

## PREPROCESAMIENTO DE DATOS

El proyecto tipreprocesamiento cienciadatos tiene como objetivo aplicar el uso de Git y Github para la gestion y colaboracion en proyectos de ciencia de datos donde se implementa un preprocesamiento completo de datasets usando la biblioteca de Pandas , incluye:

- Gestion de valores nulo
- Eliminacion de valores duplicados
- Normalizacion de las columnas
- Codificaion de variables categoricas

Nos permite preparar datasets para analisis de modelos machine learning asegurandonos datos limpios y consistentes.

### COMANDOS UTILIZADOS:

Git init : Inicializa un repositorio git en la carpeta local

Git add: Agregar todos los archivos nuevos

Glt comit -m: Guarda los cambios en el historial de git con un mensaje descriptivo

Git Push: Sube los cambios locales a la rama de GitHub

Git checkout: Crea y tambien cambia una nueva rama

Git pull origin main: Actualiza la rama local main con los ultimos cambios

Git brach : Elimina una rama local despues de fusionarla

### PULL REQUEST Y FUSION

- Primero se creo una rama feature-preprocesamiento para trabajar en las rama
- Se subio la rama a GitHub usando git push origin feature-preprocesamiento
- Se creo un Pull Request caomparando feature-preprocesamiento con main
- Revisamiento de la PR y se acepta la fusion usando Merge pull request
- Se elimino la rama feature-preprocesamiento tanto en GitHub como localmente

### AUTOMATIZACION CON GITHUB ACTIONS

Se creo un wokflow basico que ayuda a verificar la ejecucion de preprocesamiento.py cada vez que se hace un push o main

- Cada vez que se usa psuh o main se activa
- Instala Python3.10
- Instala las dependencias necesarias pandas y scikit learn
- Ejecuta el archivo

### EVIDENCIAS

```
PS C:\Users\MARIA JOSE\Downloads\UNACH-Semestres\Tercer-Semestre\CulturaDigital\Homeworks\preprocesamiento-cienciadatos> git add preprocesamiento.py
>> git commit -m "Agregar script de preprocesamiento de datos"
>> git push origin feature-preprocesamiento
>>
[feature-preprocesamiento b43e1c7] Agregar script de preprocesamiento de datos
1 file changed, 1 insertion(+), 1 deletion(-)
Enumerating objects: 5, done.
Counting objects: 100% (5/5), done.
Delta compression using up to 16 threads
```

```
PS C:\Users\MARIA JOSE\Downloads\UNACH-Semestres\Tercer-Semestre\CulturaDigital\Homeworks\preprocesamiento-cienciadatos> git status
>>
On branch feature-preprocesamiento
Your branch is up to date with 'origin/feature-preprocesamiento'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git restore <file>..." to discard changes in working directory)
        modified:   preprocesamiento.py
```

```
>> git pull origin main
>>
Switched to branch 'main'
Your branch is behind 'origin/main' by 3 commits, and can be fast-forwarded.
  (use "git pull" to update your local branch)
From https://github.com/MariaVimos/preprocesamiento-cienciadatos
 * branch                main          -> FETCH_HEAD
Updating afeb988..3ead267
Fast-forward
 preprocesamiento.py | 32 ++++++
1 file changed, 32 insertions(+)
create mode 100644 preprocesamiento.py
```

## 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](#).

base: main
compare: feature-preprocesamiento

Able to merge. These branches can be automatically merged.

Add a title

Feature preprocesamiento

Add a description

Write
Preview

Codigo de Python para realizar pre procesamiento en un data set

Markdown is supported
Paste, drop, or click to add files

Create pull request

Reviewers
No reviews

Assignees
No one—assign yourself

Labels
None yet

Projects
None yet

Milestone
No milestone


Development
Use [Closing keywords](#) in the description to automatically close issues

Conversation 0

Commits 2


Checks 0


Files changed 1



 **MariaVimos** commented now



Owner ...


Codigo de Python para realizar pre procesamiento en un data set



 **mariavimos19** added 2 commits 15 minutes ago

  preprocesamiento d9cf249

  Agregar script de preprocesamiento de datos b43e1c7

 **No conflicts with base branch**  
Merging can be performed automatically.

Merge pull request

You can also merge this with the command line. [View command line instructions.](#)

Still in progress? [Convert to draft](#)

Merged


**Feature preprocesamiento #1**  
MariaVimos merged 2 commits into `main` from `feature-preprocesa...`  now

  Agregar script de preprocesamiento de datos b43e1c7

  **MariaVimos** merged commit **3ead267** into `main` now

Revert

```
Documentacion.md  ! main.yml x
preprocesamiento-cienciadatos > workflows > ! main.yml
1  name: Preprocesamiento CI
2
3  on:
4    push:
5      branches: [ main ]
6
7  jobs:
8    test:
9      runs-on: ubuntu-latest
10     steps:
11       - uses: actions/checkout@v3
12         name: Set up Python
13       - uses: actions/setup-python@v4
14         with:
15           python-version: '3.10'
16       - name: Install dependencies
17         run: pip install pandas scikit-learn
18       - name: Run preprocesamiento
19         run: python preprocesamiento.py
20
```

 mariavimos19 WorkFlow3 cab6718 · 1 minute ago ⌚

19 lines (17 loc) · 417 Bytes

Code Blame Raw Copy Download Edit View

```
1  name: Preprocesamiento CI
2
3  on:
4    push:
5      branches: [ main ]
6
7  jobs:
8    test:
9      runs-on: ubuntu-latest
10     steps:
11       - uses: actions/checkout@v3
12         name: Set up Python
13       - uses: actions/setup-python@v4
14         with:
15           python-version: '3.10'
16       - name: Install dependencies
17         run: pip install pandas scikit-learn
18       - name: Run preprocesamiento
19         run: python preprocesamiento.py
```

## ENLACE DEL REPOSITORIO:

<https://github.com/MariaVimos/preprocesamiento-cienciadatos.git>

## Bibliografía:

• Chacon, S., & Straub, B. (2014). Pro Git. Apress. • McKinney, W. (2018). Python for Data Analysis. O'Reilly Media.