

▶ TALENTO TECH



TIC



Universidad
Tecnológica
de Bolívar

www.utb.edu.co/talento-tech

www.utb.edu.co/talento-tech

Data Wrangling end EDA with PANDAS

Ejecutor técnico: Jorge Luis Villalba Acevedo

Actividad: Exploración y Limpieza de Datos de Exportaciones Agrícolas en Colombia

Estimado **CAMPISTAS**,

Contexto: Desafíos en la Educación en Colombia

En Colombia, la educación enfrenta retos significativos, como la falta de acceso y la deserción escolar. A menudo, los estudiantes no logran alcanzar un rendimiento académico óptimo debido a factores como la asistencia irregular y la desigualdad en las oportunidades educativas. Este conjunto de datos simulado representa a estudiantes colombianos, incluyendo información sobre su rendimiento, asistencia y características demográficas. A través de este análisis, se busca resaltar cómo estas variables pueden afectar el desempeño educativo.

Descripción del Conjunto de Datos

- **student_id**: Identificación única para cada estudiante (algunos registros pueden estar duplicados).
- **age**: Edad del estudiante, variando entre 6 y 17 años.
- **score**: Calificación promedio en exámenes, con algunas calificaciones faltantes.
- **attendance**: Porcentaje de asistencia a clases, también con valores faltantes.
- **gender**: Género del estudiante, "Masculino" o "Femenino".
- **grade**: Grado escolar, ya sea "Primaria" o "Secundaria".
- **region**: Región de Colombia a la que pertenece el estudiante, incluyendo "Caribe", "Pacífico", "Andino" y "Amazonía".

Preguntas sobre Manejo de Datos Faltantes

1. ¿Cuántos valores faltantes hay en todo el DataFrame?
2. ¿Qué porcentaje de valores faltantes hay en la columna 'score'?
3. Rellena los valores faltantes de 'score' con la media.
4. Elimina las filas con valores faltantes en 'attendance'.
5. ¿Cuántos valores faltantes hay en cada columna?

Preguntas sobre Registros Duplicados

6. ¿Cuántos registros duplicados hay en el DataFrame?
7. Elimina los registros duplicados.
8. Muestra los registros duplicados.
9. ¿Cuántos estudiantes únicos hay en el DataFrame después de eliminar duplicados?

Preguntas sobre Transformación de Datos

10. Transforma la columna 'attendance' a un rango de 0 a 1.
11. Crea una nueva columna que indique si el estudiante aprobó ($\text{score} \geq 60$).
12. Crea un DataFrame que contenga solo estudiantes de secundaria.
13. Convierte la columna 'grade' a un tipo categórico.
14. ¿Cuál es la calificación promedio por grado?



Universidad
Tecnológica
de Bolívar

www.utb.edu.co/talento-tech