

Guía de Preparación

Sesión Introductoria de Web Scraping con Python

1. Objetivo del documento

Este documento tiene como objetivo que el alumnado llegue a la sesión con el entorno de trabajo correctamente instalado y configurado. Durante la sesión se trabajará con Jupyter Notebook dentro de un entorno aislado creado con Miniconda.

2. Instalación de Miniconda

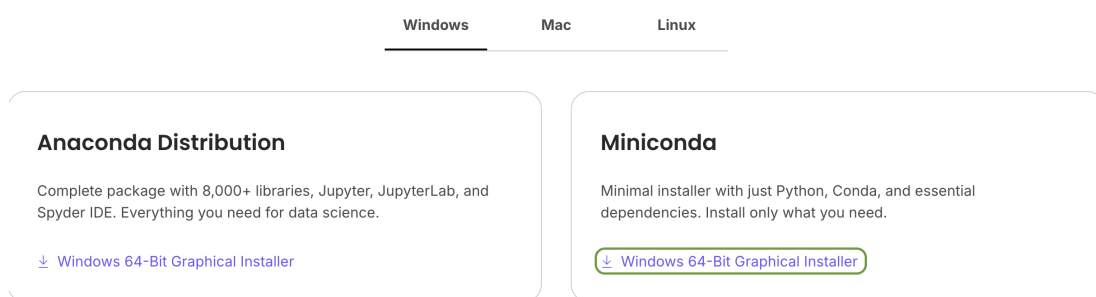
Puedes encontrar más información sobre Anaconda y Miniconda en:

<https://www.anaconda.com/download>

Para la sesión se recomienda instalar Miniconda por contener los paquetes esenciales sin llegar a pesar tanto como Anaconda. Al instalar Miniconda se instalará paralelamente Python y todo lo necesario para empezar a ejecutar código y visualizar resultados.

Puedes instalar Miniconda (sin necesidad de introducir ningún dato de contacto) en el siguiente enlace:

<https://www.anaconda.com/download/success>



Para Mac: seleccionar la descarga según si el chip es de Apple (M1, M2, M3...) o es de Intel:

WindowsMacLinux

Anaconda Distribution

Complete package with 8,000+ libraries, Jupyter, JupyterLab, and Spyder IDE. Everything you need for data science.

- 64-Bit (Apple silicon) Graphical Installer
- 64-Bit (Apple silicon) Command Line Installer

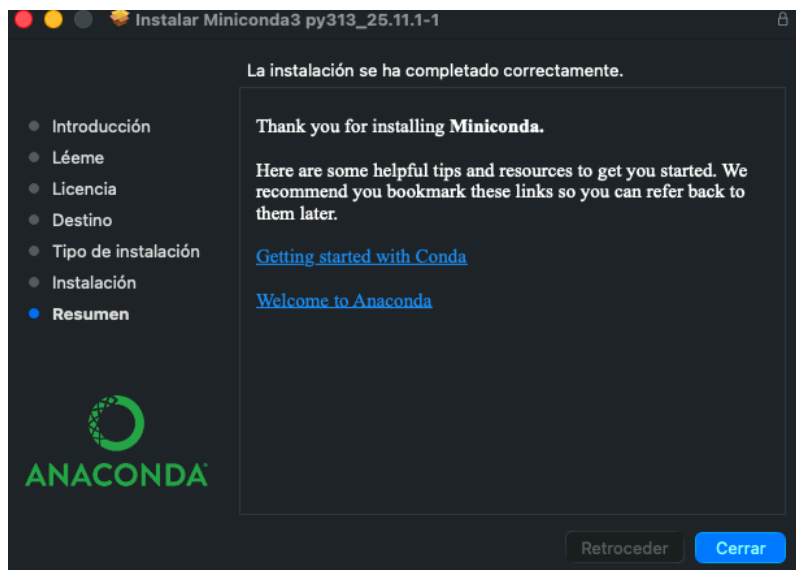
Miniconda

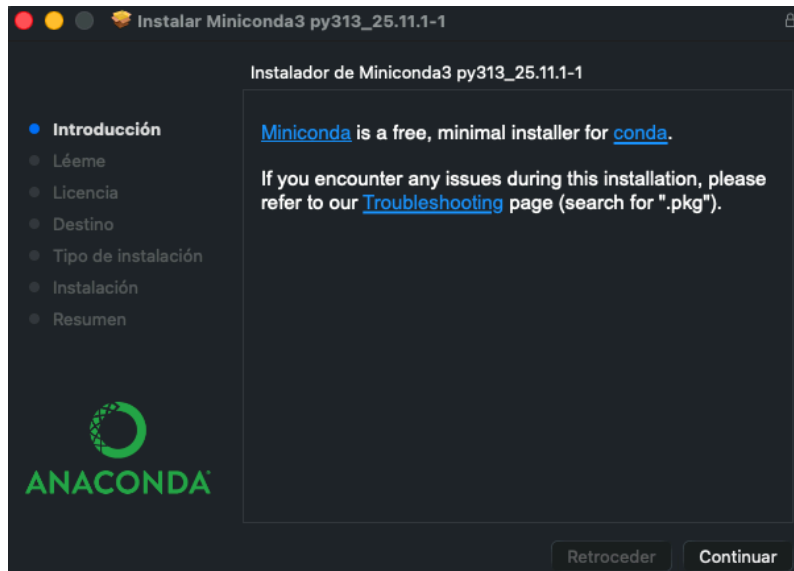
Minimal installer with just Python, Conda, and essential dependencies. Install only what you need.

- 64-Bit (Apple silicon) Graphical Installer
- 64-Bit (Apple silicon) Command Line Installer
- 64-Bit (Intel chip) Graphical Installer
- 64-Bit (Intel chip) Command Line Installer

3. Verificar la instalación

Una vez descargado, abrir el archivo .exe (Windows) o .pkg (Mac) y seguir las indicaciones del instalador. En este proceso se debe aceptar la licencia, seleccionar dónde se instala y qué usuarios tendrán acceso. La instalación ocupará aproximadamente 580 MB.





Puedes verificar que se ha instalado correctamente de la siguiente manera:

En Windows: buscar el programa `Anaconda Prompt (Miniconda3)`

En Mac: abrir el Terminal y escribir:

```
conda --version
```

Al presionar Enter debería aparecer algo como:

```
conda 25.11.1
```

4. Creación de carpeta de trabajo y clonación del repositorio

Este paso y el siguiente lo realizaremos juntos durante la sesión, pero dejo las instrucciones por si quieres probar.

Abrir Anaconda Prompt (Windows) o Terminal (macOS) y ejecutar los siguientes comandos:

```
cd Desktop
mkdir intro-scraping
cd intro-scraping
git clone https://github.com/Sergi-pmm/WebScrapingSession
cd WebScrapingSession
```

En algunos equipos con Windows, el Escritorio puede encontrarse dentro de OneDrive y llamarse Desktop o Escritorio.

Nota 1: Como alternativa a mkdir, puedes crear la carpeta `intro-scraping` manualmente desde el escritorio.

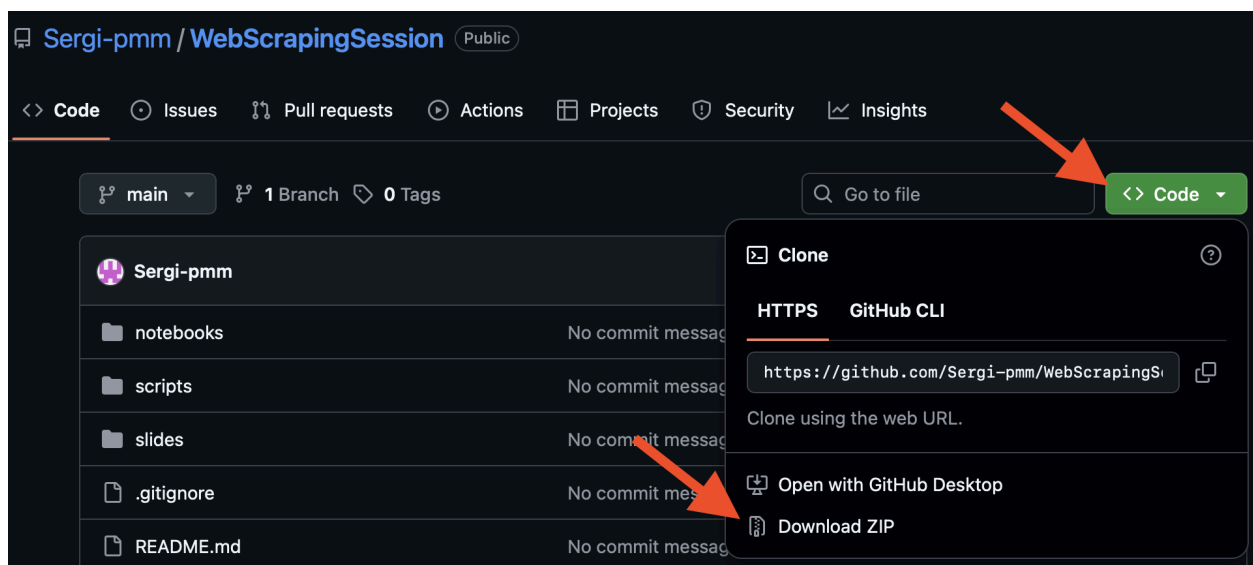
Nota 2: Para clonar el repositorio es necesario tener Git instalado.

En macOS, si Git no está instalado, el sistema suele solicitar su instalación automáticamente al ejecutar el comando git por primera vez.

En Windows, es posible que sea necesario instalar Git previamente (se indicará durante la sesión si fuera el caso).

Git no es imprescindible durante la sesión por lo que si no se desea descargar el programa se puede acceder y descargar el repositorio en:

<https://github.com/Sergi-pmm/WebScrapingSession>



5. Creación y activación del entorno virtual

Este paso lo realizaremos juntos durante la sesión, pero dejo instrucciones por si alguien lo quiere probar.

Crear el entorno con Python 3.10:

```
conda create -n intro-scraping-session python=3.10
```

Para usuarios de Mac, aparecerá el siguiente mensaje:

```
Do you accept the Terms of Service (ToS) for  
https://repo.anaconda.com/pkgs/main? [(a)ccept/(r)eject/(v)iew]: █
```

Escribir a + Intro

Para usuarios de Windows, aparecerá un indicador de que se van a instalar una serie de paquetes. Para proceder, escribir y + Intro

Finalmente, activar el entorno:

```
conda activate intro-scraping-session
```

Para desactivar el entorno usa el comando:

```
conda deactivate
```

Si tienes cualquier duda puedes contactarme a spmm3113@gmail.com