

## DESCRIPCIÓN DEL DATASET

### 1. Origen de los Datos

El dataset utilizado en este análisis fue simulado con fines educativos y representa información sobre el rendimiento académico de estudiantes universitarios. Incluye variables sociodemográficas, hábitos de estudio, participación en clase y otros factores que pueden influir en el desempeño académico.

Este conjunto de datos cuenta con 18 columnas que contienen diferentes tipos de datos como cadenas de texto (strings), números enteros y decimales (int, float), y valores booleanos (true/false). Esta diversidad permite explorar relaciones entre variables cualitativas y cuantitativas, brindando una base para el análisis exploratorio dentro del módulo *Análisis de Datos*.

### 2. Descripción de las columnas

Columna	Tipo de Dato	Descripción
student_ulid	Categorico (str)	ID único del estudiante en formato ULID
first_name	Categorico (str)	Nombre del estudiante
last_name	Categorico (str)	Apellido del estudiante
student_id	Categorico (str)	Número de identificación del estudiante
gender	Categorico (str)	Género declarado (Male, Female, Bigender, etc.)
age	Numérico (int)	Edad del estudiante
email	Categorico (str)	Correo electrónico institucional
address	Categorico (str)	Dirección del estudiante
phone	Numérico (int)	Teléfono de contacto
major	Categorico (str)	Carrera universitaria o especialización
gpa	Numérico (float)	Promedio académico (escala 0.0 a 4.0)
attendance_rate	Numérico (float)	Porcentaje de asistencia a clases (escala 0% a 100%)
mental_health_status	Categorico (string)	Estado de salud mental (Good, Average, Bad, etc.)
study_hours_per_week	Numérico (float)	Horas de estudio semanales (algunos faltantes)
extracurricular_participatio	Booleano	Participa en actividades extracurriculares (true/false)
scholarship	Booleano	Posee una beca (true/false)
internet_access	Booleano	Tiene acceso a internet en su hogar (true/false)
sleep_hours_per_night	Numérico (float)	Promedio de horas de sueño por noche (algunos valores vacíos)

### 3. Estado de los Datos

Los datos están en estado RAW, es decir, sin tratamiento previo. Esto implica:

- No se ha realizado limpieza de datos (puede haber ausentes o inconsistencias).
- Las columnas están en su forma original (nombres, formatos).

