

Evidencia nº 2 - Modulo Análisis de datos

Proyecto: EduDatálisis - Análisis del Rendimiento Académico de Estudiantes
Grupo nº 3

Descripción del Dataset

1. Origen de los Datos

El dataset utilizado para este análisis fue generado y simulado con fines educativos, y representa información sobre el rendimiento académico de estudiantes universitarios. Incluye variables socio-demográficas, hábitos de estudio y niveles de participación en clase. Su propósito principal es servir como base para el análisis exploratorio de datos en el marco del módulo "Analista de Datos".

Para la creación del conjunto de datos se utilizó la herramienta Mockaroo: Fabricate, la cual permite generar bases de datos sintéticas con integridad relacional completa. Esto asegura coherencia entre las variables simuladas y permite emular estructuras de datos reales utilizadas en entornos educativos.

El dataset consta de 18 columnas con diversos tipos de datos: cadenas de texto (strings), números enteros y decimales (int, float), y valores booleanos (true/false). Esta variedad posibilita el análisis de relaciones entre variables cualitativas y cuantitativas, en el contexto del rendimiento académico universitario.

2. Descripción de las Columnas

Columna	Tipo de Dato	Descripción
student_ulid	Categórico (str)	ID único del estudiante en formato ULID
first_name	Categórico (str)	Nombre del estudiante
last_name	Categórico (str)	Apellido del estudiante
student_id	Categórico (str)	Número de identificación del estudiante
gender	Categórico (str)	Género declarado (Male, Female, Bigender, etc.)
age	Numérico (int)	Edad del estudiante
email	Categórico (str)	Correo electrónico institucional
address	Categórico (str)	Dirección del estudiante
phone	Numérico (int)	Teléfono de contacto
major	Categórico (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	Categórico (string)	Estado de salud mental (Good, Average, Bad, etc.)
study_hours_per_week	Numérico (float)	Horas de estudio semanales (algunos faltantes)
extracurricular_participation	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).
- No se han estandarizado las categorías ni aplicado transformaciones numéricas.

