```
PS C:\Users\Aniket\OneDrive\Desktop\marvellousinfo\Assignments\Assignment_33> & C:/Users/Aniket/AppData/Local/Programs/Python/Python313/python.exe c:/Users/Aniket/OneDrive/Desktop/marvellousinfo/Assignments/Assignment_33/Marvellous_student_cluster.py

(395, 33)

  school sex  age address famsize Pstatus  Medu  Fedu     Mjob      Fjob  ... famrel freetime goout  Dalc  Walc health absences  G1  G2  G3

0     GP   F   18       U     GT3       A     4     4  at_home   teacher  ...      4        3      4     1     1      3        6   5   6   6

1     GP   F   17       U     GT3       T     1     1  at_home     other  ...      5        3      3     1     1      3        4   5   5   6

2     GP   F   15       U     LE3       T     1     1  at_home     other  ...      4        3      2     2     3      3       10   7   8  10

3     GP   F   15       U     GT3       T     4     2   health  services  ...      3        2      2     1     1      5        2  15  14  15

4     GP   F   16       U     GT3       T     3     3    other     other  ...      4        3      2     1     2      5        4   6  10  10


[5 rows x 33 columns]

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 395 entries, 0 to 394

Data columns (total 33 columns):

 #   Column      Non-Null Count  Dtype

---  ------      --------------  -----

 0   school      395 non-null    object

 1   sex         395 non-null    object

 2   age         395 non-null    int64

 3   address     395 non-null    object

 4   famsize     395 non-null    object

 5   Pstatus     395 non-null    object
```

```
 6   Medu        395 non-null    int64
 7   Fedu        395 non-null    int64
 8   Mjob        395 non-null    object
 9   Fjob        395 non-null    object
 10  reason      395 non-null    object
 11  guardian    395 non-null    object
 12  traveltime  395 non-null    int64
 13  studytime   395 non-null    int64
 14  failures    395 non-null    int64
 15  schoolsup   395 non-null    object
 16  famsup      395 non-null    object
 17  paid        395 non-null    object
 18  activities  395 non-null    object
 19  nursery     395 non-null    object
 20  higher      395 non-null    object
 21  internet    395 non-null    object
 22  romantic    395 non-null    object
 23  famrel      395 non-null    int64
 24  freetime    395 non-null    int64
 25  goout       395 non-null    int64
 26  Dalc        395 non-null    int64
 27  Walc        395 non-null    int64
 28  health      395 non-null    int64
 29  absences    395 non-null    int64
 30  G1          395 non-null    int64
 31  G2          395 non-null    int64
 32  G3          395 non-null    int64
dtypes: int64(16), object(17)
```

memory usage: 102.0+ KB

None

(395, 34)

school sex age address famsize Pstatus Medu Fedu Mjob ... goout Dalc Walc health absences G1 G2 G3 clusters

0 GP F 18 U GT3 A 4 4 at_home ... 4 1 1 3 6 5 6 6 0

1 GP F 17 U GT3 T 1 1 at_home ... 3 1 1 3 4 5 5 6 0

2 GP F 15 U LE3 T 1 1 at_home ... 2 2 3 3 10 7 8 10 2

3 GP F 15 U GT3 T 4 2 health ... 2 1 1 5 2 15 14 15 1

4 GP F 16 U GT3 T 3 3 other ... 2 1 2 5 4 6 10 10 0

[5 rows x 34 columns]

look overall clustering of data with performer groups:

school sex age address famsize Pstatus Medu Fedu Mjob ... Dalc Walc health absences G1 G2 G3 clusters performance_group

0 GP F 18 U GT3 A 4 4 at_home ... 1 1 3 6 5 6 6 0 Struggling

1 GP F 17 U GT3 T 1 1 at_home ... 1 1 3 4 5 5 6 0 Struggling

2 GP F 15 U LE3 T 1 1 at_home ... 2 3 3 10 7 8 10 2 Top Performer

3 GP F 15 U GT3 T 4 2 health ... 1 1 5 2 15 14 15 1 Average

4 GP F 16 U GT3 T 3 3 other ... 1 2 5 4 6 10 10 0 Struggling

[5 rows x 35 columns]

Silhouette Score: 0.2591861035590761

cluster centers:

	G1	G2	G3	studytime	failures	absences
0	-0.502954	-0.360518	-0.287187	0.044848	-0.298660	0.243770
1	0.999722	0.925579	0.869761	0.124895	-0.321305	-0.177139
2	-1.123843	-1.352393	-1.423886	-0.459859	1.726772	-0.259669

PS C:\Users\Aniket\OneDrive\Desktop\marvellousinfo\Assignments\Assignment_33>

Conclusion:

The K-Means clustering segmented the students into three performance groups: **Top Performer, Average, and Struggling**.

- **Top performers** have high grades and consistent study habits.
- **Average students** have moderate grades and balanced behavior.
- **Struggling students** have lower grades and slightly higher absences.

The **silhouette score of 0.26** indicates moderate clustering quality, suggesting that while the groups are meaningful, some students' performance patterns overlap across clusters.