



**PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL**  
**TALLER DE CIENCIA DE DATOS**  
(Procesamiento de Datos)

**IDENTIFICACIÓN DEL TALLER**

- **Denominación del Programa de Formación:** Tecnólogo en Análisis y Desarrollo de Software
- **Código del Programa de Formación:** 228118
- **Nombre del Proyecto:** 2417781 - Desarrollo de software como recurso para la gestión de información de las organizaciones en el contexto de la innovación, transformación digital y en el marco de políticas y requisitos de la empresa.
- **Fase del Proyecto:** Evaluación
- **Actividad de Proyecto:** Desarrollar las tareas de configuración y puesta en marcha del software
- **Competencia:** Construcción del software
- **Resultados de Aprendizaje Alcanzar:** Codificar la solución que cumpla con el diseño establecido.
- **Duración del Taller:** 04 horas

**2. DESCRIPCION DEL TALLER:**

**2.1 Contexto:**

Este es un conjunto de datos educativos que se recopila del sistema de gestión de aprendizaje (LMS) llamado Territorio. Territorio es un LMS multiagente, que ha sido diseñado para facilitar el aprendizaje mediante el uso de tecnología de punta. Dicho sistema proporciona a los usuarios un acceso síncrono a los recursos educativos desde cualquier dispositivo con conexión a Internet.

Los datos se recopilan mediante una herramienta de seguimiento de la actividad del aprendiz, que se denomina API de experiencia (xAPI). La xAPI es un componente de la arquitectura de capacitación y aprendizaje (TLA) que permite monitorear el progreso del aprendizaje y las acciones del aprendiz, como leer un artículo o ver un video de capacitación. La API de experiencia ayuda a los proveedores de actividades de aprendizaje a determinar el aprendiz, la actividad y los objetos que describen una experiencia de aprendizaje.

El conjunto de datos consta de 40 registros de aprendices y 16 características. Las características se clasifican en dos categorías principales: La primera relacionada con las características socio-demográficas y la segunda relacionada con las características académicas.



## 2.2 Información de las etiquetas:

- **Id:** identificador único de cada aprendiz.
- **Nombre:** nombres y apellidos de cada aprendiz.
- **Sexo:** genero de cada aprendiz (masculino y femenino).
- **FechaNacimiento:** día, mes, año de nacimiento de cada aprendiz.
- **Estrato:** clasificación socioeconómica de cada aprendiz.
- **Trabaja:** Si el aprendiz labora actualmente (Si - No).
- **Enfermedad:** Si el aprendiz presenta alguna enfermedad de atención permanente (Si - No).
- **Ubicación:** ubicación geográfica del aprendiz.
- **Tipo de Colegio:** procedencia académica del aprendiz.
- **Uso\_LMS:** actividad dentro del ambiente LMS.
- **Asistencia:** porcentaje de asistencia en el ambiente LMS.
- **Entregas\_Completas:** porcentaje de entrega de las actividades del LMS.
- **Sistemas:** calificación final del área de conocimiento.
- **Matemáticas:** calificación final del área de conocimiento.
- **Inglés:** calificación final del área de conocimiento.
- **Comunicación:** calificación final del área de conocimiento.

## 2.3 Procesamiento de Datos:

1. Carga e Integración de los datos
2. Descripción estadística inicial de los datos
3. Eliminar o transformar variables irrelevantes, redundantes y nulos
4. Limpieza de datos
5. Descripción estadística final de los datos

## 2.4 Requerimientos

Para tener un mejor entendimiento del negocio y de los datos; esta información puede ser utilizada para realizar un informe estadístico que sería de gran interés para la empresa en los siguientes desafíos:

- Aprendices aprobados y no aprobados según el uso del LMS
- Características de los aprendices que usan el LMS y aprueban el programa.
- Características de los aprendices que usan el LMS y no aprueban el programa.
- Patrones de comportamiento de los aprendices en el ambiente LMS.

Esta información sería útil para el SENA Regional Antioquia para tomar decisiones sobre su negocio, tales como:



- Definir su nicho de mercado.
- Desarrollar estrategias de marketing.
- Ofrecer servicios personalizados a sus clientes.

### 2.5 Restricciones

- Los aprendices aprueban el programa de formación con un promedio ponderado superior a 3,5 entre las áreas de conocimiento.

### 2.6 Dataset

- <https://github.com/majash29/CienciaDatosSENAV02/tree/b0d2203d4888142dd2b3896126dd6b2728b78d55/02-ExploracionDatos/TallerProcesamientoDatos/Dataset>

## 3. ENTREGABLES DEL TALLER

- En el LMS se debe entregar la URL del repositorio de GitHub con los siguientes documentos:
  - Informe final de resultados (portada, introducción, desarrollo metodológico y conclusiones)
  - Notebook de Python con los scripts utilizados.

## 4. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
<b>Autor (es)</b>	Alvaro Pérez Niño	Instructor	ADSO	17 agosto de 2023