Grado 11° Guía 2

TIC

Sesión 1

Aprendizajes esperados

Al final de esta sesión verifica que puedas:

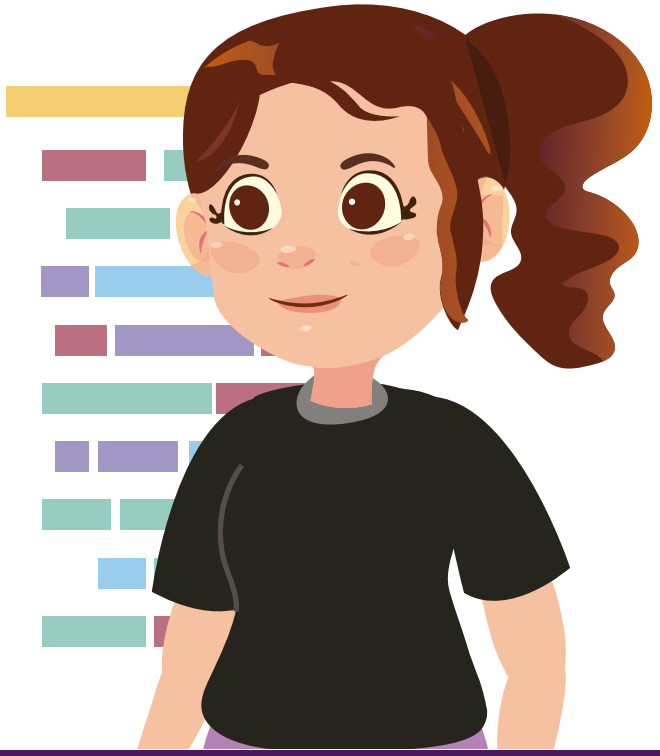
- Reconocer varias estructuras de datos y archivos, incluyendo hojas de cálculo, archivos CSV y *Dataframes*.
- Leer, depurar y escribir código utilizando el paquete Pandas de *Python* para leer datos de un archivo CSV a un *Dataframe* de Pandas.

Material para la clase

- ○ *Anexos 1.1 y 1.2* según el editor de *Python* que utilicen en clase

Duración sugerida

15% 70% 15%



Lo que sabemos, lo que debemos saber

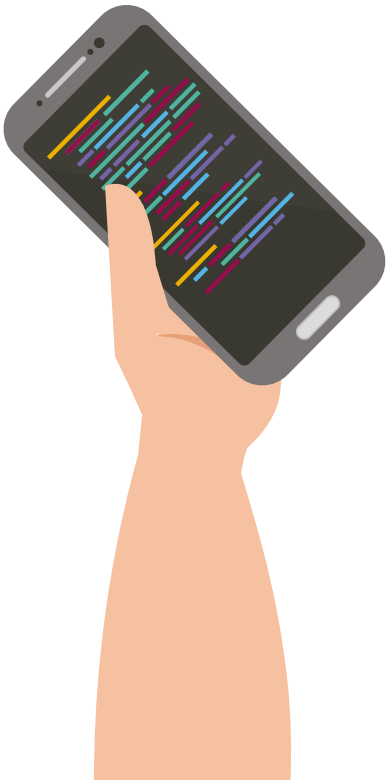
Esta sección corresponde al 15% de avance de la sesión

Inicia la sesión participando en una lluvia de ideas sobre *Python*.

> *¿Qué recuerdas acerca de su uso?*
> *¿Puedes mencionar algunas funciones y características?*

Como ya sabes, *Python* es un lenguaje de programación versátil y muy popular en diferentes áreas. En esta guía aprenderás acerca de su uso en el análisis de datos. *Python* es uno de los lenguajes más utilizados en el mundo del análisis de datos por su simplicidad y su amplio ecosistema de herramientas. A lo largo de esta guía, descubrirás cómo leer archivos y trabajar con datos de manera estructurada.

En grados anteriores, ya has utilizado *Python* para automatizar tareas sencillas, programar juegos interactivos y manipular variables como cadenas de texto y listas de una y dos dimensiones. En esta sesión aprenderás a utilizar un nuevo tipo de variable, pero antes, observa las siguientes imágenes y responde:



> *¿Qué tienen en común las diferentes representaciones?*
> *¿En qué se diferencian?*
> *¿Qué ventajas y desventajas tiene cada tipo de representación?*

Figura 1. Archivo CSV

```
Rango de edad,Sí sabe,No sabe,No informa,No informa.1
De 0 a 9 años*,40.6,10.31,1.43,
De 10 a 19 años*,97.3,1.37 1.33,
De 20 a 29 años*,96.57 1.71,1.72,
De 30 a 39 años*,95.96,2.69,1.35,
```

Figura 2. Datos en un formato de tabla

Rango de edad	Sí sabe	No sabe	No informa	No informa.1
De 0 a 9 años*	40.6	10.31	1.43	
De 10 a 19 años	97.3	1.37	1.33	
De 20 a 29 años	96.57	1.71	1.72	
De 30 a 39 años	95.96	2.69	1.35	
De 40 a 49 años	94.5	4.24	1.26	
De 50 a 59 años	92.47	6.28	1.25	
De 60 a 69 años	88.71	10.34	0.96	

Figura 3. Archivo CSV

```
Nombre,Indicador,Año,Valor
Ingresos Altos,Realizó un pago digital,2000,
Ingresos Altos,"Realizó un pago digital, mujer", 2000,
Ingresos Altos,"Realizó un pago digital, hombre",2000,
Ingresos Altos,% mujeres empleadas,2000,88.4271771152308
Ingresos Altos,% hombres empleados,2000,83.8033567952988
Ingresos Altos,% de empleados total,2000,85.8145356958302
Ingresos bajos y medios ,Realizo un pago digital,2000,
```

Figura 4. Datos en un formato de tabla

Nombre	Indicador	Año	Valor
Ingresos Altos	Realizó un pago digital	2000	
Ingresos Altos	Realizó un pago digital, mujer	2000	
Ingresos Altos	Realizó un pago digital, hombre	2000	
Ingresos Altos	% mujeres empleadas	2000	88.42717712
Ingresos Altos	% hombres empleados	2000	83.8033568
Ingresos Altos	% de empleados total	2000	85.8145357
Ingresos bajos y medios	Realizó un pago digital	2000	
Ingresos bajos y medios	Realizó un pago digital, mujer	2000	
Ingresos bajos y medios	Realizó un pago digital, hombre	2000	

Como puedes notar, las *Figuras 1 y 2* representan el mismo conjunto de datos. Asimismo, las *Figuras 3 y 4* son dos representaciones distintas de otro conjunto de datos.

En este caso, las *Figuras 1 y 3* muestran los datos en su formato **separado por comas** (o CSV por las siglas en inglés de "comma separated values"). Este formato es muy usado para guardar grandes archivos de datos porque ocupa poco espacio en memoria.

En cambio, las *Figuras 2 y 4* representan los datos en un formato de tabla. Las tablas, por su parte, permiten visualizar los datos de una forma clara y ordenada.

En esta sesión, usarás un paquete de *Python* muy popular llamado **pandas,** el cual es esencial para leer, procesar y analizar grandes volúmenes de datos. El paquete **pandas** permite leer archivos y guardarlos en un nuevo tipo de estructura de datos llamada ***dataframe***.

Un *dataframe* se representa como una tabla que organiza la información en filas y columnas, similar a una hoja de cálculo de Excel. Cada columna puede contener un tipo específico de dato, como números, texto o fechas y cada fila representa un registro. Con un dataframe, puedes realizar operaciones como filtrar, ordenar, sumar, promediar y agrupar datos de forma eficiente. Se presenta en la *Figura 5*.

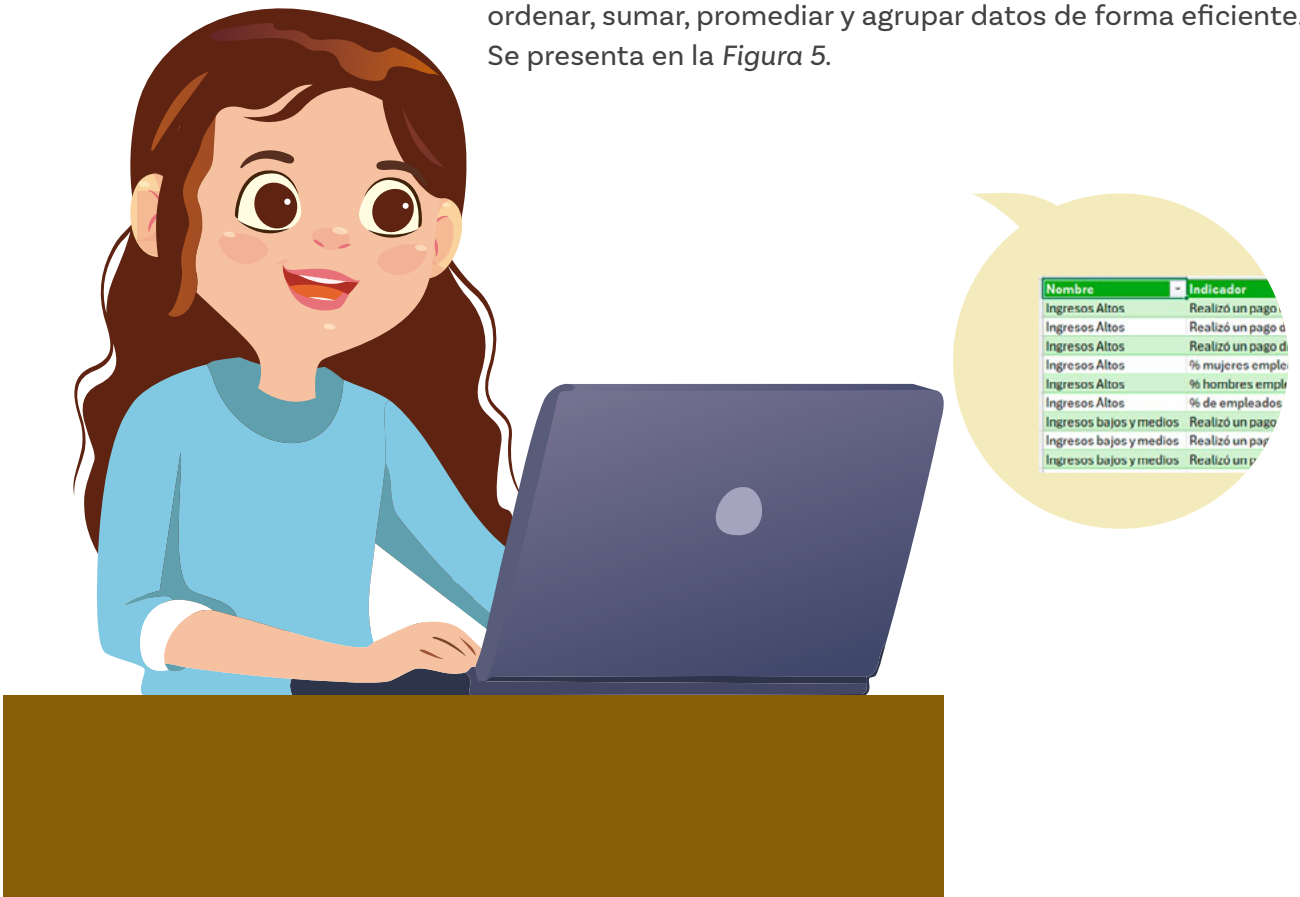
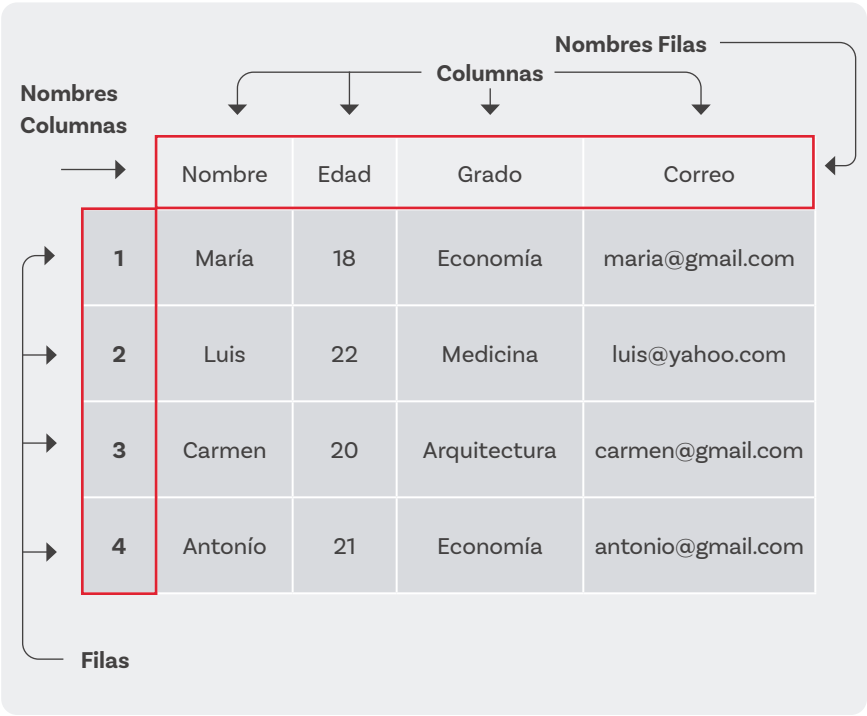


Figura 5. Dataframe



En la siguiente actividad leerás tu primer archivo csv en *Python*.

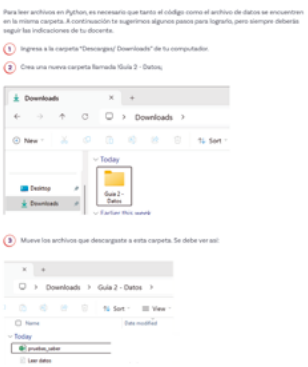
Glosario

- **Archivo CSV:** archivo de texto donde los datos están separados por comas, ideal para almacenar tablas de datos.
- **Dataframe:** estructura de datos en forma de tabla con filas y columnas, utilizada para organizar y analizar información.
- **Paquete:** conjunto de herramientas predefinidas en *Python* que facilitan tareas específicas.
- **Pandas:** paquete de *Python* para manejar, procesar y analizar datos de manera eficiente.

Anexos

Anexo 1.1

Anexo 1.2



Enlace

Archivos de esta sesión:
Leer datos.ipynb / leer datos.py pruebas_saber.csv

Manos a la obra Conectadas

Esta sección corresponde al 85% de avance de la sesión

En esta actividad, aprenderás algunos comandos básicos para leer e interactuar con datos en *Python*. Trabajarás en parejas, utilizando un editor de *Python* como *Anaconda*, *Google Colab* o *Python-Online*, dependiendo de lo que tu docente te indique. Si tienes dudas sobre cómo acceder a alguno de estos editores, consulta el *Anexo 1.1*, que contiene tutoriales detallados para cada opción.

Para empezar, descarguen los archivos con los que van a trabajar durante la primera parte de la actividad. Sigan el enlace o utilicen el QR y guarden el archivo en el escritorio del computador. Recuerden que los archivos *.ipynb* se refieren a cuadernos interactivos que pueden utilizar en *Anaconda* o *Google Colab*, y los archivos *.py* se refieren a código de *Python* que pueden utilizar en editores como *Python-Online*.

Abran el archivo pruebas_saber.csv haciendo doble clic.

¿Qué observan?
¿Qué esperan ver en Python cuando lo utilicen?

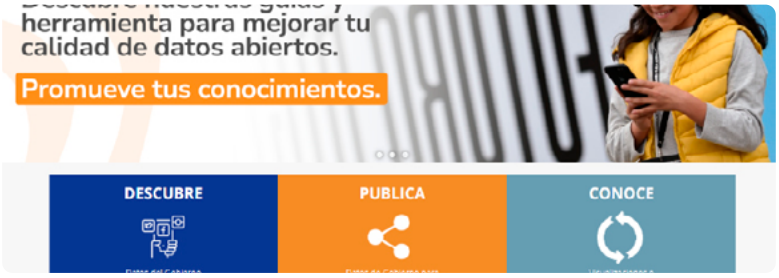
A continuación, tu docente entregará una copia del *Anexo 1.2* que corresponda con el editor que están utilizando en el salón. Sigan las indicaciones para cargar el archivo de datos y ejecutar los primeros comandos.

Cuando terminen, es el momento de buscar y explorar datos que les interesen. Para esto, van a utilizar la página de Datos Abiertos del gobierno de Colombia. Recuerden que los Datos Abiertos son información pública que se puede usar, reutilizar y redistribuir libremente.

Sigan los pasos para encontrar y descargar un archivo de datos que te interese a ti y a tu compañero o compañera.

1. Ingresen a datos.gov.co y den clic en “Descubre” como se presenta en la *Figura 6*.

Figura 6. Sitio web de datos.gov.co



2. Seleccionen el filtro “Conjuntos de datos” como se presenta en la *Figura 7*.

Figura 7. Conjunto de datos

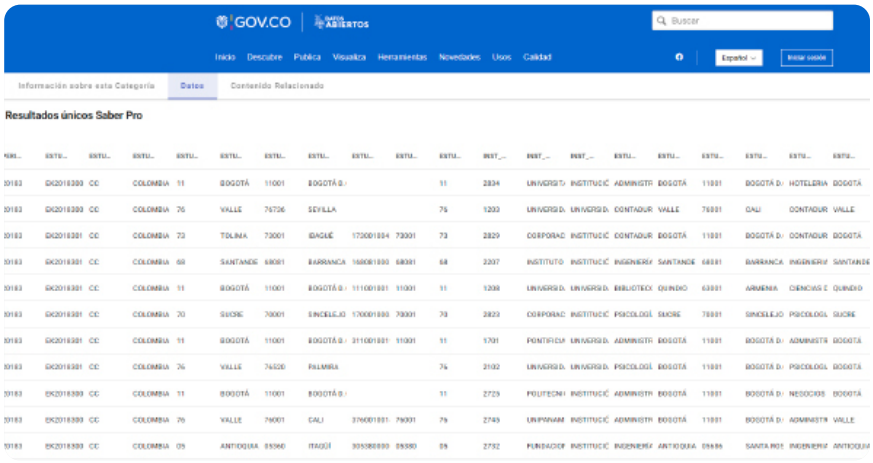
3. Seleccionen un tema de su interés en la sección “Clasificación” como en la lista que se presenta en la *Figura 8*.

Figura 8. Seleccion temática de interés

4. Exploren los diferentes conjuntos de datos y hagan clic en uno que capte su atención.

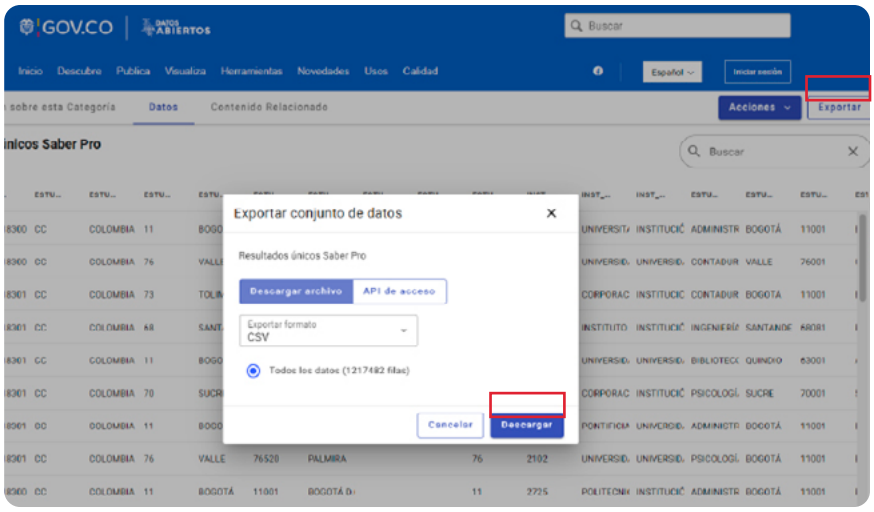
5. Den clic en la pestaña “Datos” y observa los resultados como se presenta en la *Figura 9*.

Figura 9. Resultados y presentaciónes de datos



6. Exporten los datos como archivo CSV tal y como se visualiza en la *Figura 10*.

Figura 10. Exportar datos como CSV



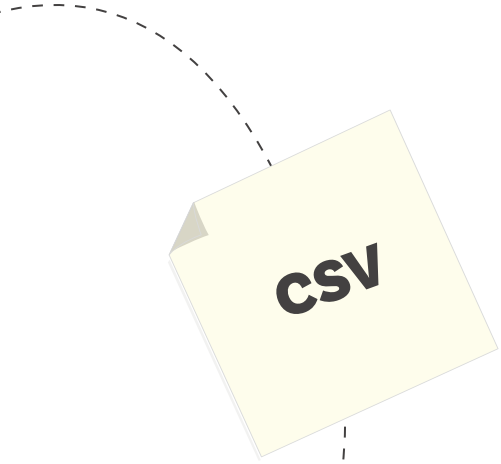
Ahora regresen al código, carguen y exploren los datos en *Python*.

Prepárate para compartir tus hallazgos con el resto de la clase al terminar la sesión.

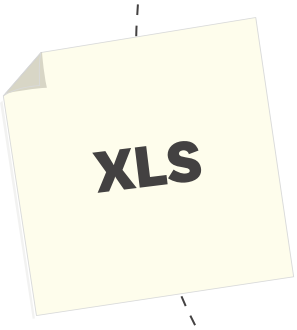
Antes de irnos

Esta sección corresponde al 100% de avance de la sesión

Revisa los aprendizajes de la sesión de forma individual respondiendo las preguntas de forma que mejor reflejen tu progreso:

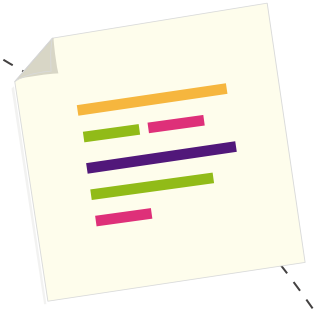


1. ¿Puedes reconocer varias estructuras de datos y archivos, incluyendo hojas de cálculo, archivos CSV y dataframes?
   - ◯ Sí
   - ◯ Parcialmente
   - ◯ Aún no



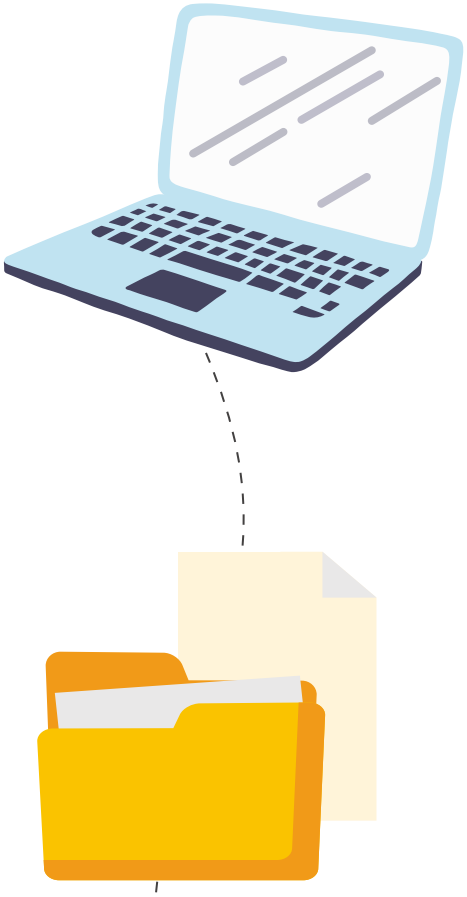
2. ¿Puedes leer, depurar y escribir código utilizando la biblioteca Pandas de *Python* para leer datos de un archivo CSV a un *dataframe* de Pandas?
   - ◯ Sí
   - ◯ Parcialmente
   - ◯ Aún no

Si tus respuestas fueron "Parcialmente" o "Aún no", vuelve a leer los contenidos. Resalta o subraya los términos que no hayas comprendido. Luego, discute con tus compañeras y compañeros de grupo lo que se hizo en cada momento de la actividad y el rol al que correspondía. Si todavía te quedan dudas, consúltale a tu docente.



Para cerrar la sesión, completa el siguiente tiquete de salida:

Fecha: _____



Piensa en el conjunto de datos que descargaste y exploraste en *Python*. ¿Qué preguntas podrías responder con estos datos?

¿Cómo relacionas lo aprendido en esta clase con tus conocimientos previos de *Python*?