Deviation of ParentschoolSatisfaction



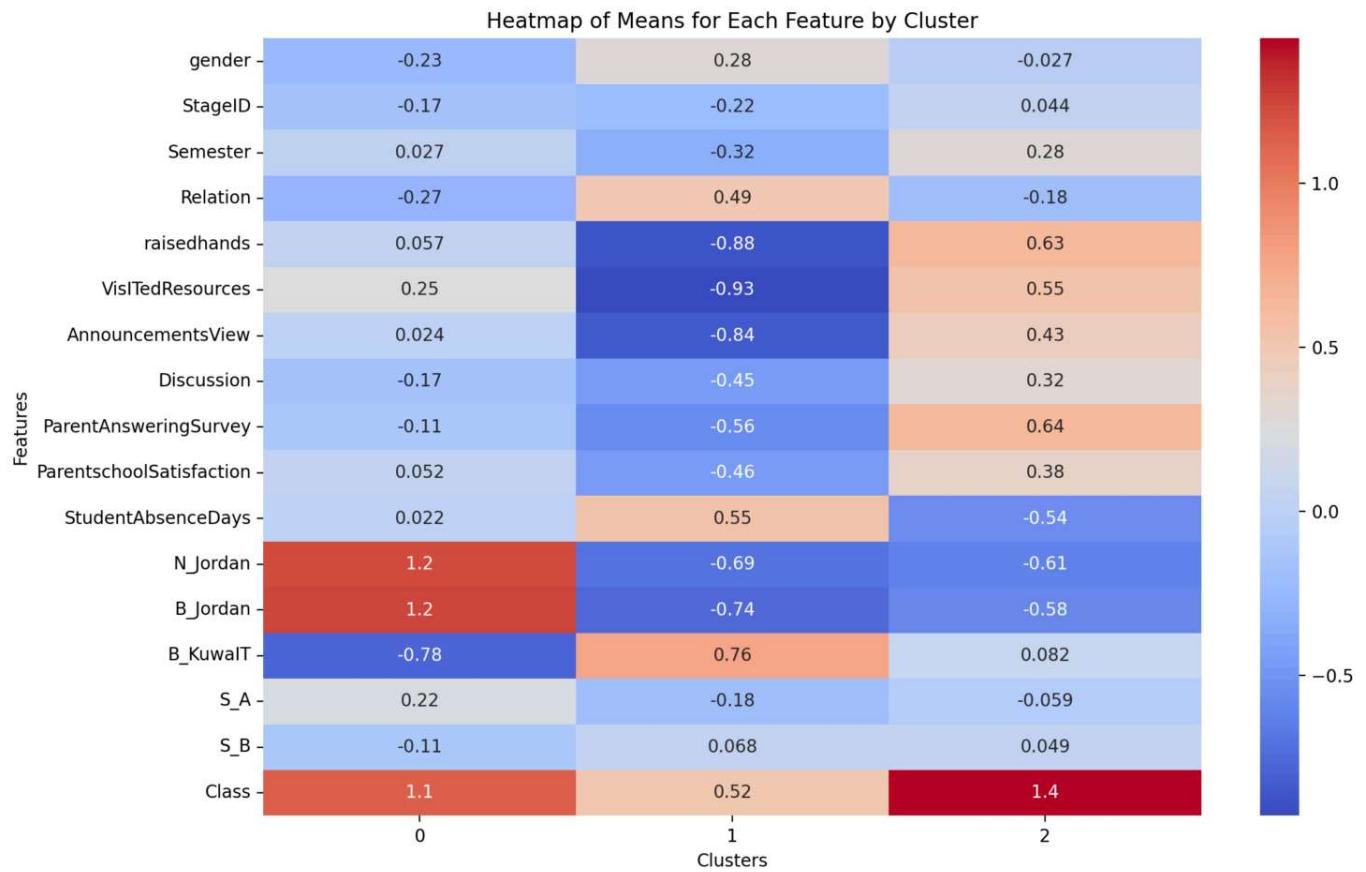
Mean and Standard Deviation of StudentAbsenceDays



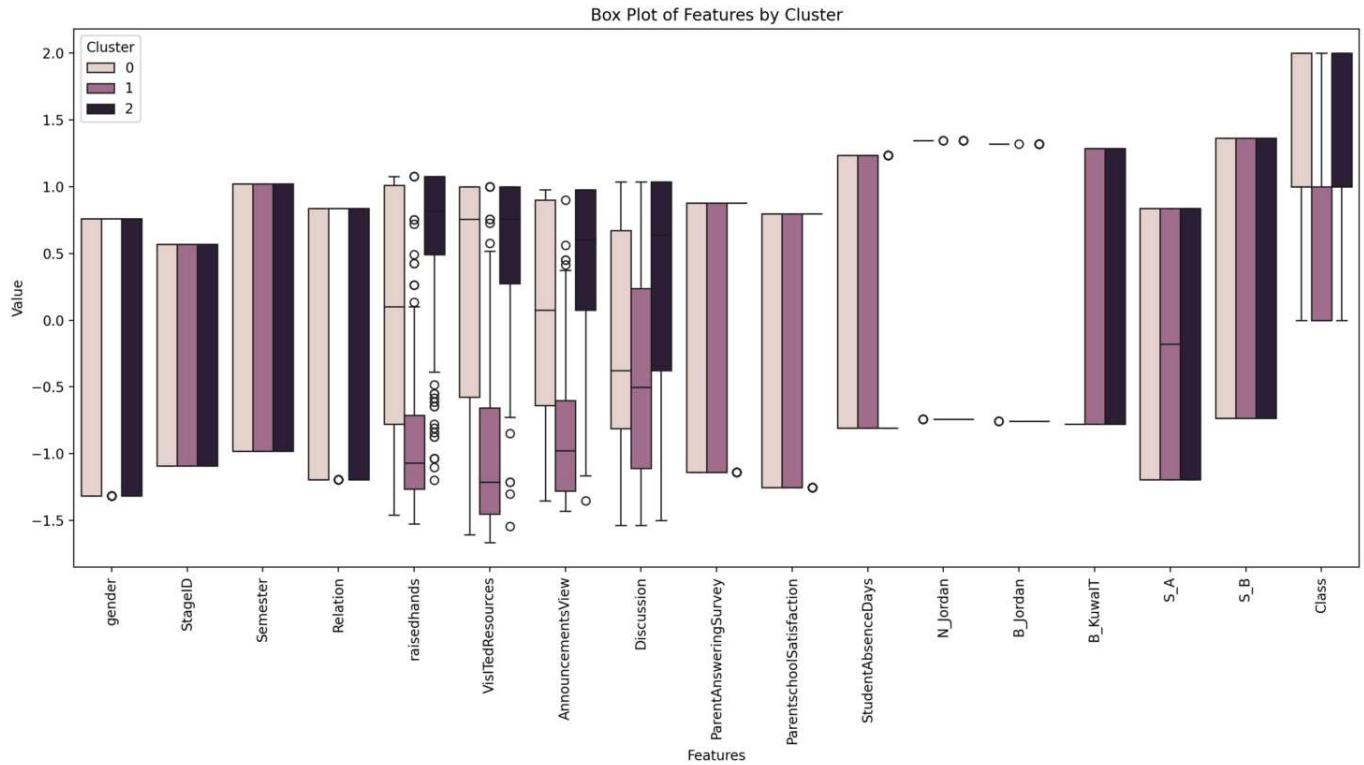




Heatmap of Means for Each Feature by Cluster:



Box Plot of Features by Cluster:



Check for missing values in each column of vif_dropped_data:

Missing Values:

	0
gender	0
StageID	0
Semester	0
Relation	0
raisedhands	0
VisITedResources	0
AnnouncementsView	0
Discussion	0
ParentAnsweringSurvey	0
ParentschoolSatisfaction	0

Check for missing values in each column:

Missing Values:

	0
gender	0
StageID	0
Semester	0
Relation	0
raisedhands	0
VisitedResources	0
AnnouncementsView	0
Discussion	0
ParentAnsweringSurvey	0
ParentschoolSatisfaction	0

Result of Feature Importances from Random Forest

Accuracy: 0.972027972027972

Classification Report:

precision recall f1-score support

0	0.98	1.00	0.99	50
1	0.96	0.98	0.97	48
2	0.98	0.93	0.95	45
accuracy			0.97	143

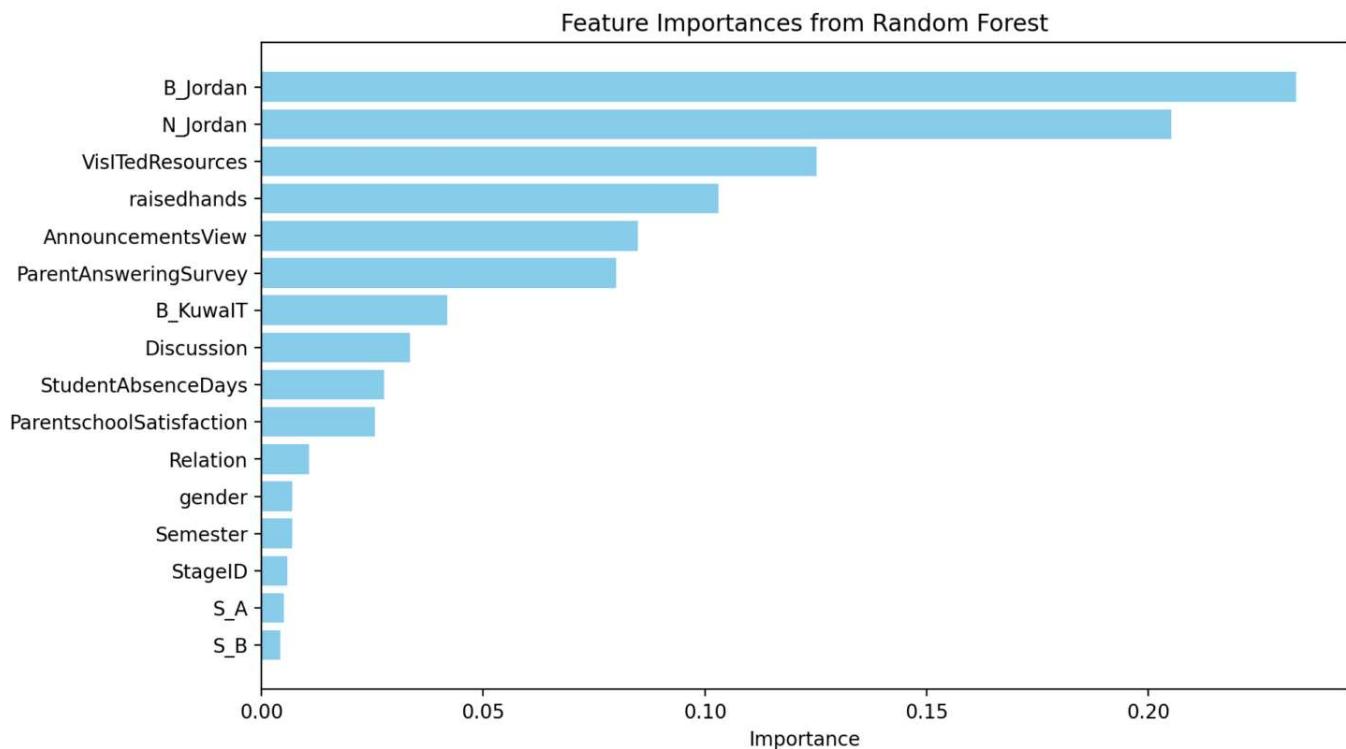
macro avg 0.97 0.97 0.97 143 weighted avg 0.97 0.97 0.97 143

Confusion Matrix:

0	1	2
50	0	0
0	47	1
1	2	42

Feature Importances:

	Importance
B_Jordan	0.2333
N_Jordan	0.2052
VisITedResources	0.1252
raisedhands	0.1031
AnnouncementsView	0.0849
ParentAnsweringSurvey	0.08
B_KuwaitT	0.0418
Discussion	0.0334
StudentAbsenceDays	0.0277
ParentschoolSatisfaction	0.0256



Classification Report for Decision Tree:

Accuracy (Decision Tree): 0.9230769230769231

Classification Report (Decision Tree):

precision recall f1-score support

0	0.98	1.00	0.99	50
1	0.85	0.96	0.90	48
2	0.95	0.80	0.87	45
accuracy			0.92	143

macro avg 0.93 0.92 0.92 143 weighted avg 0.93 0.92 0.92 143

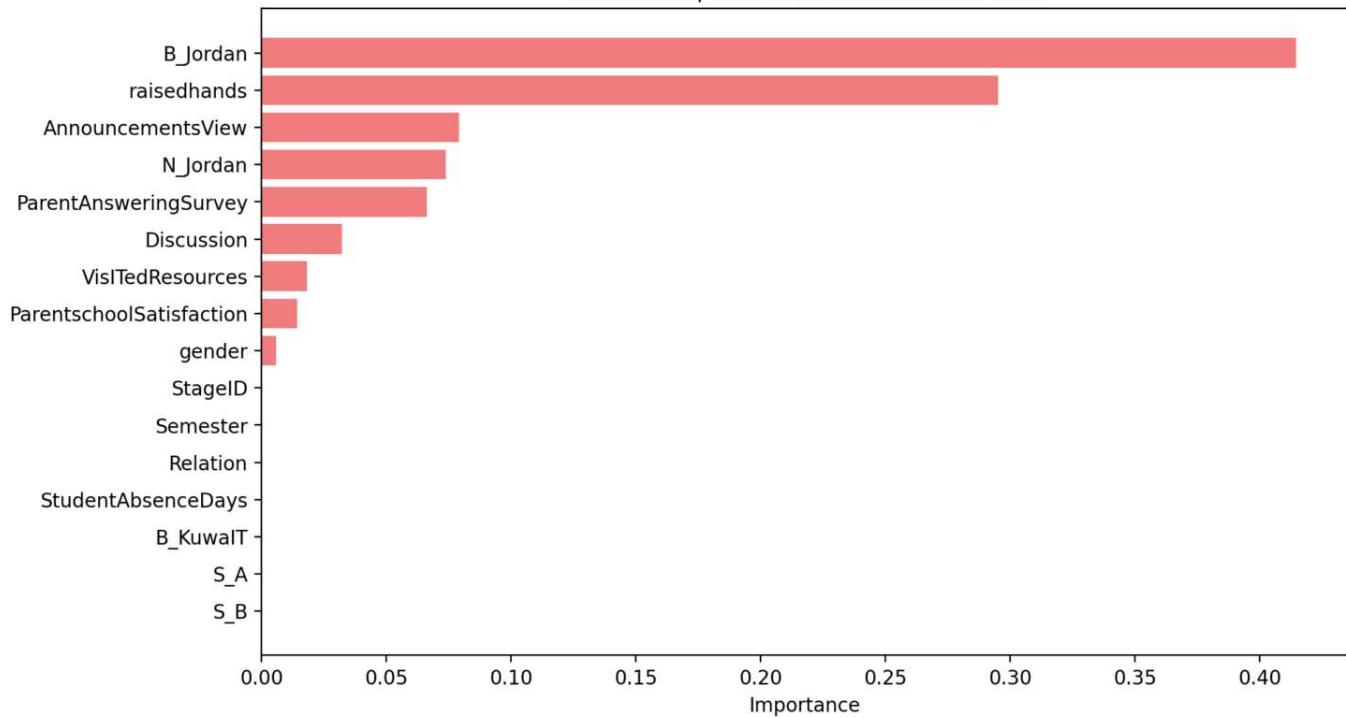
Confusion Matrix (Decision Tree):

0	1	2	
0	50	0	0
1	0	46	2
2	1	8	36

Feature Importances (Decision Tree):

	Importance
B_Jordan	0.4147
raisedhands	0.2951
AnnouncementsView	0.0791
N_Jordan	0.0737
ParentAnsweringSurvey	0.0664
Discussion	0.0322
VisitedResources	0.0184
ParentschoolSatisfaction	0.0143
gender	0.006
StageID	0

Feature Importances from Decision Tree



Classification Report for Logistic Regression:

Accuracy (Logistic Regression): 0.965034965034965

Classification Report (Logistic Regression):

precision recall f1-score support

0	0.98	1.00	0.99	50
1	0.94	0.98	0.96	48
2	0.98	0.91	0.94	45
accuracy			0.97	143

macro avg 0.97 0.96 0.96 143 weighted avg 0.97 0.97 0.96 143

Confusion Matrix (Logistic Regression):

0	1	2	
0	50	0	0
1	0	47	1
2	1	3	41

Feature Importances (Logistic Regression):

	Importance
B_Jordan	1.9054
N_Jordan	1.8465
S_A	0.1793
AnnouncementsView	0.0938
StageID	0.061
raisedhands	0.0482
VisitedResources	0.0045
S_B	0.0026
StudentAbsenceDays	-0.0594
Discussion	-0.0631

Feature Importances from Logistic Regression

