In [27]:
```
import pandas as pd
import numpy as np
```

In [28]:
```
df=pd.read_csv('Student_performance_data _ - Copy.csv')
```

In [29]:
```
df.head()
```

Out[29]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absences
0	1001	17	1	0	2	19.833723	7
1	1002	18	0	0	1	15.408756	0
2	1003	15	0	2	3	4.210570	26
3	1004	17	1	0	3	10.028829	14
4	1005	17	1	0	2	4.672495	17

In [30]:
```
df.tail()
```

Out[30]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTimeWeekly	Absenc
2387	3388	18	1	0	3	10.680555	
2388	3389	17	0	0	1	7.583217	
2389	3390	16	1	0	2	6.805500	
2390	3391	16	1	1	0	12.416653	
2391	3392	16	1	0	2	17.819907	

In [31]:
```
df.describe()
```

Out[31]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	StudyTim
count	2392.000000	2392.000000	2392.000000	2392.000000	2392.000000	239
mean	2196.500000	16.468645	0.510870	0.877508	1.746237	
std	690.655244	1.123798	0.499986	1.028476	1.000411	
min	1001.000000	15.000000	0.000000	0.000000	0.000000	
25%	1598.750000	15.000000	0.000000	0.000000	1.000000	
50%	2196.500000	16.000000	1.000000	0.000000	2.000000	
75%	2794.250000	17.000000	1.000000	2.000000	2.000000	
max	3392.000000	18.000000	1.000000	3.000000	4.000000	

In [32]:
```
df.isna().sum()
```

Out[32]:
```
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
StudyEfficiency       0
ActiveParticipation   0
GradeClass            0
dtype: int64
```

In [33]:
```
df.corr()
```

Out[33]:

	StudentID	Age	Gender	Ethnicity	ParentalEducation	Study
StudentID	1.000000	-0.042255	-0.014625	-0.012990	-0.002307	
Age	-0.042255	1.000000	0.044895	-0.028473	0.025099	
Gender	-0.014625	0.044895	1.000000	0.016010	0.006771	
Ethnicity	-0.012990	-0.028473	0.016010	1.000000	0.033595	
ParentalEducation	-0.002307	0.025099	0.006771	0.033595	1.000000	
StudyTimeWeekly	0.026976	-0.006800	0.011469	0.007184	-0.011051	
Absences	0.014841	-0.011511	0.021479	-0.025712	0.036518	
Tutoring	-0.007834	-0.012076	-0.031597	-0.017440	-0.017340	
ParentalSupport	0.003016	0.033197	0.008065	0.020922	-0.017463	
Extracurricular	-0.003611	-0.025061	-0.005964	-0.008927	0.007479	
Sports	-0.020703	-0.046320	-0.008897	-0.004484	0.002029	
Music	-0.005468	-0.003492	0.007109	-0.014627	0.039439	
Volunteering	0.008011	0.013074	-0.000200	0.013468	0.011960	
GPA	-0.002697	0.000275	-0.013360	0.027760	-0.035854	
StudyEfficiency	0.022726	-0.029107	-0.018161	-0.015992	-0.015603	
ActiveParticipation	-0.012400	-0.035449	-0.004987	-0.008613	0.028967	
GradeClass	-0.098500	-0.006250	0.022998	-0.023326	0.041031	

In [34]:
```
correlation=df.corr()['GradeClass']
correlation
```

Out[34]:
```
StudentID             -0.098500
Age                   -0.006250
Gender                 0.022998
Ethnicity             -0.023326
ParentalEducation      0.041031
StudyTimeWeekly       -0.134131
Absences               0.728633
Tutoring              -0.111695
ParentalSupport       -0.136823
Extracurricular       -0.069733
Sports                -0.026654
Music                 -0.036065
Volunteering           0.013156
GPA                   -0.782835
StudyEfficiency       -0.016915
ActiveParticipation   -0.065587
GradeClass             1.000000
Name: GradeClass, dtype: float64
```

In [35]:
```
x=df.drop(['StudentID','Age','Ethnicity','ParentalEducation','StudyEfficiency','
```

In [ ]:
```

```

In [36]:
```
x
```

Out[36]:

	StudyTimeWeekly	Absences	Tutoring	ParentalSupport	Volunteering	GPA
0	19.833723	7	1	2	0	2.929196
1	15.408756	0	0	1	0	3.042915
2	4.210570	26	0	2	0	0.112602
3	10.028829	14	0	3	0	2.054218
4	4.672495	17	1	3	0	1.288061
...	...	...	...	...	...	...
2387	10.680555	2	0	4	0	3.455509
2388	7.583217	4	1	4	0	3.279150
2389	6.805500	20	0	2	1	1.142333
2390	12.416653	17	0	2	0	1.803297
2391	17.819907	13	0	2	1	2.140014

2392 rows × 7 columns

In [37]:
```
y=df['GradeClass']
y
```

Out[37]:
```
0       2
1       1
2       4
3       3
4       4
       ..
2387    0
2388    4
2389    2
2390    1
2391    1
Name: GradeClass, Length: 2392, dtype: int64
```

In [38]:
```
x.size
```

Out[38]:
```
16744
```

In [39]:
```
y.size
```

Out[39]:
```
2392
```

In [40]:
```
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42
```

In [41]:
```
x_train.shape
```

Out[41]:
```
(1913, 7)
```

In [42]:
```
x_test.shape
```

Out[42]:
```
(479, 7)
```

In [43]:
```
y_train.shape
```

Out[43]:
```
(1913,)
```

In [44]:
```
y_test.shape
```

Out[44]:
```
(479,)
```

In [45]:
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

RFC = RandomForestClassifier()
```

In [46]:
```
RFC.fit(x_train,y_train)
```

Out[46]:
```
▾ RandomForestClassifier
RandomForestClassifier()
```

In [47]:
```
test_predict=RFC.predict(x_test)

Accuracy=accuracy_score(y_test,test_predict)
```

In [48]:
```
Accuracy
```

Out[48]:
```
1.0
```

In [ ]:

In [ ]:

In [ ]:

In [49]:
```
import pickle
```

In [50]:
```
with open('RFC_std_1.pkl','wb') as file:
    pickle.dump(RFC,file)
```