```
# 1.) Importing the neccessary files.
import pandas as pd
```

```
# 2.) Loading the dataset.
df = pd.read_csv("/content/Student_performance_data _.csv")
```

```
# 3.) Check for the missing values and removing the duplicates.
print(" 🔹 Missing values before cleaning:")
print(df.isnull().sum())

duplicates = df.duplicated().sum()
print("\n 🔹 Duplicate rows:", duplicates)

df = df.drop_duplicates()
```

```
 🔹 Missing values before cleaning:
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

 🔹 Duplicate rows: 0
```

```
# 4.) Converting the categoricals to numericals.
for col in df.columns:
    if df[col].dtype == 'object':
        df[col] = df[col].astype('category').cat.codes

print("\n 🔹 After converting categorical columns:")
print(df.head())
```

```

 🔹 After converting categorical columns:
   StudentID  Age  Gender  Ethnicity  ParentalEducation  StudyTimeWeekly  \
0       1001   17       1          0                  2        19.833723
1       1002   18       0          0                  1        15.408756
2       1003   15       0          2                  3         4.210570
3       1004   17       1          0                  3        10.028829
4       1005   17       1          0                  2         4.672495

   Absences  Tutoring  ParentalSupport  Extracurricular  Sports  Music  \
0         7         1                2                0       0      1
1         0         0                1                0       0      0
2        26         0                2                0       0      0
3        14         0                3                1       0      0
4        17         1                3                0       0      0

   Volunteering       GPA  GradeClass
0             0  2.929196         2.0
1             0  3.042915         1.0
2             0  0.112602         4.0
3             0  2.054218         3.0
4             0  1.288061         4.0
```

```
# 5.) Feature & target selection.
X = df.iloc[:, :-1]
y = df.iloc[:, -1]

print("\n 🔹 Features shape:", X.shape)
print(" 🔹 Target shape:", y.shape)
```

```

 🔹 Features shape: (2392, 14)
 🔹 Target shape: (2392,)
```

```
# 6.) Saving the cleaned dataset.
df.to_csv("cleaned_student_data.csv", index=False)
print("\n✅ Cleaned dataset saved")
```

```
✅ Cleaned dataset saved
```