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

In [ ]:
```
df=pd.read_csv('college_student_management_data.csv')
```

In [ ]:
```
df
```

Out [13]:

| | student_id | age | gender | major | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **1540** | S1541 | 24 | Male | Business | 3.61 | 6 | 70.5 | 0.77 | Leave | 26 | 64 |
| **1541** | S1542 | 24 | Other | Business | 2.80 | 4 | 77.4 | 0.96 | Graduated | 12 | 19 |
| **1542** | S1543 | 21 | Male | Computer Science | 3.61 | 5 | 66.3 | 0.99 | Leave | 1 | 89 |
| **1543** | S1544 | 22 | Male | Arts | 3.18 | 4 | 97.9 | 0.85 | Graduated | 20 | 15 |
| **1544** | S1545 | 25 | Other | Engineering | 2.33 | 4 | 90.7 | 0.66 | Leave | 32 | 57 |

1545 rows × 15 columns

In [ ]:
```
df.head()
```

Out [14]:

| | student_id | age | gender | major | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |

In [ ]:
```
df.tail()
```

Out [15]:

| | student_id | age | gender | major | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1540** | S1541 | 24 | Male | Business | 3.61 | 6 | 70.5 | 0.77 | Leave | 26 | 64 |
| **1541** | S1542 | 24 | Other | Business | 2.80 | 4 | 77.4 | 0.96 | Graduated | 12 | 19 |
| **1542** | S1543 | 21 | Male | Computer Science | 3.61 | 5 | 66.3 | 0.99 | Leave | 1 | 89 |
| **1543** | S1544 | 22 | Male | Arts | 3.18 | 4 | 97.9 | 0.85 | Graduated | 20 | 15 |
| **1544** | S1545 | 25 | Other | Engineering | 2.33 | 4 | 90.7 | 0.66 | Leave | 32 | 57 |

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

Out [16]:

| | 0 |
|---|---|
| **student_id** | 0 |
| **age** | 0 |
| **gender** | 0 |
| **major** | 0 |
| **GPA** | 0 |
| **course_load** | 0 |
| **avg_course_grade** | 0 |
| **attendance_rate** | 0 |
| **enrollment_status** | 0 |
| **lms_logins_past_month** | 0 |
| **avg_session_duration_minutes** | 0 |
| **assignment_submission_rate** | 0 |
| **forum_participation_count** | 0 |

| | 0 |
|---|---|
| video_completion_rate | 0 |
| risk_level | 0 |

**dtype:** int64

In [ ]:
```
df.loc[10:15][['student_id','age','major','avg_course_grade']]
```

Out [17]:

| | student_id | age | major | avg_course_grade |
|---|---|---|---|---|
| **10** | S011 | 20 | Arts | 69.6 |
| **11** | S012 | 24 | Engineering | 75.5 |
| **12** | S013 | 20 | Arts | 63.2 |
| **13** | S014 | 20 | Computer Science | 65.9 |
| **14** | S015 | 25 | Computer Science | 71.9 |
| **15** | S016 | 22 | Arts | 63.6 |

In [ ]:
```
df.iloc[10:15,0:4]
```

Out [19]:

| | student_id | age | gender | major |
|---|---|---|---|---|
| **10** | S011 | 20 | Other | Arts |
| **11** | S012 | 24 | Female | Engineering |
| **12** | S013 | 20 | Other | Arts |
| **13** | S014 | 20 | Other | Computer Science |
| **14** | S015 | 25 | Male | Computer Science |

In [ ]:
```
df.columns
```

Out [20]:
```
Index(['student_id', 'age', 'gender', 'major', 'GPA', 'course_load',
       'avg_course_grade', 'attendance_rate', 'enrollment_status',
       'lms_logins_past_month', 'avg_session_duration_minutes',
       'assignment_submission_rate', 'forum_participation_count',
       'video_completion_rate', 'risk_level'],
      dtype='object')
```

In [ ]:
```
df.shape
```

Out [21]:
```
(1545, 15)
```

In [ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1545 entries, 0 to 1544
Data columns (total 15 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   student_id                    1545 non-null   object 
 1   age                           1545 non-null   int64  
 2   gender                        1545 non-null   object 
 3   major                         1545 non-null   object 
 4   GPA                           1545 non-null   float64
 5   course_load                   1545 non-null   int64  
 6   avg_course_grade              1545 non-null   float64
 7   attendance_rate               1545 non-null   float64
 8   enrollment_status             1545 non-null   object 
 9   lms_logins_past_month         1545 non-null   int64  
 10  avg_session_duration_minutes  1545 non-null   int64  
 11  assignment_submission_rate    1545 non-null   float64
 12  forum_participation_count     1545 non-null   int64  
 13  video_completion_rate         1545 non-null   float64
 14  risk_level                    1545 non-null   object 
dtypes: float64(5), int64(5), object(5)
memory usage: 181.2+ KB
```

In [ ]:
```
df.describe()
```

Out [24]:

| | age | GPA | course_load | avg_course_grade | attendance_rate | lms_logins_past_month | avg_session_duration_minutes | assignm |
|---|---|---|---|---|---|---|---|---|
| **count** | 1545.000000 | 1545.000000 | 1545.000000 | 1545.000000 | 1545.000000 | 1545.000000 | 1545.000000 | 1545.00 |
| **mean** | 21.482848 | 3.012324 | 4.487379 | 79.914887 | 0.794460 | 19.344984 | 49.365049 | 0.75220 |
| **std** | 2.300287 | 0.571395 | 1.098163 | 11.528625 | 0.115142 | 11.515003 | 23.195192 | 0.14435 |
| **min** | 18.000000 | 2.000000 | 3.000000 | 60.000000 | 0.600000 | 0.000000 | 10.000000 | 0.50000 |
| **25%** | 19.000000 | 2.520000 | 4.000000 | 69.900000 | 0.690000 | 10.000000 | 29.000000 | 0.63000 |
| **50%** | 22.000000 | 3.030000 | 5.000000 | 79.600000 | 0.790000 | 19.000000 | 49.000000 | 0.76000 |
| **75%** | 23.000000 | 3.510000 | 5.000000 | 89.700000 | 0.890000 | 29.000000 | 70.000000 | 0.88000 |
| **max** | 25.000000 | 4.000000 | 6.000000 | 100.000000 | 1.000000 | 39.000000 | 89.000000 | 1.00000 |

In [ ]:
```
df.dtypes
```

Out [25]:

| | 0 |
|---|---|
| student_id | object |

| | 0 |
|---|---|
| **age** | int64 |
| **gender** | object |
| **major** | object |
| **GPA** | float64 |
| **course_load** | int64 |
| **avg_course_grade** | float64 |
| **attendance_rate** | float64 |
| **enrollment_status** | object |
| **lms_logins_past_month** | int64 |
| **avg_session_duration_minutes** | int64 |
| **assignment_submission_rate** | float64 |
| **forum_participation_count** | int64 |
| **video_completion_rate** | float64 |
| **risk_level** | object |

**dtype:** object

In [ ]:
```
df.head(11)
```

Out [27]:

| | student_id | age | gender | major | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_se |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |
| **5** | S006 | 25 | Male | Engineering | 2.53 | 4 | 61.0 | 0.81 | Active | 18 | 65 |
| **6** | S007 | 22 | Female | Engineering | 2.04 | 3 | 87.6 | 0.91 | Active | 27 | 39 |
| **7** | S008 | 22 | Other | Computer Science | 3.17 | 4 | 72.6 | 0.82 | Active | 22 | 31 |
| **8** | S009 | 24 | Other | Computer Science | 3.68 | 6 | 62.2 | 0.82 | Active | 3 | 84 |
| **9** | S010 | 19 | Other | Engineering | 3.10 | 4 | 84.4 | 0.65 | Active | 19 | 22 |
| **10** | S011 | 20 | Other | Arts | 2.44 | 4 | 69.6 | 0.65 | Leave | 18 | 80 |

In [ ]:
```
df['major'].value_counts()
```

Out [28]:

| major | count |
|---|---|
| **Arts** | 404 |
| **Computer Science** | 389 |
| **Engineering** | 384 |
| **Business** | 368 |

**dtype:** int64

In [ ]:
```
df['gender'].value_counts()
```

Out [31]:

| gender | count |
|---|---|
| **Male** | 531 |
| **Female** | 525 |
| **Other** | 489 |

**dtype:** int64

In [ ]:
```
df['course_load'].isnull()
```

Out [32]:

| | course_load |
|---|---|
| **0** | False |
| **1** | False |
| **2** | False |
| **3** | False |

| | course_load |
|---|---|
| **4** | False |
| **...** | ... |
| **1540** | False |
| **1541** | False |
| **1542** | False |
| **1543** | False |
| **1544** | False |

1545 rows × 1 columns

**dtype:** bool

In [ ]:
```
df.notnull()
```

Out [36]:

| | student_id | age | gender | major | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | True | True | True | True | True | True | True | True | True | True | True |
| **1** | True | True | True | True | True | True | True | True | True | True | True |
| **2** | True | True | True | True | True | True | True | True | True | True | True |
| **3** | True | True | True | True | True | True | True | True | True | True | True |
| **4** | True | True | True | True | True | True | True | True | True | True | True |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **1540** | True | True | True | True | True | True | True | True | True | True | True |
| **1541** | True | True | True | True | True | True | True | True | True | True | True |
| **1542** | True | True | True | True | True | True | True | True | True | True | True |
| **1543** | True | True | True | True | True | True | True | True | True | True | True |
| **1544** | True | True | True | True | True | True | True | True | True | True | True |

1545 rows × 15 columns

In [ ]:
```
df.size
```

Out [37]: 23175

In [ ]:
```
(df.isnull().sum().sum() / df.size)*100
```

Out [40]: np.float64(0.0)

In [ ]:
```
df.dropna(inplace = True)
```

In [ ]:
```
df.fillna('na',inplace = True)
```

In [ ]:
```
df.rename(columns = {'major':'field'},inplace = True)
```

In [ ]:
```
df.head()
```

Out [46]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |

In [ ]:
```
df['lms_logins_past_month'].dtype
```

Out [49]: dtype('int64')

In [ ]:
```
df['avg_session_duration_minutes'].dtype
```

Out [50]: dtype('int64')

In [ ]:
```
df['DataTime'] = pd.to_datetime(df['lms_logins_past_month'])
```

In [ ]:
```
df.head(2)
```

Out [52]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |

In [ ]:
```
df['DataTime'].dt.month
```

Out [53]:

| | DataTime |
|---|---|
| **0** | 1 |
| **1** | 1 |
| **2** | 1 |
| **3** | 1 |
| **4** | 1 |
| **...** | ... |
| **1540** | 1 |
| **1541** | 1 |
| **1542** | 1 |
| **1543** | 1 |
| **1544** | 1 |

1545 rows × 1 columns

**dtype:** int32

In [ ]:
```
df['GPA'].unique()
```

Out [54]:
```
array([2.42, 3.73, 2.8 , 2.59, 2.3 , 2.53, 2.04, 3.17, 3.68, 3.1 , 2.44,
       2.07, 2.43, 2.83, 3.52, 2.38, 3.44, 2.91, 3.64, 3.11, 3.27, 3.05,
       2.93, 2.47, 2.61, 2.58, 3.13, 2.24, 3.83, 2.32, 3.76, 3.65, 2.54,
       2.21, 3.57, 3.43, 3.19, 3.79, 2.68, 3.88, 2.17, 3.86, 2.26, 3.14,
       3.81, 3.89, 2.84, 2.75, 3.87, 3.35, 2.23, 2.08, 3.69, 2.28, 2.51,
       3.75, 2.37, 2.56, 2.95, 4.  , 3.  , 2.77, 2.22, 3.54, 2.01, 3.72,
       2.74, 3.74, 3.46, 3.3 , 3.59, 3.51, 2.79, 2.25, 3.47, 3.95, 2.1 ,
       3.36, 2.94, 3.16, 2.09, 3.33, 3.56, 3.09, 2.2 , 3.38, 2.48, 3.82,
       3.42, 3.9 , 2.06, 2.66, 3.45, 3.71, 3.15, 2.15, 2.36, 2.87, 3.48,
       2.45, 3.28, 3.66, 3.21, 2.41, 3.8 , 2.12, 3.34, 3.98, 3.06, 3.29,
       3.12, 3.55, 3.24, 2.63, 2.55, 3.96, 2.33, 3.62, 2.67, 3.08, 2.99,
       2.46, 2.6 , 2.71, 3.92, 2.92, 2.96, 2.27, 3.99, 2.78, 3.32, 2.81,
       3.91, 3.84, 2.35, 2.03, 3.4 , 3.93, 2.62, 3.53, 3.67, 3.61, 3.03,
       2.49, 2.  , 2.31, 2.18, 2.89, 3.5 , 2.86, 2.65, 2.29, 3.78, 2.73,
       2.19, 3.18, 2.72, 2.4 , 2.02, 3.77, 3.22, 2.76, 3.04, 3.26, 2.34,
       3.94, 2.9 , 3.7 , 2.69, 2.97, 3.01, 3.6 , 3.23, 3.41, 3.07, 3.25,
       3.63, 2.14, 2.5 , 2.82, 3.2 , 3.85, 3.39, 3.49, 2.88, 3.31, 2.85,
       2.57, 2.64, 2.16, 2.7 , 2.98, 3.58, 3.97, 3.02, 2.52, 2.13, 2.11,
       3.37, 2.05, 2.39])
```

In [ ]:
```
df['Year'] = df['DataTime'].dt.year
```

In [ ]:
```
df.sort_values('course_load')
```

Out [56]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **504** | S505 | 24 | Male | Business | 2.90 | 3 | 98.7 | 0.77 | Leave | 33 | 48 |
| **457** | S458 | 25 | Male | Arts | 3.36 | 3 | 92.4 | 0.97 | Active | 6 | 46 |
| **459** | S460 | 20 | Male | Computer Science | 2.41 | 3 | 63.7 | 0.74 | Graduated | 15 | 66 |
| **469** | S470 | 21 | Female | Computer Science | 2.65 | 3 | 71.5 | 0.60 | Graduated | 38 | 89 |
| **471** | S472 | 25 | Male | Computer Science | 3.83 | 3 | 81.3 | 0.77 | Leave | 3 | 16 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **25** | S026 | 23 | Female | Engineering | 2.61 | 6 | 89.7 | 0.73 | Active | 12 | 37 |
| **13** | S014 | 20 | Other | Computer Science | 2.83 | 6 | 65.9 | 0.61 | Leave | 3 | 74 |
| **1529** | S1530 | 21 | Female | Business | 2.25 | 6 | 94.3 | 0.93 | Active | 28 | 23 |
| **45** | S046 | 21 | Male | Computer Science | 2.93 | 6 | 89.8 | 0.99 | Leave | 7 | 40 |
| **11** | S012 | 24 | Female | Engineering | 2.07 | 6 | 75.5 | 0.81 | Active | 26 | 25 |

1545 rows × 17 columns

In [ ]:
```
import matplotlib.pyplot as plt
```

In [ ]:
```
df.groupby('student_id')['avg_course_grade'].sum()
```

Out [62]:

| student_id | avg_course_grade |
|---|---|
| **S001** | 67.2 |
| **S002** | 64.4 |
| **S003** | 95.3 |
| **S004** | 73.7 |
| **S005** | 87.4 |
| **...** | ... |
| **S995** | 86.8 |
| **S996** | 77.9 |
| **S997** | 79.5 |
| **S998** | 68.2 |
| **S999** | 85.1 |

1545 rows × 1 columns

**dtype:** float64

In [ ]:
```
df.head(2)
```

Out [65]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |

In [ ]:
```
newdata = df.loc[0:50]
```

In [ ]:
```
newdata.head()
```

Out [67]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |

In [ ]:
```
newdata2 = df.loc[100:150]
```

In [ ]:
```
newdata2.head()
```

Out [69]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ses |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **100** | S101 | 24 | Other | Computer Science | 2.79 | 3 | 93.6 | 0.75 | Leave | 24 | 34 |
| **101** | S102 | 19 | Female | Business | 3.86 | 4 | 82.0 | 0.94 | Leave | 22 | 59 |
| **102** | S103 | 19 | Female | Business | 2.25 | 5 | 66.1 | 0.82 | Graduated | 21 | 23 |
| **103** | S104 | 21 | Female | Arts | 3.47 | 3 | 89.1 | 0.96 | Graduated | 3 | 15 |
| **104** | S105 | 19 | Male | Business | 3.95 | 5 | 84.0 | 0.97 | Active | 33 | 85 |

In [ ]:
```
N = pd.concat([newdata,newdata2],ignore_index = True)
```

In [ ]:
```
N.tail()
```

Out [72]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ses |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **97** | S147 | 20 | Male | Computer Science | 2.45 | 6 | 64.9 | 0.63 | Leave | 27 | 25 |
| **98** | S148 | 20 | Other | Computer Science | 3.28 | 5 | 72.1 | 0.88 | Graduated | 32 | 30 |

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_ses |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **99** | S149 | 18 | Other | Business | 3.14 | 3 | 81.3 | 0.97 | Leave | 25 | 15 |
| **100** | S150 | 25 | Other | Business | 3.14 | 5 | 82.6 | 0.71 | Graduated | 28 | 50 |
| **101** | S151 | 20 | Female | Computer Science | 3.33 | 6 | 84.0 | 0.75 | Leave | 26 | 24 |

In [ ]:
```
N.shape
```

Out [73]: (102, 17)

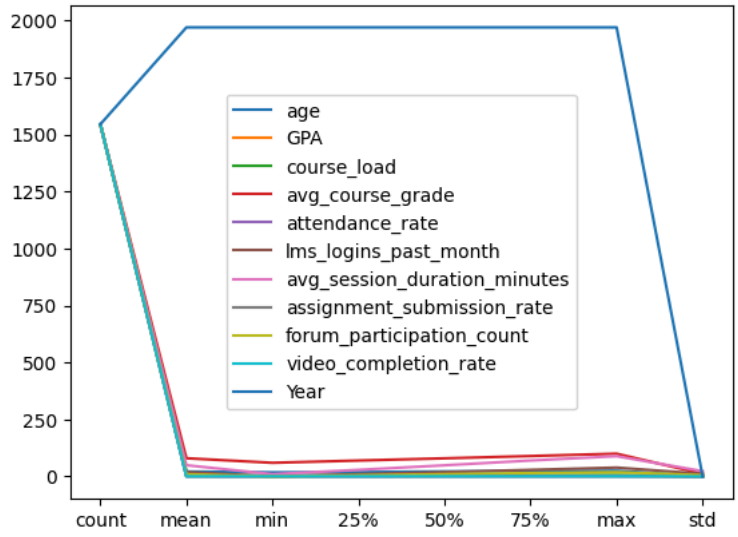
In [ ]:
```
N.loc[66:68]
```

Out [74]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **66** | S116 | 21 | Other | Business | 2.20 | 6 | 92.6 | 0.71 | Graduated | 16 | 81 |
| **67** | S117 | 18 | Male | Arts | 3.38 | 6 | 83.9 | 0.80 | Graduated | 0 | 23 |
| **68** | S118 | 23 | Male | Business | 2.48 | 5 | 75.7 | 0.63 | Active | 5 | 40 |

In [ ]:
```
df.describe().plot()
```

Out [79]: <Axes: >




In [ ]:
```
df.head()
```

Out [80]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | lms_logins_past_month | avg_sessi |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 32 | 33 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 29 | 53 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 34 | 69 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 22 | 18 |
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 9 | 11 |

In [ ]:
```
list = ['lms_logins_past_month','avg_session_duration_minutes','forum_participation_count','video_completion_r
```

In [ ]:
```
df.drop(list, axis = 1,inplace = True)
```

In [ ]:
```
df.head()
```

Out [84]:

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | assignment_submission_rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | S001 | 24 | Other | Computer Science | 2.42 | 5 | 67.2 | 0.71 | Graduated | 0.70 | 1970 00:0 |
| **1** | S002 | 21 | Male | Arts | 3.73 | 6 | 64.4 | 0.84 | Leave | 0.91 | 1970 00:0 |
| **2** | S003 | 22 | Male | Computer Science | 2.80 | 3 | 95.3 | 0.89 | Graduated | 0.58 | 1970 00:0 |
| **3** | S004 | 24 | Male | Arts | 2.59 | 4 | 73.7 | 0.98 | Graduated | 0.91 | 1970 00:0 |

| | student_id | age | gender | field | GPA | course_load | avg_course_grade | attendance_rate | enrollment_status | assignment_submission_rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **4** | S005 | 20 | Other | Computer Science | 2.30 | 4 | 87.4 | 0.95 | Active | 0.77 | 1970 00:0 |

In [ ]:
```
df.describe().plot()
```

Out [86]: <Axes: >