

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

```
In [2]: df=pd.read_csv('Student_performance_data _.csv')
df
```

Out[2]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular
0	1001	17	1	0	2	19.833723	7	1	2	
1	1002	18	0	0	1	15.408756	0	0	1	
2	1003	15	0	2	3	4.210570	26	0	2	
3	1004	17	1	0	3	10.028829	14	0	3	
4	1005	17	1	0	2	4.672495	17	1	3	
...
2387	3388	18	1	0	3	10.680555	2	0	4	
2388	3389	17	0	0	1	7.583217	4	1	4	
2389	3390	16	1	0	2	6.805500	20	0	2	
2390	3391	16	1	1	0	12.416653	17	0	2	
2391	3392	16	1	0	2	17.819907	13	0	2	

2392 rows x 15 columns

```
In [3]: df.shape
```

Out[3]: (2392, 15)

```
In [4]: df.head
```

Out[4]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular	Sports	Music	Volunteering	GPA	GradeClass
0	1001	17	1	0	2	19.833723	7	1	2		0	0	0	2.929196	2.0
1	1002	18	0	0	1	15.408756	0	0	1		0	0	0	3.042915	1.0
2	1003	15	0	2	3	4.210570	26	0	2		0	0	0	0.112602	4.0
3	1004	17	1	0	3	10.028829	14	0	3		1	0	0	2.054218	3.0
4	1005	17	1	0	2	4.672495	17	1	3		0	0	0	1.288061	4.0
...
2387	3388	18	1	0	3	10.680555	2	0	4		1	0	0	3.455509	0.0
2388	3389	17	0	0	1	7.583217	4	1	4		0	1	0	3.279150	4.0
2389	3390	16	1	0	2	6.805500	20	0	2		0	0	0	1.142333	2.0
2390	3391	16	1	1	0	12.416653	17	0	2		0	1	1	1.803297	1.0
2391	3392	16	1	0	2	17.819907	13	0	2		0	0	0	2.140014	1.0

[2392 rows x 15 columns]>

```
In [5]: df.tail()
```

Out[5]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular
2387	3388	18	1	0	3	10.680555	2	0	4	
2388	3389	17	0	0	1	7.583217	4	1	4	
2389	3390	16	1	0	2	6.805500	20	0	2	
2390	3391	16	1	1	0	12.416653	17	0	2	
2391	3392	16	1	0	2	17.819907	13	0	2	

In [6]:

df.count()

Out[6]:

StudentID2392
Age2392
Gender2392
Ethnicity2392
ParentalEducation2392
StudyTimeWeekly2392
Absences2392
Tutoring2392
ParentalSupport2392
Extracurricular2392
Sports2392
Music2392
Volunteering2392
GPA2392
GradeClass2392
dtype: int64

In [7]:

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2392 entries, 0 to 2391
Data columns (total 15 columns):
Column Non-Null Count Dtype

0 StudentID 2392 non-null int64
1 Age 2392 non-null int64
2 Gender 2392 non-null int64
3 Ethnicity 2392 non-null int64
4 ParentalEducation 2392 non-null int64
5 StudyTimeWeekly 2392 non-null float64
6 Absences 2392 non-null int64
7 Tutoring 2392 non-null int64
8 ParentalSupport 2392 non-null int64
9 Extracurricular 2392 non-null int64
10 Sports 2392 non-null int64
11 Music 2392 non-null int64
12 Volunteering 2392 non-null int64
13 GPA 2392 non-null float64
14 GradeClass 2392 non-null float64
dtypes: float64(3), int64(12)
memory usage: 280.4 KB

In [8]:

df.isnull()

Out[8]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...
2387	False	False	False	False	False	False	False	False	False	False
2388	False	False	False	False	False	False	False	False	False	False
2389	False	False	False	False	False	False	False	False	False	False
2390	False	False	False	False	False	False	False	False	False	False
2391	False	False	False	False	False	False	False	False	False	False

2392 rows × 15 columns

In [9]:

df.isnull().sum()

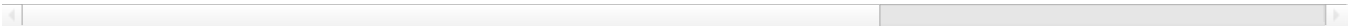
Out[9]: StudentID 0
Age 0
Gender 0
Ethnicity 0
ParentalEducation 0
StudyTimeWeekly 0
Absences 0
Tutoring 0
ParentalSupport 0
Extracurricular 0
Sports 0
Music 0
Volunteering 0
GPA 0
GradeClass 0
dtype: int64

In [10]: df.dropna()

Out[10]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular
0	1001	17	1	0	2	19.833723	7	1	2	
1	1002	18	0	0	1	15.408756	0	0	1	
2	1003	15	0	2	3	4.210570	26	0	2	
3	1004	17	1	0	3	10.028829	14	0	3	
4	1005	17	1	0	2	4.672495	17	1	3	
...
2387	3388	18	1	0	3	10.680555	2	0	4	
2388	3389	17	0	0	1	7.583217	4	1	4	
2389	3390	16	1	0	2	6.805500	20	0	2	
2390	3391	16	1	1	0	12.416653	17	0	2	
2391	3392	16	1	0	2	17.819907	13	0	2	

2392 rows × 15 columns

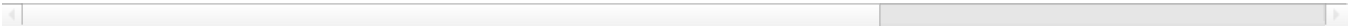


In [11]: df.fillna(0)

Out[11]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Extracurricular
0	1001	17	1	0	2	19.833723	7	1	2	
1	1002	18	0	0	1	15.408756	0	0	1	
2	1003	15	0	2	3	4.210570	26	0	2	
3	1004	17	1	0	3	10.028829	14	0	3	
4	1005	17	1	0	2	4.672495	17	1	3	
...
2387	3388	18	1	0	3	10.680555	2	0	4	
2388	3389	17	0	0	1	7.583217	4	1	4	
2389	3390	16	1	0	2	6.805500	20	0	2	
2390	3391	16	1	1	0	12.416653	17	0	2	
2391	3392	16	1	0	2	17.819907	13	0	2	

2392 rows × 15 columns



In [12]: df.describe()

Out[12]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences	Tutoring	ParentalInvolvement
count	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000
mean	2196.500000	16.468645	0.510870	0.877508	1.746237	9.771992	14.541388	0.301421	0.000000
std	690.655244	1.123798	0.499986	1.028476	1.000411	5.652774	8.467417	0.458971	0.000000
min	1001.000000	15.000000	0.000000	0.000000	0.000000	0.001057	0.000000	0.000000	0.000000
25%	1598.750000	15.000000	0.000000	0.000000	1.000000	5.043079	7.000000	0.000000	0.000000
50%	2196.500000	16.000000	1.000000	0.000000	2.000000	9.705363	15.000000	0.000000	0.000000
75%	2794.250000	17.000000	1.000000	2.000000	2.000000	14.408410	22.000000	1.000000	0.000000
max	3392.000000	18.000000	1.000000	3.000000	4.000000	19.978094	29.000000	1.000000	0.000000

In [13]:

df['Extracurricular'].mean()

Out[13]:

0.3833612040133779

In [14]:

df['Music'].mode()

Out[14]:

0 0
Name: Music, dtype: int64

In [15]:

df['Sports'].median()

Out[15]:

0.0

In []: