

Dataset Description

This dataset captures academic, demographic, and behavioral information from students at two secondary schools. The data was collected via school reports and structured questionnaires and is intended for understanding patterns in student life and academic performance.

Each row represents a unique student. Below is a brief overview of the columns:

Demographic Information

- **school**: School attended by the student (GP or MS)
- **sex**: Gender of the student (F for female, M for male)
- **address**: Type of home address (U for urban, R for rural)
- **famsize**: Size of the student's family (LE3 for ≤ 3 members, GT3 for > 3)
- **Pstatus**: Parental cohabitation status (T for together, A for apart)
- **guardian**: Primary guardian (mother, father, or other)

Parental Background

- **Medu**: Mother's level of education (0 to 4)
 - **Fedu**: Father's level of education (0 to 4)
 - **Mjob**: Mother's job (teacher, health, services, at_home, or other)
 - **Fjob**: Father's job (teacher, health, services, at_home, or other)
-

School and Study Related

- **reason**: Reason for choosing the school (**home**, **reputation**, **course**, or **other**)
 - **traveltime**: Travel time from home to school (1: <15 min to 4: >1 hour)
 - **failures**: Number of past class failures (1 to 4)
 - **schoolsup**: Receiving extra educational support (**yes** or **no**)
 - **famsup**: Receiving family educational support (**yes** or **no**)
 - **paid**: Attending extra paid classes (**yes** or **no**)
 - **activities**: Participation in extracurricular activities (**yes** or **no**)
 - **nursery**: Attended nursery school (**yes** or **no**)
 - **higher**: Aspiration to pursue higher education (**yes** or **no**)
 - **internet**: Internet access at home (**yes** or **no**)
-

Lifestyle & Behavior

- **famrel**: Quality of family relationships (1: very bad to 5: excellent)
- **freetime**: Free time after school (1: very low to 5: very high)
- **goout**: Frequency of going out with friends (1 to 5)
- **Dalc**: Weekday alcohol consumption (1 to 5)
- **health**: Current health status (1 to 5)
- **absences**: Number of school absences (0 to 93)

Academic Performance

- **G1**: Grade in the first term (0 to 20)
 - **G2**: Grade in the second term (0 to 20)
 - **G3**: Final grade for the year (0 to 20)
-

Feature Columns

- **Feature_1, Feature_2, Feature_3**: Anonymized columns — their original meanings are intentionally hidden and should be uncovered as part of your task. Use EDA and logic to infer what they might represent.
-

Target Column

- **romantic**: Indicates whether the student is in a romantic relationship (**yes** or **no**)