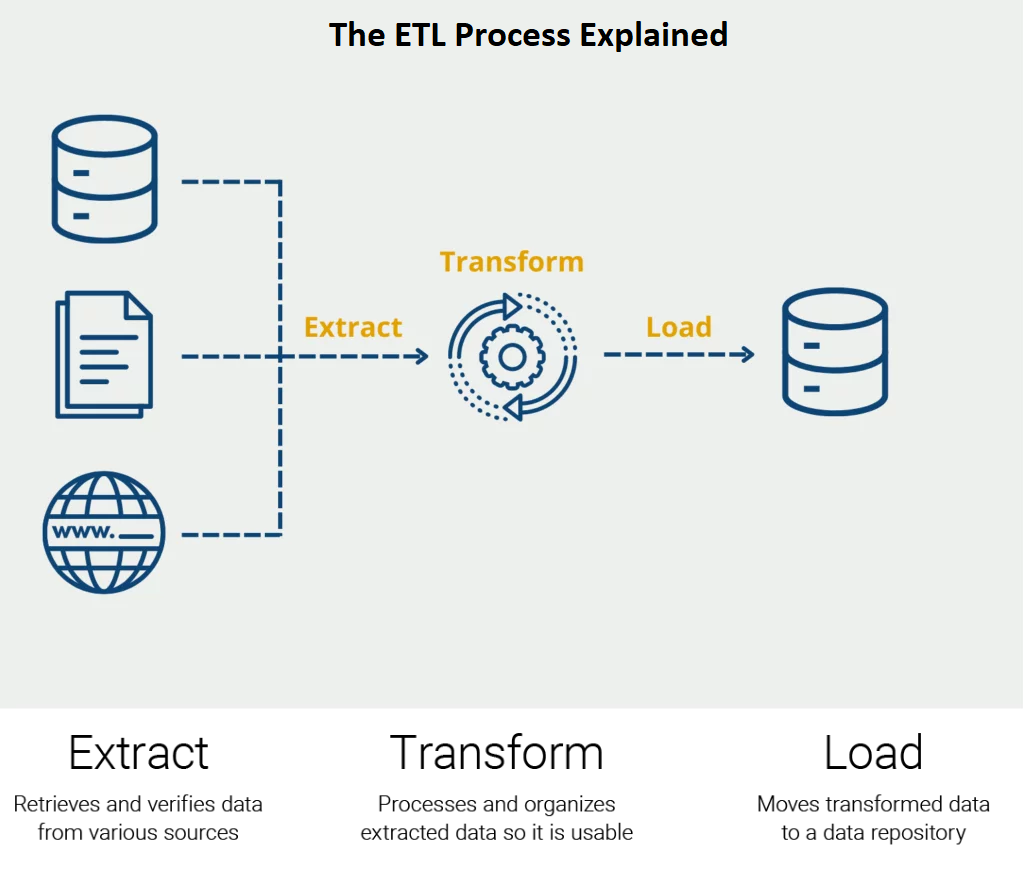
Laboratorio Base de datos

* Diseño de base de datos (Modelos y diccionario de datos)
* Construcción de bases de datos a partir del modelo de datos
* Limpieza de datos

****

1. **Fecha:** Semana No 5 (Lunes, 26 de mayo de 2025 al Viernes, 30 de Mayo de 2025)
2. **ETL (Extract, Transform, Load):**



1. **Entregables para cada base de datos:**

Al proporcionar esta documentación, se brinda a los usuarios una comprensión completa de la estructura y el funcionamiento de la base de datos, lo que facilita su uso y mantenimiento adecuados.

**1. Entregable: Modelo conceptual**

* Realizar el modelo entidad relación que permita representar la situación anteriormente planteada e identifique de manera clara las **entidades**, **relaciones** y **cardinalidades** del sistema (Imágen PNG o JPG del modelo).
* Puede usar **DIA** o cualquier otra herramienta adecuada.

**2. Entregable: Modelo físico**

* Un modelo de datos físico es un modelo específico de bases de datos que representa objetos de datos relacionales (por ejemplo, tablas, columnas, claves principales y claves externas) y sus relaciones. Un modelo de datos físico se puede utilizar para generar sentencias DDL.
* Crear el modelo relacional utilizando la herramienta MySQL Workbench. Entregar la imágen y el proyecto mysql workbench.

**3. Entregable: Script de la base de datos (Archivo .sql)**

Elabore el archivo ***.sql*** con las instrucciones para:

* Creación de los esquemas necesarios
* [Inserción de datos](https://docs.google.com/document/d/1BMh4zuS9SGC0iGhYfw9-Zq9MwEu_i74x1VUKvMDer_E/edit?usp=drive_link) utilizando el origen entregado por el instructor (*extract - transform*).
* Imprima la lista de tablas
* Imprima la descripción de cada campo de cada tabla de la base de datos
* Imprima los datos de cada tabla

**4. Entregable:** Archivos *.XLS* o .*CSV*

* Después realizar la limpieza y posible transformación de los datos para el proceso de carga de los mismos (*load data*).
* **Ejemplo**: [***Ejemplo Importar (CSV a MySQL)***](https://drive.google.com/drive/folders/1O4Tah8Q3gzHGAhiVY0BQC7imEyvWkn3h?usp=drive_link)

**5. Diccionario de datos**

Diccionario de datos con campos documentados (Generado).

**6. Entregable:** Tabla de Consultas

1. **Bases de datos**

* **Contexto**

1. **Norma ISO 27001**

La ISO (International Organization for Standardization) es la Organización Internacional de Normalización, cuya principal actividad es la elaboración de normas técnicas internacionales, tanto de productos como de servicios, a través de normas voluntarias que usan las empresas para su eficiencia y rentabilidad.

La norma ISO/IEC 27001 es un estándar para la seguridad de la información aprobado y publicado como estándar internacional en octubre de 2005 y busca cumplir con principios de la tríada de la seguridad: confidencialidad, integridad y disponibilidad de la información. Su objetivo es identificar los riesgos, definiciones de estrategias para evitarlos e implementar salvaguardas.

En ese sentido, ayuda a las organizaciones a establecer la política y los objetivos de gestión de la seguridad de la información y a comprender cómo se pueden gestionar los aspectos importantes, aplicar los controles necesarios y establecer objetivos claros para mejorar la seguridad de la información.

1. **Activos de información**

La cláusula 8 de la ISO 27001 especifica todos los requisitos que deben cumplirse al implementar la seguridad de la información.

En la Gestión de la Seguridad de la Información, los activos se refieren a cualquier elemento valioso para una organización que debe ser protegido del acceso no autorizado, uso, divulgación, modificación, destrucción o compromiso. Esta definición incluye no sólo herramientas tangibles, como hardware, software y equipos de red, sino también los intangibles, como información, conocimientos, propiedad intelectual y reputación.

[*https://blog.invgate.com/es/iso-27001-gestion-activos*](https://blog.invgate.com/es/iso-27001-gestion-activos)

1. **Ley 1712 de 2014 (Ley de Transparencia)**

El objetivo de esta ley es regular el derecho de acceso a la información pública, los procedimientos para el ejercicio y garantía del derecho y las excepciones a la publicidad de información.

1. **Datos abiertos**

"**Datos abiertos**" es una expresión que ganó popularidad con el movimiento de **transparencia** y gobierno **abierto** en todo el mundo, aunque no siempre es tratado con la debida claridad.

Los **datos abiertos** corresponden a información pública dispuesta en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento.

En Colombia la **Ley de Transparencia** y Acceso a la Información, define los datos abiertos como *“todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos”.*

Por lo anterior, podemos resaltar que todos los **datos** que están bajo la guarda del gobierno estatal o de entidades vinculadas a él pueden ser publicados.

Esta información se publica en el sitio oficial *Datos Abiertos Colombia* <https://www.datos.gov.co/>

* **Actividad INDIVIDUAL**

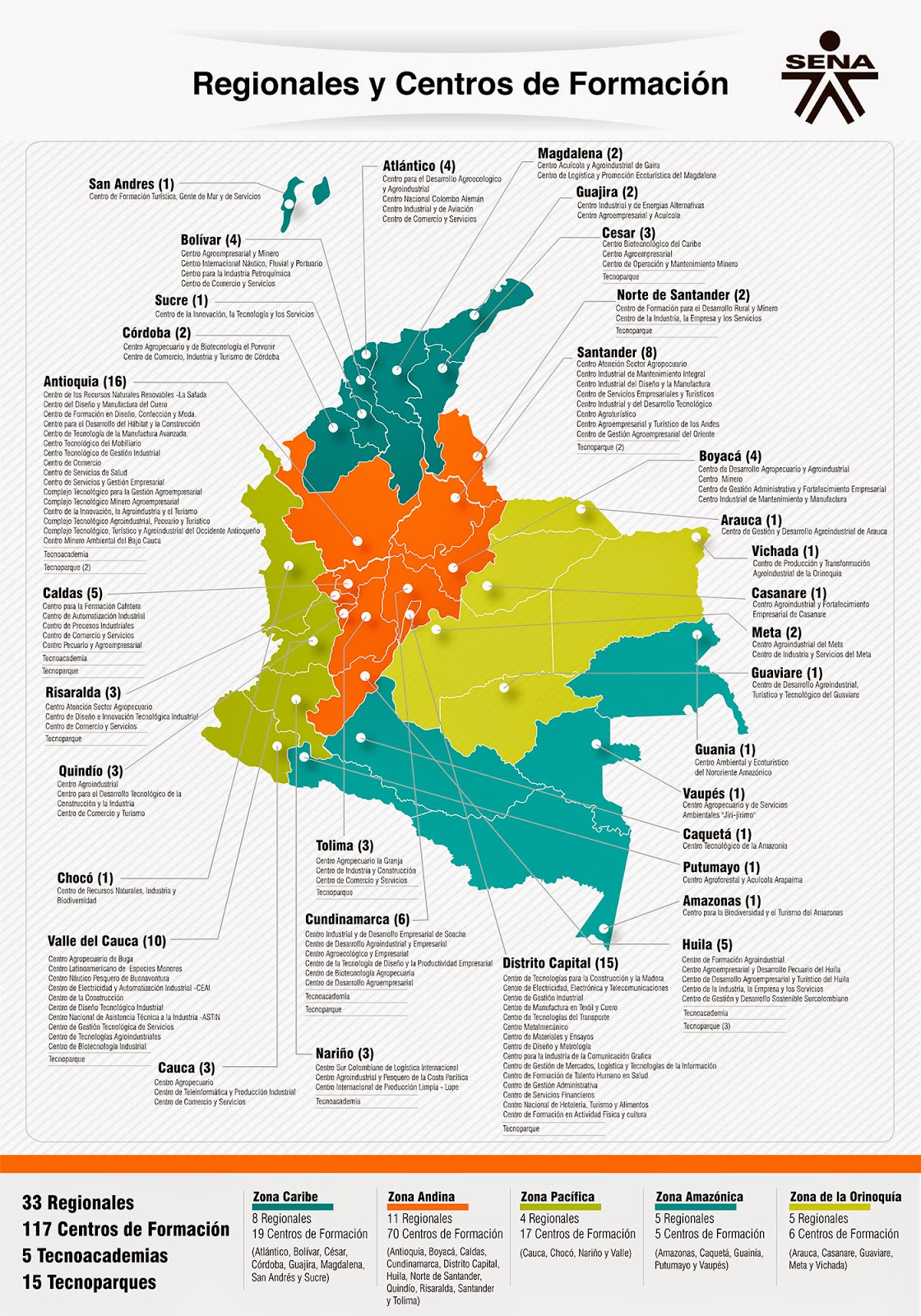
**Base de datos No 1 - Centros de formación SENA**

****

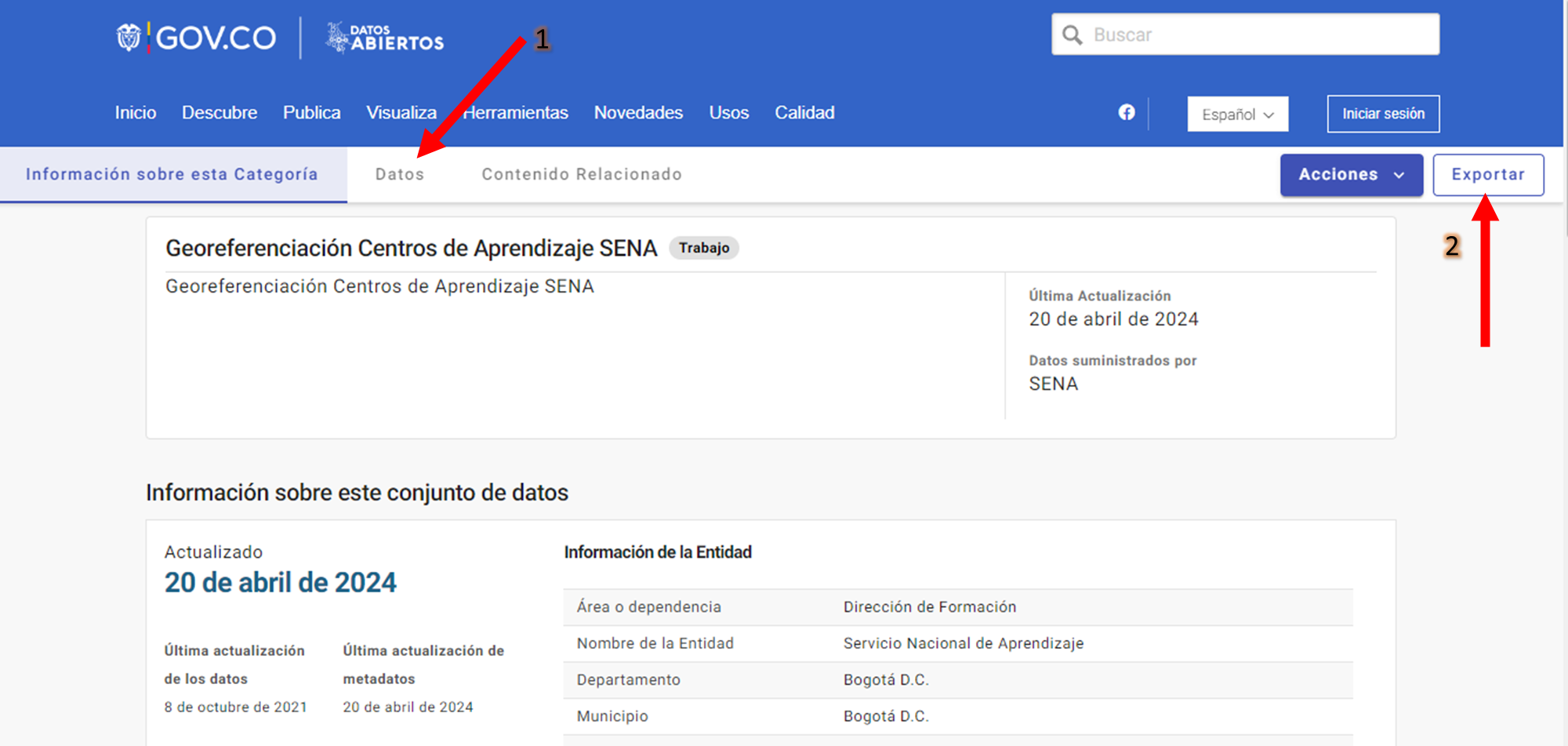
El Servicio Nacional de Aprendizaje (SENA) de Colombia desea crear una base de datos para gestionar la información de sus centros de formación en todo el país, incluyendo su ubicación geográfica.

Los centros ofrecen formación en diversos campos para cubrir las necesidades de la Industria nacional e internacional. La dirección de formación de la institución, ha sido la encargada de publicar los datos referentes a la georreferenciación de todos los centros de formación de cada una de sus 33 regionales.

Es importante comprender que mediante esta base de datos es posible realizar consultas que permitan encontrar los centros de formación dentro de una determinada área geográfica. Así mismo, se abre la posibilidad de integrar la base de datos con un sistema de información geográfica (GIS) para visualizar los centros en un mapa interactivo.

**** [**https://senavaupes.blogspot.com/2015/03/nuestros-centros-de-formacion-por-todo.html**](https://senavaupes.blogspot.com/2015/03/nuestros-centros-de-formacion-por-todo.html)

Se solicita diseñar e implementar una base de datos relacional que permita almacenar los datos pertinentes de cada centro, así como su georreferencia para facilitar su localización en un mapa a partir de los datos publicados por la institución en formatos interoperables y estandarizados, en el portal de datos abiertos del estado colombiano([www.datos.gov.co](http://www.datos.gov.co)).

****

Enlace publicación:

<https://www.datos.gov.co/Trabajo/Georeferenciaci-n-Centros-de-Aprendizaje-SENA/8cnh-7asj/about_data>

**Queries: Tabla consultas.**

Texto de la *consulta* y *capturas de pantalla* de resultados obtenidos de cada consulta.

|  | **Consultas** | **Query** |
| --- | --- | --- |
| **0** | Instrucciones DDL:  Creación de los esquemas necesarios | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **1** | Imprima la lista de tablas de la base de datos | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **2** | Describa cada tabla identificada en el modelo de la base de datos. | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **3** | Inserción de datos utilizando el origen entregado por el instructor (extract - transform). | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **4** | Imprima los registros de cada tabla identificada en el modelo de la base de datos. | *Texto de la consulta:*  Select …  *Captura de pantalla:* |

**::::::::::::::::::::::::::::::::::::::::**

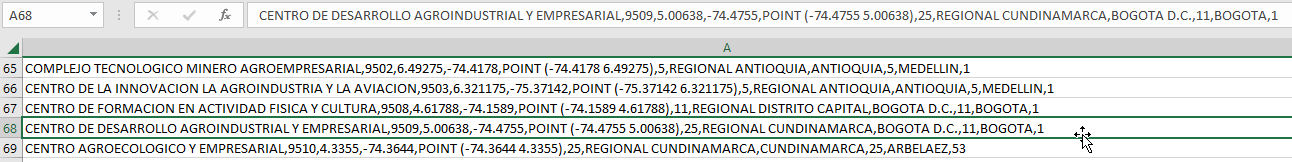
**Tenga en cuenta:**

**::::::::::::::::::::::::::::::::::::::::**

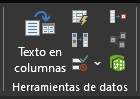
* **Limpieza de los datos:**

La limpieza de datos se considera una etapa previa y separada a los procesos ETL. La limpieza de datos elimina errores y asigna datos de origen al formato de datos de destino. Por ejemplo, si tiene campos numéricos vacíos podría asignar el número 0, si tiene un campo cuyo valor es “P” podría asignar el valor de datos “Principal” o asignar “Secundario” a “S”.

En nuestro archivo, por ejemplo, al revisar los datos de la fila 68 se puede identificar un dato erróneo: el CENTRO DE DESARROLLO AGROINDUSTRIAL Y EMPRESARIAL está ubicado en Villeta (Cundinamarca) y no en Bogotá.



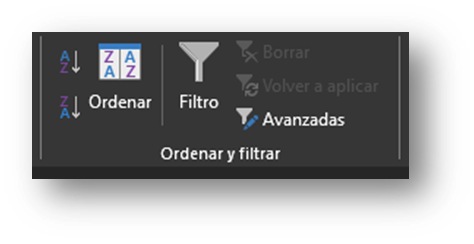
* **Excel:**

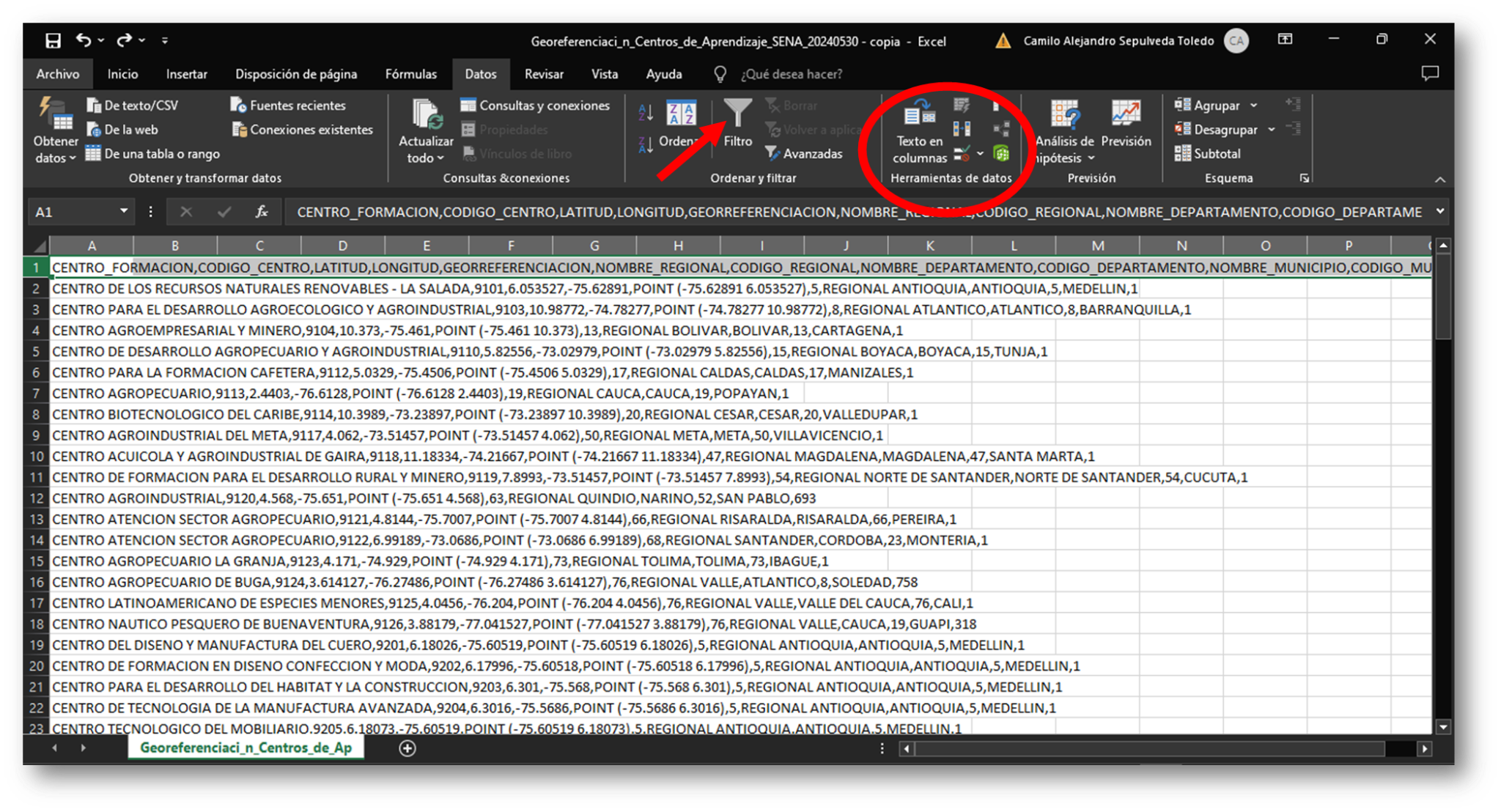
****

Recuerde que es posible dividir texto del archivo .CSV en diferentes columnas con el asistente para convertir texto en columnas y así manipular más fácilmente los datos.

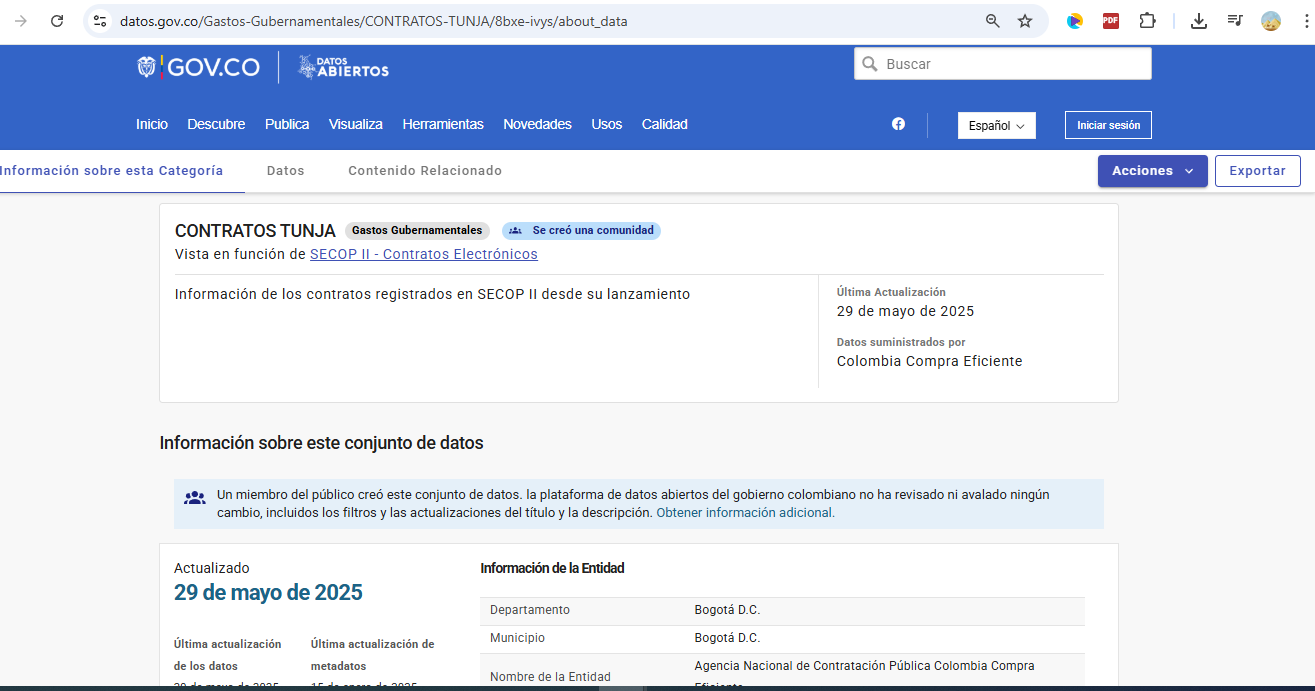
* **Excel:**

No olvide que puede filtrar los datos para poder analizar de mejor manera los datos.



****

**Base de datos No 2 - Contratos Tunja**



Colombia Compra Eficiente es una Entidad descentralizada de la rama ejecutiva del orden nacional, con personería jurídica, patrimonio propio y autonomía administrativa y financiera, adscrita al Departamento Nacional de Planeación ([DNP](https://www.dnp.gov.co/Paginas/inicio.aspx)) que es el rector del Sistema de Compra Pública de Colombia.

Colombia Compra Eficiente ofrece a los partícipes del sistema de compra pública herramientas para facilitar los procesos y fortalecer sus capacidades para obtener mayor valor por el dinero público en el Sistema de Compra Pública colombiana.

Se cuenta con el SECOP II que funciona como una plataforma transaccional con cuentas para las Entidades Estatales y los Proveedores. Cada cuenta tiene unos usuarios asociados a ella. Desde sus cuentas las Entidades Estatales crean, evalúan y adjudican Procesos de Contratación. Los Proveedores pueden hacer comentarios a los Documentos del Proceso, presentar ofertas y seguir el proceso de selección en línea.

La Agencia Nacional de Contratación Pública Colombia Compra Eficiente ha publicado los datos de los contratos registrados en el SECOP II desde su lanzamiento en la ciudad de Tunja.

Se solicita diseñar e implementar una base de datos relacional que permita almacenar los datos pertinentes a cada contrato ejecutado en la alcaldía de Tunja utilizando como fuente de datos el archivo publicado en el portal de datos abiertos del estado colombiano ([www.datos.gov.co](http://www.datos.gov.co)).

Enlace publicación: <https://www.datos.gov.co/Gastos-Gubernamentales/CONTRATOS-TUNJA/8bxe-ivys/about_data>

**Queries: Tabla consultas.**

Texto de la *consulta* y *capturas de pantalla* de resultados obtenidos de cada consulta.

|  | **Consultas** | **Query** |
| --- | --- | --- |
| **0** | Instrucciones DDL:  Creación de los esquemas necesarios | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **1** | Imprima la lista de tablas de la base de datos | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **2** | Describa cada tabla identificada en el modelo de la base de datos. | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **3** | Inserción de datos utilizando el origen entregado por el instructor (extract - transform). | *Texto de la consulta:*  Select …  *Captura de pantalla:* |
| **4** | Imprima los registros de cada tabla identificada en el modelo de la base de datos. | *Texto de la consulta:*  Select …  *Captura de pantalla:* |

**Referencias:**

[1]. Ley 1712 de 2014

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882>

[2]. Datos abiertos

<https://gobiernodigital.mintic.gov.co/portal/Iniciativas/Datos-abiertos/>

[3]. ¿Qué son los datos abiertos?

<https://www.datos.gov.co/stories/s/-T-primera-vez-con-Datos-Abiertos-Aqu-te-contamos-/smn2-7atz/>

[4]. Introducción a los datos abiertos

<https://herramientas.datos.gov.co/sites/default/files/2021-08/20210821IntroduccionDatosAbiertos.pdf>

[5]. Curso Virtual de Datos Abiertos

<https://lms.mintic.gov.co/login/index.php>

[6]. Iniciativa Datos Abiertos de Colombia 2024

<https://www.datos.gov.co/stories/s/Iniciativa-Datos-Abiertos-de-Colombia-2024/nvf3-jun6>

[7]. Informes resultados índices internacionales de datos abiertos

<https://www.datos.gov.co/stories/s/Iniciativa-Datos-Abiertos-Colombia-en-el-Mundo-ndi/jyq6-x6nc>

[8]. Activos de información

<https://blog.invgate.com/es/iso-27001-gestion-activos>

[9]. Guía de clasificación de activos de información

<https://gobiernodigital.mintic.gov.co/692/articles-150528_G5_Gestion_Clasificacion.pdf>

[10]. Dividir texto en diferentes columnas con el Asistente para convertir texto en columnas

<https://support.microsoft.com/es-es/office/dividir-texto-en-diferentes-columnas-con-el-asistente-para-convertir-texto-en-columnas-30b14928-5550-41f5-97ca-7a3e9c363ed7#:~:text=Seleccione%20la%20celda%20o%20columna,los%20Delimitadores%20para%20sus%20datos>.

[11]. Excel - Separar texto en columnas.

<https://www.youtube.com/watch?v=HVxrByb7k8g>

[12]. Filtrar los datos de una tabla o un rango

<https://support.microsoft.com/es-es/office/filtrar-los-datos-de-una-tabla-o-un-rango-01832226-31b5-4568-8806-38c37dcc180e>

[13]. ¿Qué es extracción, transformación y carga (ETL)?

<https://aws.amazon.com/es/what-is/etl/#:~:text=ETL%20combina%20bases%20de%20datos,de%20grandes%20conjuntos%20de%20datos>.

[14]. ¿QUÉ ES EL SECOP II?

<https://www.colombiacompra.gov.co/secop-ii/que-es-el-secop-ii>