



**PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL**  
**TALLER DE CIENCIA DE DATOS**  
(Limpieza 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:** 03 horas

**2. DESCRIPCION DEL TALLER:**

**2.1 Contexto:**

El crecimiento de los supermercados en la mayoría de las ciudades pobladas es cada vez mayor y la competencia en el mercado también es alta. El conjunto de datos es una de las ventas históricas de la empresa de supermercados que ha registrado en 3 sucursales diferentes durante 3 meses de datos. Los métodos de análisis de datos predictivos son fáciles de aplicar con este conjunto de datos.

**2.2 Información de atributo:**

- **Factura:** número de identificación de factura de comprobante de venta generado por computadora
- **Cliente:** nombres y apellidos del cliente
- **Ubicación:** Ubicación de los supercentros
- **Genero:** tipo de género del cliente
- **Producto:** nombre de la categoría de producto que adquiere el cliente en la compra
- **TipoVenta:** opción de venta utilizado por el cliente para la compra (hay 3 métodos disponibles: física, catálogo y en línea)
- **Cantidad:** número de productos comprados por cliente
- **ValorUnitario:** precio de cada producto en \$- PESOS COP



### 2.3 Proceso de Preparación 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
6. Transformación de tipo de datos según el método

### 3. ENTREGABLES DEL TALLER

- Lectura y carga del dataset contenido en la siguiente URL:  
[https://github.com/majash29/CienciaDatosSENAV02/raw/main/02-ExploracionDatos/TallerLimpiezaDatos/04\\_TallerLimpiezaDatos.xlsx](https://github.com/majash29/CienciaDatosSENAV02/raw/main/02-ExploracionDatos/TallerLimpiezaDatos/04_TallerLimpiezaDatos.xlsx)
- Implementar los scripts necesarios que cumplan con el proceso de Preparación de los Datos.
- Generar una nueva columna calculada denominada SUBTOTAL correspondiente a multiplicar la CANTIDAD por VALORUNITARIO
- Realizar una grafica que permita mostrar la cantidad de productos vendidos por ubicación
- Realizar una grafica que permita mostrar total de ventas por genero de los clientes
- Generar dos consultas de agrupamiento de valor agregado a la informacion obtenida.
- Exportar el dataset final a un documento de hoja de calculo – Excel.

### 4. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
<b>Autor (es)</b>	Alvaro Pérez Niño	Instructor	ADSO	10 agosto de 2023