



Coordinación de
Educación Abierta y a Distancia
VICERRECTORADO ACADÉMICO



CULTURA DIGITAL Y SOCIEDAD

Actividad Autónoma 1

Unidad 2: Herramientas y Metodologías en Ciencia de Datos



FACULTAD DE
Ingeniería

Nombres: Ana Cristina Lima Merino

Fecha: 08/11/2025

Carrera: Ingeniería Ciencia de Datos e Inteligencia Artificial

Periodo académico: 1S 2025

Semestre: TERCERO

Lik repositorio: <https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git>

INFORME

En el presente documento se resumirá el proceso realizado para la creación de un repositorio en GitHub. El objetivo es demostrar lo aprendido en clases desde la organización de las carpetas , la creación y clonación del repositorio y también de sus ramas.

GitHub es una plataforma que permite colaborar entre distintas personas en él se alojan proyectos en la nube, en él tiene un sistema que ayuda a controlar las versiones que es el Git. Esta plataforma nos ayuda a administrar el código sobre todo porque se lleva el registro de los cambios que se han realizado(Saavedra, 2023).

La primera parte consistió en crear un repositorio en GitHub el cual lo llamamos “preprocesamiento-ciencia-datos”. Debemos saber que un repositorio es una carpeta en la nube en donde se van guardando todos los archivos que están relacionados con el proyecto. De esta manera podemos acceder desde cualquier computadora y también conocer todo lo que se va modificando.

Create a new repository

Repositories contain a project's files and version history. Have a project elsewhere? [Import a repository](#).

Required fields are marked with an asterisk (*).

1 General

Owner *	Repository name *
AnaCristina2308	/ preprocesamiento-ciencia-datos
<small>✓ preprocesamiento-ciencia-datos is available.</small>	

Dentro de este repositorio creamos un archivo llamado gitignore, el que nos sirve para indicarle al Git lo que debe ignorar o no queremos que vayan en el repositorio. En este caso excluimos las carpetas pycache y .pcy de Python ya que estas se generan de manera automática. También excluimos la carpeta .vscocore porque este contiene configuraciones personales y el archivo .env que puede almacenar información personal o contraseñas.

Así mismo, sabemos que siempre es necesario crear un archivo llamado README en este se debe describir el proyecto. En este caso colocamos el nombre, el objetivo y la organización de las carpetas. Luego hice el primer comité para guardar estos documentos en el repositorio.

```

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ git commit -m "Subir los archivos git ignore y README"
[main (root-commit) 1fd184c] Subir los archivos git ignore y README
 2 files changed, 52 insertions(+)
  create mode 100644 .gitignore
  create mode 100644 README.md

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ git push -u origin main
Enumerating objects: 4, done.
Counting objects: 100% (4/4), done.
Delta compression using up to 4 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 891 bytes | 297.00 KiB/s, done.
Total 4 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
To https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
 * [new branch]      main -> main
branch 'main' set up to track 'origin/main'.

```

En la segunda parte de este trabajo vamos a clonar el proyecto que ya creamos, para esto utilizamos otra carpeta en mi caso llamada Tareas FundamentosBD. Entonces lo clonamos para trabajar con una copia local y no trabajarla directamente en la versión en línea. Entonces, utilizamos el comando git clone y agregamos la dirección de nuestro repositorio. Aquí también procedí a iniciar sesión configurando mi nombre de usuario y correo, para esto se ocupó los comandos “git config --global analima2308” y “git config --global ana.limam2000@gmail.com”

```

Cris@DESKTOP-S8IQI8A MINGW64 /
$ cd c:

Cris@DESKTOP-S8IQI8A MINGW64 /c
$ cd c:Tareas_FundamentosBD/

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD
$ git clone https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
Cloning into 'preprocesamiento-ciencia-datos'...
remote: Enumerating objects: 4, done.
remote: Counting objects: 100% (4/4), done.
remote: Compressing objects: 100% (4/4), done.
remote: Total 4 (delta 0), reused 4 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (4/4), done.

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD
$ cd preprocesamiento-ciencia-datos
bash: cd: preprocesamiento-ciencia-datos: No such file or directory

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD
$ cd preprocesamiento-ciencia-datos

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos
(main)
$ git config --global user.name "AnaCristina2308"
$ git config --global user.email "ana.limam2000@gmail.com"

```

Posteriormente creó una nueva rama con el comando “git checkout -b feature-preprocesamiento”. Esta nos sirve para que trabajemos sin afectar directamente nuestro proyecto principal, y si cometemos errores no se lo hagan directamente en el main.

```

$ git checkout -b feature-preprocesamiento
Switched to a new branch 'feature-preprocesamiento'

```

Dentro de esta carpeta creamos el archivo [preprocesamiento.py](#) en el vamos a poner todas las funciones básicas de Python que nos ayuda para el procesamiento de datos como cargar el archivo, eliminar los nulos, normalizar datos y otros.

Después de esto vamos a guardar los cambios para esto vamos a ocupar dos comandos. El primero es el git add éste guarda en la memoria los cambios que hicimos, y “git commit” que nos sirve para poner un comentario de los cambios que hicimos. y luego subimos estos cambios al GitHub a través de un “git push” este sube los cambios que guardamos de manera local a la nube en nuestro repositorio.

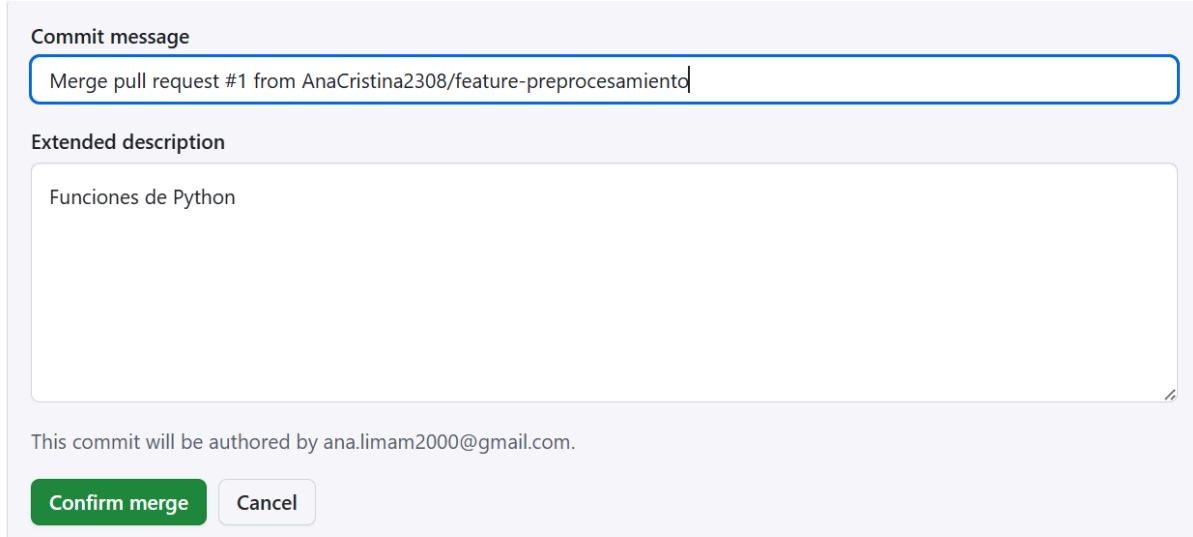
```
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git add .

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git commit -m "Agregue el archivo con las funciones de Python necesarias para el procesamiento de una base de datos completa"
[feature-preprocesamiento f8bd0ee] Agregue el archivo con las funciones de Python necesarias para el procesamiento de una base de datos completa
 1 file changed, 13 insertions(+)
 create mode 100644 src/procesamiento/preprocesamiento.py

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git push -u origin feature-preprocesamiento
Enumerating objects: 6, done.
Counting objects: 100% (6/6), done.
Delta compression using up to 4 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (5/5), 634 bytes | 634.00 KiB/s, done.
Total 5 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
remote:
remote: Create a pull request for 'feature-preprocesamiento' on GitHub by visiting:
remote:   https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos/pull/
new/feature-preprocesamiento
remote:
To https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
 * [new branch]      feature-preprocesamiento -> feature-preprocesamiento
branch 'feature-preprocesamiento' set up to track 'origin/feature-preprocesamiento'.

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ |
```

Para la tercera parte del trabajo debemos regresar al GITHUB, en este vamos a hacer un Pull Request, esto hace que se unan los cambios que hicimos en nuestra rama con la rama principal. Aquí se puede solicitar que sea revisado por cualquier persona con acceso al repositorio. Aquí hacemos el Merge que es fusionar y a la vez eliminamos la rama porque esta ya no es necesaria.



CAPTURAS DE PANTALLA

Create a new repository

Repositories contain a project's files and version history. Have a project elsewhere? [Import a repository.](#)
Required fields are marked with an asterisk (*).

1 General

Owner * Repository name *

AnaCristina2308 / preprocesamiento-ciencia-datos

preprocesamiento-ciencia-datos is available.

Great repository names are short and memorable. How about [reimagined-couscous?](#)

Description

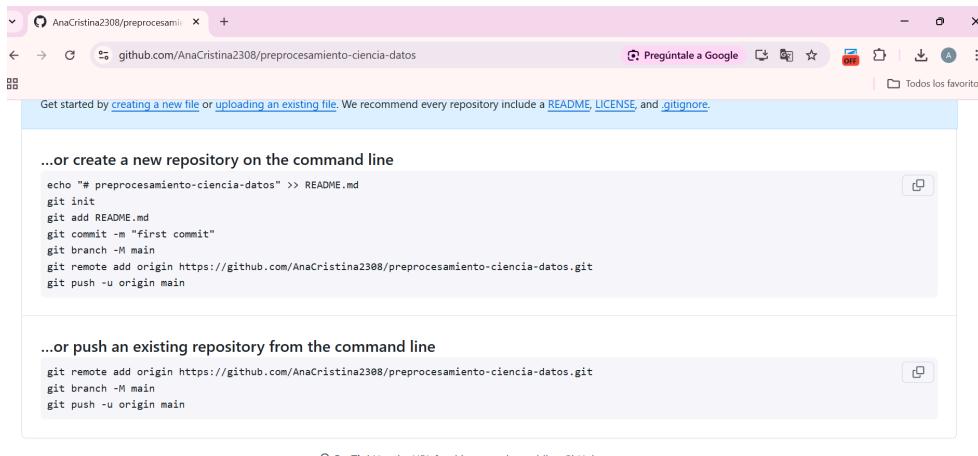
0 / 350 characters

2 Configuration

Choose visibility *

Choose who can see and commit to this repository

Public



Crear carpeta

```
Cris@DESKTOP-S8IQI8A MINGW64 ~
$ cd c:

Cris@DESKTOP-S8IQI8A MINGW64 /c
$ cd FundamentosBD
bash: cd: FundamentosBD: No such file or directory

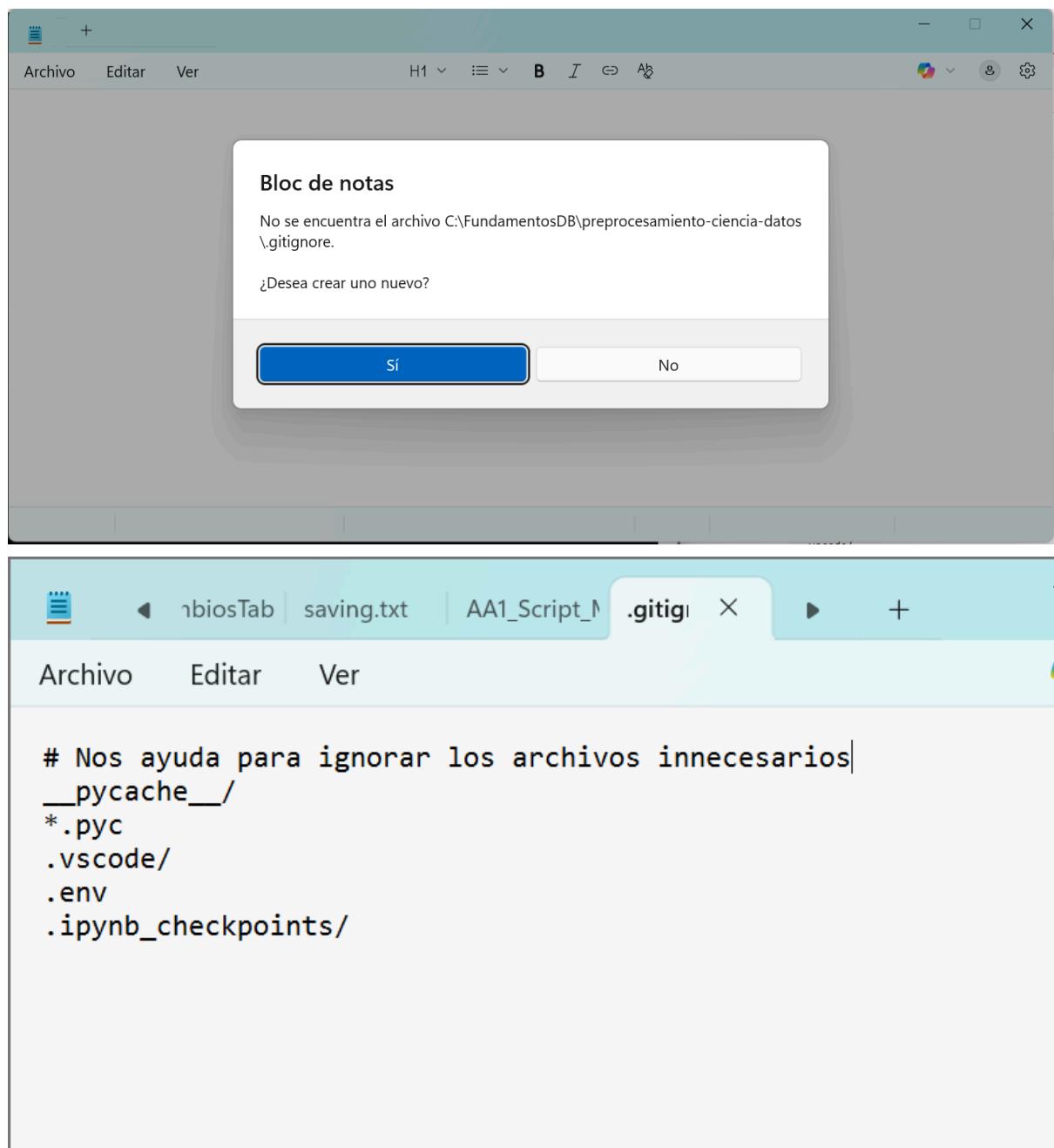
Cris@DESKTOP-S8IQI8A MINGW64 /c
$ cd c:FundamentosDB/

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB
$ mkdir preprocesamiento-ciencia-datos

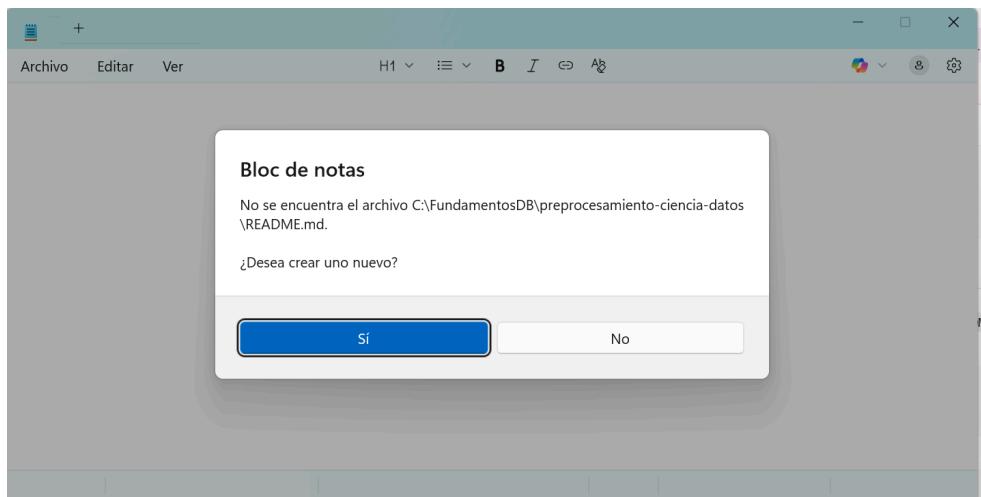
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB
$ cd preprocesamiento-ciencia-datos/

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos
$ git init
Initialized empty Git repository in C:/FundamentosDB/preprocesamiento-ciencia-datos/.git/
```

Configurar el gitignore



Para crear el archivo README



```
Cris@DESKTOP-S8IQI8A MINGW64 ~
$ cd c:
Cris@DESKTOP-S8IQI8A MINGW64 /c
$ cd FundamentosBD
bash: cd: FundamentosBD: No such file or directory

Cris@DESKTOP-S8IQI8A MINGW64 /c
$ cd c:FundamentosDB/
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB
$ mkdir preprocesamiento-ciencia-datos

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB
$ cd preprocesamiento-ciencia-datos/

Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos
$ git init
Initialized empty Git repository in C:/FundamentosDB/preprocesamiento-ciencia-datos/.git/
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (master)
$ git remote add origin https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (master)
$ git branch -M main
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ notepad .gitignore
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ notepad README.md
```

```
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos
$ git init
Initialized empty Git repository in C:/FundamentosDB/preprocesamiento-ciencia-datos/.git/
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (master)
$ git remote add origin https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (master)
$ git branch -M main
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ notepad .gitignore
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ notepad README.md
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ git add .
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ git commit -m "Subir los archivos ignore y README"
[master 1fdd14c] Subir los archivos git ignore y README
 2 files changed, 52 insertions(+)
  create mode 100644 .gitignore
  create mode 100644 README.md
Cris@DESKTOP-S8IQI8A MINGW64 /c/FundamentosDB/preprocesamiento-ciencia-datos (main)
$ git push -u origin main
Enumerating objects: 4, done.
Counting objects: 100% (4/4), done.
Delta compression using up to 4 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 891 bytes | 297.00 KiB/s, done.
Total 4 (delta 0), reused 0 delta(s), pack-reused 0 (from 0)
To https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
 * [new branch]      main -> main
branch 'main' set up to track 'origin/main'.
```

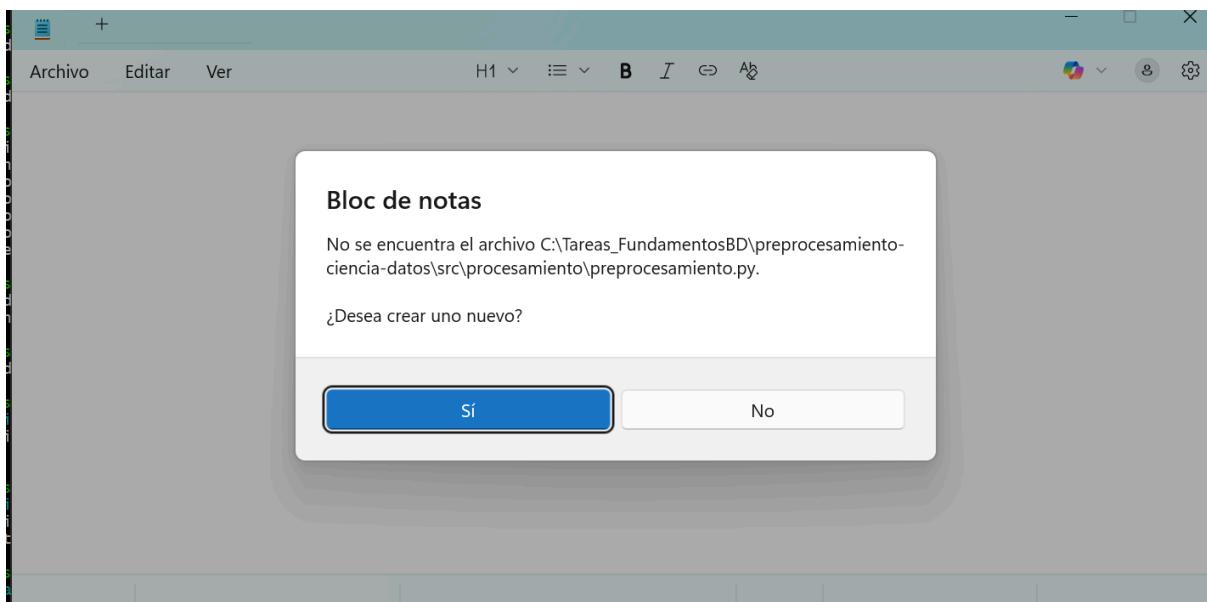
PARTE 2: Implementación de Git y GitHub Instrucciones: 1. Clonación y Configuración: o Clona el repositorio creado usando git clone. o Configura tu usuario y correo electrónico:

```
Cris@DESKTOP-S8IQI8A MINGW64 /  
$ cd c:  
  
Cris@DESKTOP-S8IQI8A MINGW64 /c  
$ cd c:Tareas_FundamentosBD/  
  
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD  
$ git clone https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git  
Cloning into 'preprocesamiento-ciencia-datos'...  
remote: Enumerating objects: 4, done.  
remote: Counting objects: 100% (4/4), done.  
remote: Compressing objects: 100% (4/4), done.  
remote: Total 4 (delta 0), reused 4 (delta 0), pack-reused 0 (from 0)  
Receiving objects: 100% (4/4), done.  
  
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD  
$ cd preprocesamiento-ciencia-datos  
bash: cd: preprocesamiento-ciencia-datos: No such file or directory  
  
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos  
(main)  
$ git config --global user.name "AnaCristina2308"  
git config --global user.email "ana.limam2000@gmail.com"  
  
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos  
(main)  
$
```

2. Creación y Gestión de Ramas: o Crea una rama llamada feature-preprocesamiento. o En esta rama, crea un archivo preprocesamiento.py que contenga funciones en Python para realizar el preprocesamiento completo de un dataset. o Realiza un commit y sube los cambios a GitHub:

```
$ git checkout -b feature-preprocesamiento  
Switched to a new branch 'feature-preprocesamiento'
```

En esta rama, crea un archivo preprocesamiento.py que contenga funciones en Python para realizar el preprocesamiento completo de un dataset.



Realiza un commit y sube los cambios a GitHub:

```
Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git add .

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git commit -m "Agregue el archivo con las funciones de Python necesarias para el procesamiento de una base de datos completa"
[feature-preprocesamiento f8bd0ee] Agregue el archivo con las funciones de Python necesarias para el procesamiento de una base de datos completa
 1 file changed, 13 insertions(+)
 create mode 100644 src/procesamiento/preprocesamiento.py

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ git push -u origin feature-preprocesamiento
Enumerating objects: 6, done.
Counting objects: 100% (6/6), done.
Delta compression using up to 4 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (5/5), 634 bytes | 634.00 KiB/s, done.
Total 5 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
remote:
remote: Create a pull request for 'feature-preprocesamiento' on GitHub by visiting:
remote:   https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos/pull/
new/feature-preprocesamiento
remote:
To https://github.com/AnaCristina2308/preprocesamiento-ciencia-datos.git
 * [new branch]      feature-preprocesamiento -> feature-preprocesamiento
branch 'feature-preprocesamiento' set up to track 'origin/feature-preprocesamiento'.

Cris@DESKTOP-S8IQI8A MINGW64 /c/Tareas_FundamentosBD/preprocesamiento-ciencia-datos (feature-preprocesamiento)
$ |
```

3. Pull Request y Fusión: Crea una pull request en GitHub para fusionar feature-preprocesamiento con main.

Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks. [Learn more about diff comparisons here.](#)

Add a title

Add a description

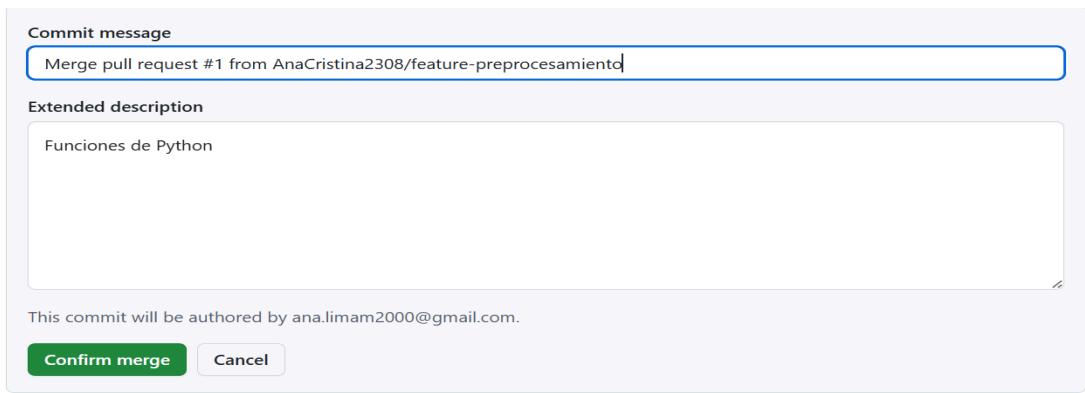
Write Preview

Agregue el archivo con las funciones de Python necesarias para el procesamiento de una base de datos completa

Reviewers
No reviews

Assignees
No one—[assign yourself](#)

Labels
None yet



Solicita la revisión de un compañero (simulado) y acepta la pull request. o Elimina la rama feature-preprocesamiento tras la fusión.

Pull request successfully merged and closed Delete branch
You're all set — the `feature-preprocesamiento` branch can be safely deleted.

Configuracion del Git Action

```

1 name: Verificación de preprocesamiento
2
3 on:
4   push:
5   pull_request:
6
7 jobs:
8   test:
9     runs-on: ubuntu-latest
10
11 steps:
12   - name: Descargar el repositorio
13     uses: actions/checkout@v3

```

Referencias

Saavedra, J. A. (2023, 4 de junio). *Qué es Github: para qué sirve, cómo funciona y cómo usarlo*. EBAC. <https://ebac.mx/blog/que-es-github>