



HU_SM4_01: Estadística Descriptiva aplicada a un dataset con Python

Nombre de la HU:		Análisis de Rendimiento Estudiantil
Objetivo de la HU		Como director académico quiero obtener un análisis confiable sobre el rendimiento de los estudiantes, para identificar patrones de asistencia, esfuerzo y resultados finales, con el fin de mejorar las estrategias de acompañamiento y apoyo escolar.
TASK1	Descripción de la Tarea: Exploración inicial de la información	
	Criterios de Aceptación: CA_01: Cargar la base de datos de estudiantes. CA_02: Seleccionar columnas numéricas de interés (age, studytime, failures, absences, G1, G2, G3). CA_03: Filtrar y mostrar estudiantes mayores de 18 años. CA_04: Filtrar y mostrar estudiantes con más de 20 ausencias.	
	Story Points: 1	
TASK2	Descripción de la Tarea: Revisión de calidad de los datos	
	Criterios de Aceptación: CA_05: Identificar y eliminar registros repetidos. CA_06: Detectar valores extremos en (absences y G3) notas finales - (outliers) CA_07: Validar que las notas (G1, G2, G3) estén dentro del rango oficial 0–20.	
	Story Points: 1	
TASK3	Descripción de la Tarea: Indicadores básicos de rendimiento	
	Criterios de Aceptación: CA_07: Promedio, mínimo y máximo de la nota final (G3). CA_08: Cantidad de estudiantes por nivel de horas de estudio (studytime). CA_09: Promedio general de ausencias (absences).	
	Story Points: 1	
TASK4	Descripción de la Tarea: Indicadores complementarios	
	Criterios de Aceptación: CA_11: Porcentaje de estudiantes aprobados (nota final ≥ 10) y reprobados . Distribución de ausencias en niveles (bajo, medio, alto). Normalización de las calificaciones finales para análisis comparativo. CA_11: (Si quieres, añadir CA_13: Percentiles de ausencias).	



Story Points: 1

Cierre de la actividad:

1. Entregables:

Notebook con nombre HU_SM4_01_Apellido_Nombre.ipynb que debe incluir:

- Dataset limpio y depurado.
- Tabla con indicadores básicos.
- Tabla con indicadores complementarios.
- Código documentado (Uso de Markdown)
- Resultados de los KPIs solicitados:
 - Promedio de ausencias por grupo de edad.
 - Total de estudiantes por sexo.
 - Promedio de G3 en estudiantes con studytime alto vs bajo.
 - Distribución de G3: mínimo, máximo, media, mediana.

2. Requerimientos Funcionales (RF)

- **RF_01:** Permitir la carga y validación de la base de datos de estudiantes.
- **RF_02:** Seleccionar y mostrar únicamente las variables numéricas de interés.
- **RF_03:** Aplicar filtros de población (mayores de 18 años, más de 20 ausencias).
- **RF_04:** Identificar y eliminar registros duplicados.
- **RF_05:** Detectar y depurar valores atípicos en ausencias y calificaciones.
- **RF_06:** Validar que las calificaciones estén en el rango oficial (0–20).
- **RF_07:** Calcular indicadores básicos (promedio, mínimo, máximo, conteos).
- **RF_08:** Calcular indicadores complementarios (aprobados/reprobados, percentiles, normalización).
- **RF_09:** Generar reporte final con dataset limpio e indicadores solicitados.

3. Requerimientos No Funcionales (RNF)

- **RNF_01:** El análisis debe ejecutarse en segundos para bases pequeñas y escalar razonablemente a bases medianas.
- **RNF_02:** Los resultados deben ser reproducibles y consistentes.
- **RNF_03:** Los indicadores generados deben presentarse en un formato claro y fácil de interpretar.
- **RNF_04:** La solución debe ser sencilla de ejecutar, sin configuraciones técnicas complejas para el usuario final.