Clases Interactivas con Google Colab, Mkdocs y Github Actions

Francisco Alfaro

26 de Agosto del 2022

Tabla de Contenidos

Introducción

- Motivación

Herramientas

- Google Colab + Mkdocs + Github

Manos a la Obra

- Caso de Estudio + Ejemplos

Conclusiones

- Resultados

Motivación

- Clases desordenadas
- Replicabilidad
- Simplicidad

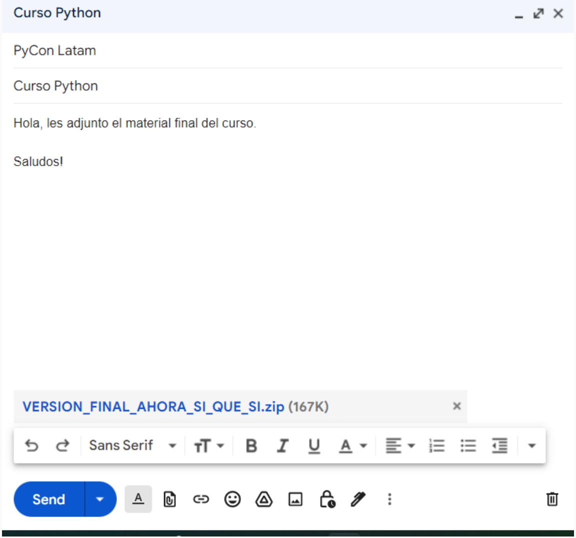
Motivación

- Clases desordenadas
- Replicabilidad
- Simplicidad

Motivación

- Clases desordenadas
- Replicabilidad
- Simplicidad

Problema ✗



Curso Python

PyCon Latam

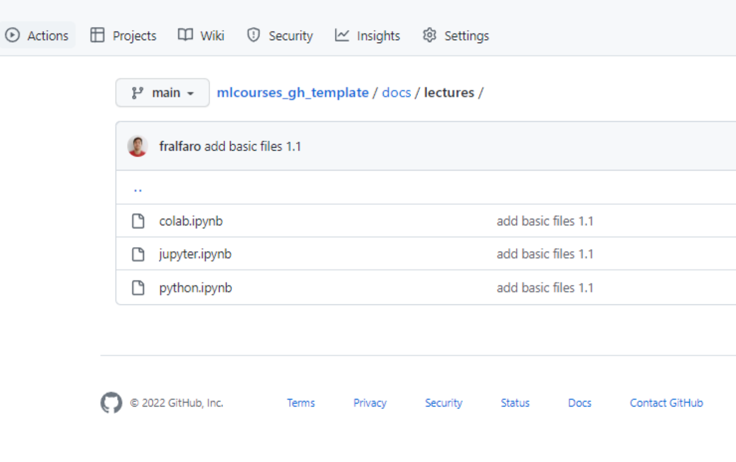
Curso Python

Hola, les adjunto el material final del curso.

Saludos!

VERSION_FINAL_AHORA_SI_QUE_SI.zip (167K)

Problema ✗



main · mlcourses_gh_template / docs / lectures /

fralfaro add basic files 1.1

..

colab.ipynb	add basic files 1.1
jupyter.ipynb	add basic files 1.1
python.ipynb	add basic files 1.1

© 2022 GitHub, Inc.

Solución

Solución: Ocupar de manera creativa Google Colab, Mkdocs y Github Actions.

Tabla de Contenidos

- Introducción
  - Motivación
- Herramientas
  - Google Colab + Mkdocs + Github
- Manos a la Obra
  - Caso de Estudio + Ejemplos
- Conclusiones
  - Resultados

Google Colab

- **Google Colab** permite escribir y ejecutar código de Python en el navegador. Es adecuado para tareas de aprendizaje automático, análisis de datos y educación.
- No requiere configuración y que ofrece acceso sin coste adicional a recursos informáticos, como GPUs.
- Alternativas: SageMaker, DeepNote.

Google Colab

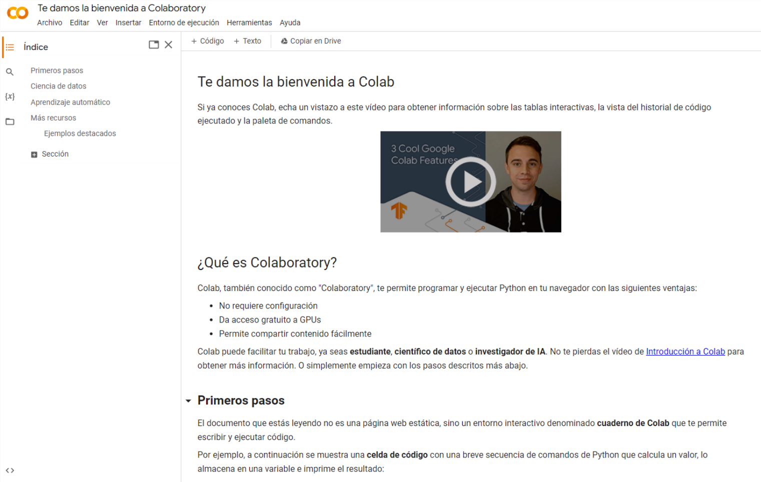
- **Google Colab** permite escribir y ejecutar código de Python en el navegador. Es adecuado para tareas de aprendizaje automático, análisis de datos y educación.
- No requiere configuración y que ofrece acceso sin coste adicional a recursos informáticos, como GPUs.
- Alternativas: SageMaker, DeepNote.

Google Colab

- **Google Colab** permite escribir y ejecutar código de Python en el navegador. Es adecuado para tareas de aprendizaje automático, análisis de datos y educación.
- No requiere configuración y que ofrece acceso sin coste adicional a recursos informáticos, como GPUs.
- Alternativas: SageMaker, DeepNote.

Google Colab

https://colab.research.google.com


Te damos la bienvenida a Colaboratory
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda
Índice
Primeros pasos
Ciencia de datos
Aprendizaje automático
Más recursos
Ejemplos destacados
Sección
+ Código + Texto
Copiar en Drive
3 Cool Google Colab Features

Te damos la bienvenida a Colab

Si ya conoces Colab, echa un vistazo a este vídeo para obtener información sobre las tablas interactivas, la vista del historial de código ejecutado y la paleta de comandos.

¿Qué es Colaboratory?

Colab, también conocido como "Colaboratory", te permite programar y ejecutar Python en tu navegador con las siguientes ventajas:

- No requiere configuración
- Da acceso gratuito a GPUs
- Permite compartir contenido fácilmente

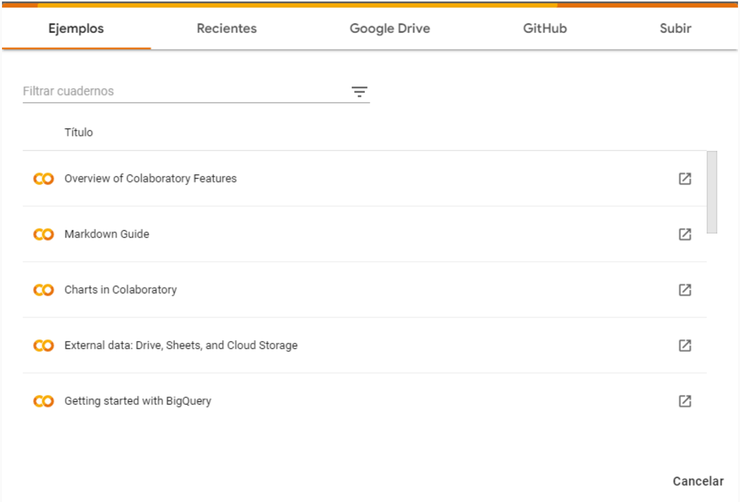
Colab puede facilitar tu trabajo, ya seas **estudiante**, **científico de datos** o **investigador de IA**. No te pierdas el vídeo de Introducción a Colab para obtener más información. O simplemente empieza con los pasos descritos más abajo.

Primeros pasos

El documento que estás leyendo no es una página web estática, sino un entorno interactivo denominado **cuaderno de Colab** que te permite escribir y ejecutar código.

Por ejemplo, a continuación se muestra una **celda de código** con una breve secuencia de comandos de Python que calcula un valor, lo almacena en una variable e imprime el resultado:

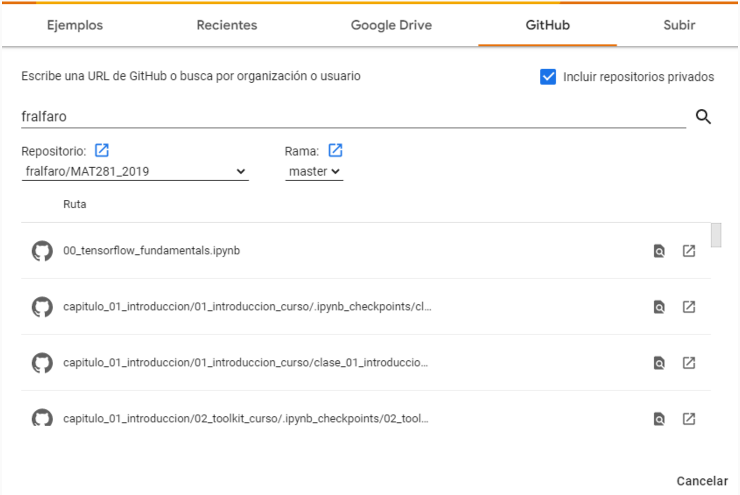
Archivos



Título

- Overview of Colaboratory Features
- Markdown Guide
- Charts in Colaboratory
- External data: Drive, Sheets, and Cloud Storage
- Getting started with BigQuery

Archivos



Escribe una URL de GitHub o busca por organización o usuario

☑ Incluir repositorios privados

fralfaro

Repositorio: fralfaro/MAT281_2019

Rama: master

Ruta

00_tensorflow_fundamentals.ipynb

capitulo_01_introduccion/01_introduccion_curso/.ipynb_checkpoints/cl...

capitulo_01_introduccion/01_introduccion_curso/clase_01_introduccio...

capitulo_01_introduccion/02_toolkit_curso/.ipynb_checkpoints/02_tool...

Mkdocs 📖

- **Mkdocs** es una librería para documentar código en python mediante archivos markdown.
- Mayor popularidad que **Sphinx**.
- Útil para **.ipynb** con: **mkdocs-material** y **mkdocs-jupyter**.
- Alternativas: JB Jupyter-book, Fastpages.

Mkdocs 📖

- **Mkdocs** es una librería para documentar código en python mediante archivos markdown.
- Mayor popularidad que **Sphinx**.
- Útil para **.ipynb** con: **mkdocs-material** y **mkdocs-jupyter**.
- Alternativas: JB Jupyter-book, Fastpages.

Mkdocs 📖

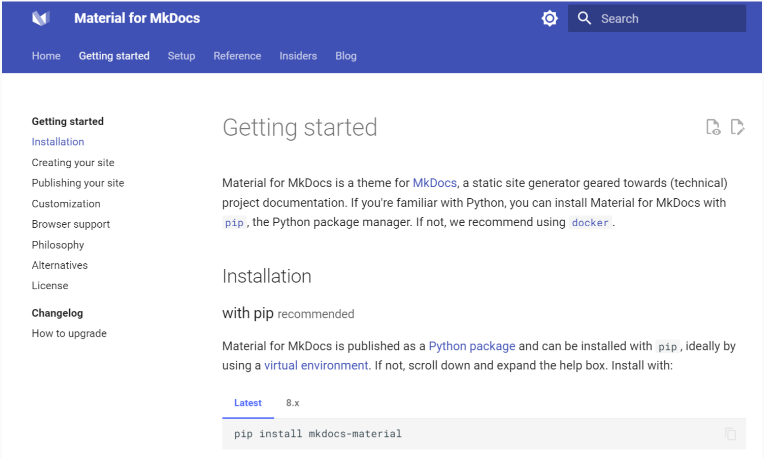
- **Mkdocs** es una librería para documentar código en python mediante archivos markdown.
- Mayor popularidad que **Sphinx**.
- Útil para **.ipynb** con: **mkdocs-material** y **mkdocs-jupyter**.
- Alternativas: Jupyter-book, Fastpages.

Mkdocs 📖

- **Mkdocs** es una librería para documentar código en python mediante archivos markdown.
- Mayor popularidad que **Sphinx**.
- Útil para **.ipynb** con: **mkdocs-material** y **mkdocs-jupyter**.
- Alternativas: JB Jupyter-book, Fastpages.

Mkdocs 📖

https://squidfunk.github.io/mkdocs-material



Getting started

Material for MkDocs is a theme for MkDocs, a static site generator geared towards (technical) project documentation. If you're familiar with Python, you can install Material for MkDocs with `pip`, the Python package manager. If not, we recommend using `docker`.

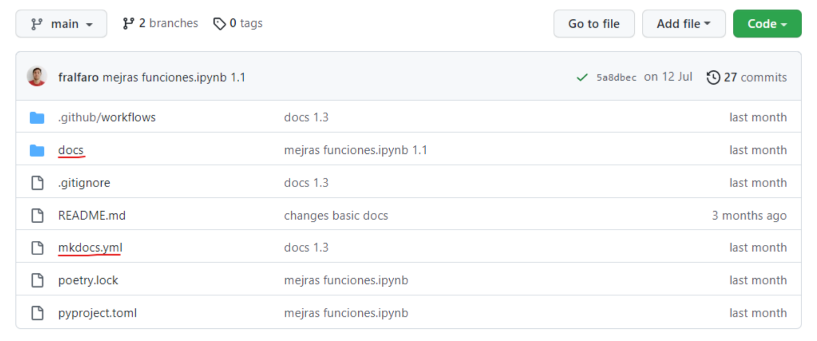
Installation

with pip recommended

Material for MkDocs is published as a Python package and can be installed with `pip`, ideally by using a virtual environment. If not, scroll down and expand the help box. Install with:

```
pip install mkdocs-material
```

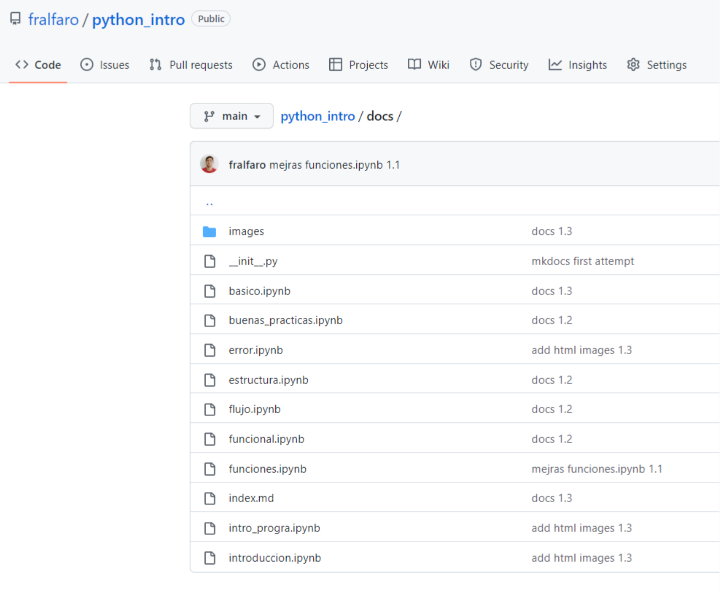
Ejemplo 📖



fralfaro mejras funciones.ipynb 1.1 ✓ 5a8dbec on 12 Jul 27 commits

.github/workflows	docs 1.3	last month
docs	mejras funciones.ipynb 1.1	last month
.gitignore	docs 1.3	last month
README.md	changes basic docs	3 months ago
mkdocs.yml	docs 1.3	last month
poetry.lock	mejras funciones.ipynb	last month
pyproject.toml	mejras funciones.ipynb	last month

docs 📖

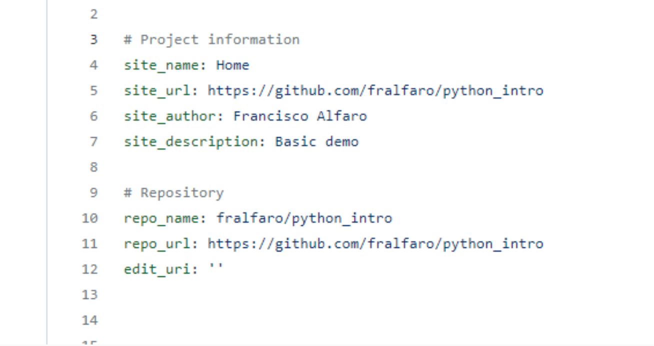


fralfaro mejras funciones.ipynb 1.1

..

images	docs 1.3
__init__.py	mkdocs first attempt
basico.ipynb	docs 1.3
buenas_practicas.ipynb	docs 1.2
error.ipynb	add html images 1.3
estructura.ipynb	docs 1.2
flujo.ipynb	docs 1.2
funcional.ipynb	docs 1.2
funciones.ipynb	mejras funciones.ipynb 1.1
index.md	docs 1.3
intro_progra.ipynb	add html images 1.3
introduccion.ipynb	add html images 1.3

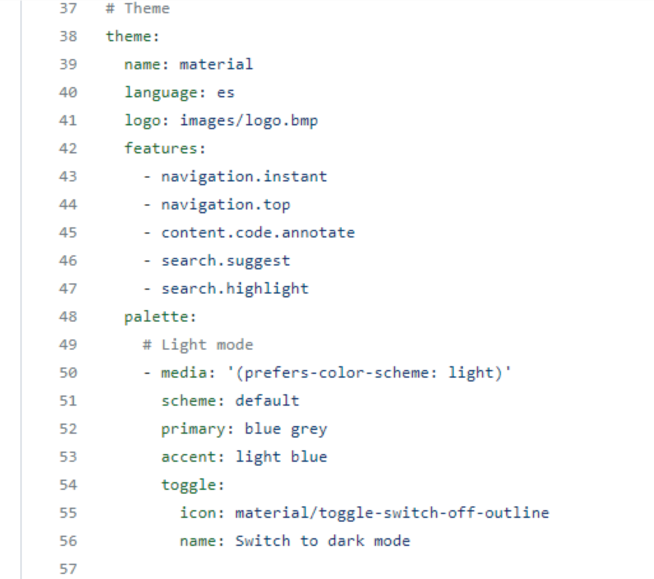
mkdocs.yml 📖



```yaml
# Project information
site_name: Home
site_url: https://github.com/fralfaro/python_intro
site_author: Francisco Alfaro
site_description: Basic demo

# Repository
repo_name: fralfaro/python_intro
repo_url: https://github.com/fralfaro/python_intro
edit_uri: ''
```

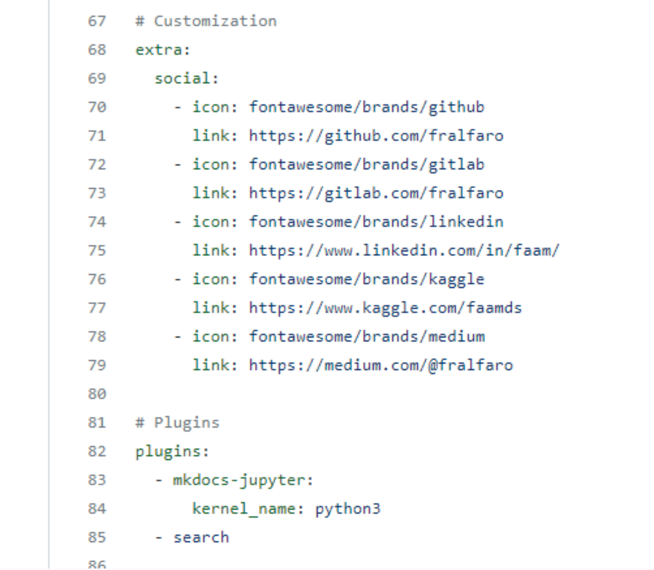
mkdocs.yml 📖



```yaml
# Theme
theme:
  name: material
  language: es
  logo: images/logo.bmp
  features:
    - navigation.instant
    - navigation.top
    - content.code.annotate
    - search.suggest
    - search.highlight
  palette:
    # Light mode
    - media: '(prefers-color-scheme: light)'
      scheme: default
      primary: blue grey
      accent: light blue
      toggle:
        icon: material/toggle-switch-off-outline
        name: Switch to dark mode

```

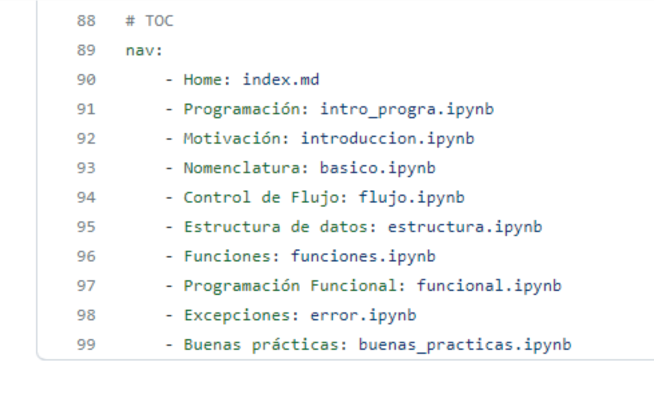
mkdocs.yml 📖



```yaml
# Customization
extra:
  social:
    - icon: fontawesome/brands/github
      link: https://github.com/fralfaro
    - icon: fontawesome/brands/gitlab
      link: https://gitlab.com/fralfaro
    - icon: fontawesome/brands/linkedin
      link: https://www.linkedin.com/in/faam/
    - icon: fontawesome/brands/kaggle
      link: https://www.kaggle.com/faamds
    - icon: fontawesome/brands/medium
      link: https://medium.com/@fralfaro

# Plugins
plugins:
  - mkdocs-jupyter:
      kernel_name: python3
  - search
```

mkdocs.yml 📖

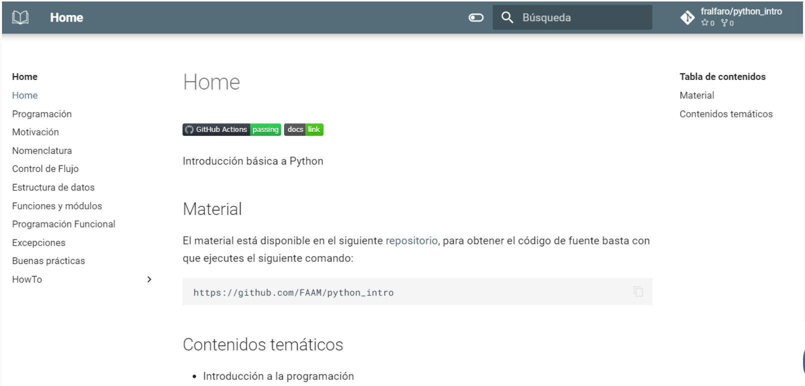


```yaml
# TOC
nav:
    - Home: index.md
    - Programación: intro_progra.ipynb
    - Motivación: introduccion.ipynb
    - Nomenclatura: basico.ipynb
    - Control de Flujo: flujo.ipynb
    - Estructura de datos: estructura.ipynb
    - Funciones: funciones.ipynb
    - Programación Funcional: funcional.ipynb
    - Excepciones: error.ipynb
    - Buenas prácticas: buenas_practicas.ipynb
```

mkdocs.yml 📖

> *mkdocs serve*
>
> *INFO -[12:00:00] Browser connected:*

`http://127.0.0.1:8000/fralfaro/python_intro/`



Home

- Home
- Programación
- Motivación
- Nomenclatura
- Control de Flujo
- Estructura de datos
- Funciones y módulos
- Programación Funcional
- Excepciones
- Buenas prácticas
- HowTo ›

Home

Introducción básica a Python

Material

El material está disponible en el siguiente repositorio, para obtener el código de fuente basta con que ejecutes el siguiente comando:

```
https://github.com/FAAM/python_intro
```

Contenidos temáticos

- Introducción a la programación

Tabla de contenidos

- Material
- Contenidos temáticos

Github

- **Github** es un sitio para albergar código más popular a nivel mundial.
- Recursos gratuitos (pero limitados) para CI/CD mediante Github Actions.
- Nos sirve para ocupar **GITHUB PAGE** -> generar sitios estáticos.
- Alternativas: Gitlab, Bitbucket.

Github

- **Github** es un sitio para albergar código más popular a nivel mundial.
- Recursos gratuitos (pero limitados) para CI/CD mediante Github Actions.
- Nos sirve para ocupar **GITHUB PAGE** -> generar sitios estáticos.
- Alternativas: Gitlab, Bitbucket.

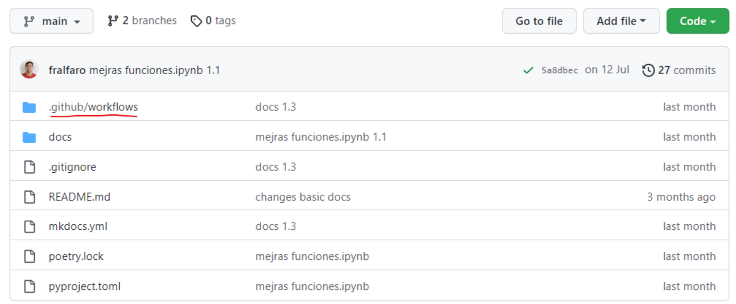
Github

- **Github** es un sitio para albergar código más popular a nivel mundial.
- Recursos gratuitos (pero limitados) para CI/CD mediante Github Actions.
- Nos sirve para ocupar **GITHUB PAGE** -> generar sitios estáticos.
- Alternativas: Gitlab, Bitbucket.

Github

- **Github** es un sitio para albergar código más popular a nivel mundial.
- Recursos gratuitos (pero limitados) para CI/CD mediante Github Actions.
- Nos sirve para ocupar **GITHUB PAGE** -> generar sitios estáticos.
- Alternativas: Gitlab, Bitbucket.

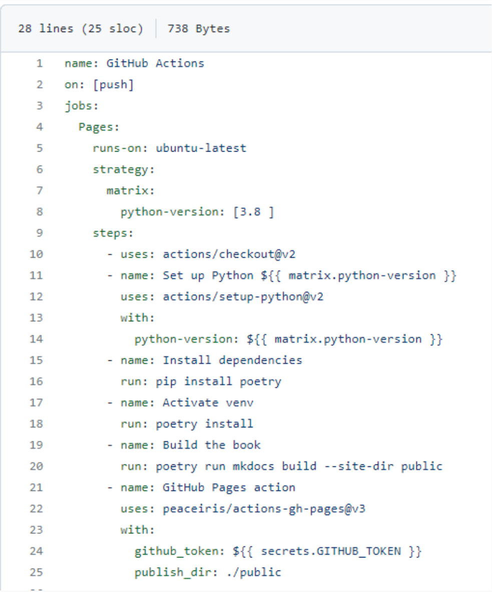
Github



fralfaro mejras funciones.ipynb 1.1 ✓ 5a8dbec on 12 Jul 27 commits

.github/workflows	docs 1.3	last month
docs	mejras funciones.ipynb 1.1	last month
.gitignore	docs 1.3	last month
README.md	changes basic docs	3 months ago
mkdocs.yml	docs 1.3	last month
poetry.lock	mejras funciones.ipynb	last month
pyproject.toml	mejras funciones.ipynb	last month

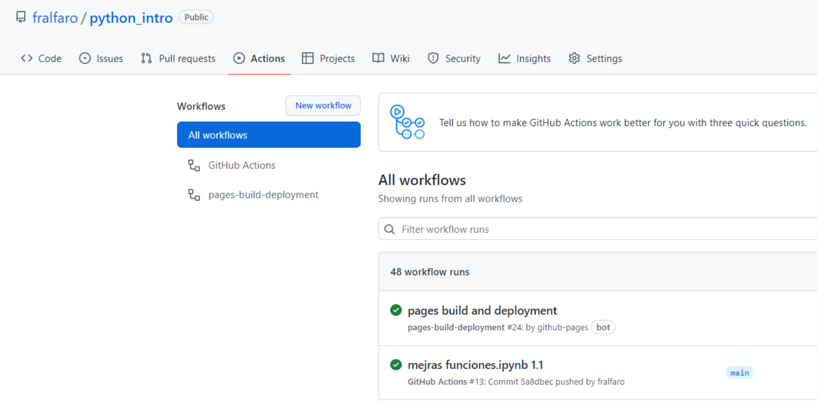
documentation.yml



28 lines (25 sloc) | 738 Bytes

```yaml
name: GitHub Actions
on: [push]
jobs:
  Pages:
    runs-on: ubuntu-latest
    strategy:
      matrix:
        python-version: [3.8 ]
    steps:
      - uses: actions/checkout@v2
      - name: Set up Python ${{ matrix.python-version }}
        uses: actions/setup-python@v2
        with:
          python-version: ${{ matrix.python-version }}
      - name: Install dependencies
        run: pip install poetry
      - name: Activate venv
        run: poetry install
      - name: Build the book
        run: poetry run mkdocs build --site-dir public
      - name: GitHub Pages action
        uses: peaceiris/actions-gh-pages@v3
        with:
          github_token: ${{ secrets.GITHUB_TOKEN }}
          publish_dir: ./public
```

Actions



fralfaro / python_intro Public

Tell us how to make GitHub Actions work better for you with three quick questions.

All workflows

Showing runs from all workflows

Filter workflow runs

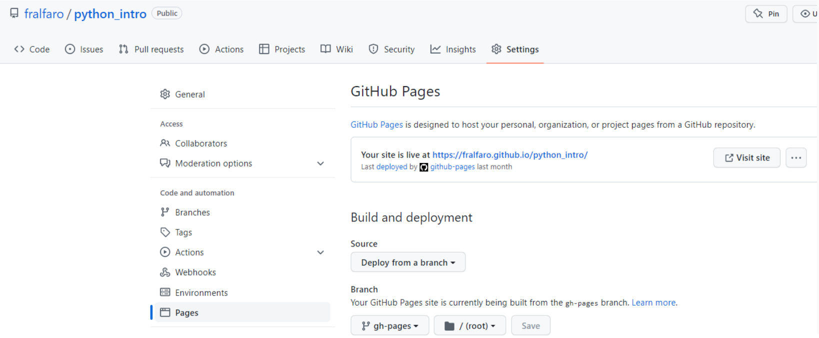
48 workflow runs

✓ pages build and deployment
pages-build-deployment #24: by github-pages bot

✓ mejras funciones.ipynb 1.1
GitHub Actions #13: Commit 5a8dbec pushed by fralfaro
main

Pages

https://fralfaro.github.io/python_intro/



fralfaro / python_intro Public

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Your site is live at https://fralfaro.github.io/python_intro/

Last deployed by github-pages last month

Build and deployment

Source

Deploy from a branch

Branch

Your GitHub Pages site is currently being built from the `gh-pages` branch. Learn more.

gh-pages / (root) Save

Tabla de Contenidos

Introducción

- Motivación

Herramientas

- Google Colab + Mkdocs + Github

Manos a la Obra

- Caso de Estudio + Ejemplos

Conclusiones

- Resultados

Caso de Estudio

Realizar el curso de **Introducción a Python** mediante notebooks.

- Código este en Github
- Documentación ordenada del curso
- Replicabilidad del código.

Ocuparemos el repositorio: fralfaro/python_intro

Caso de Estudio

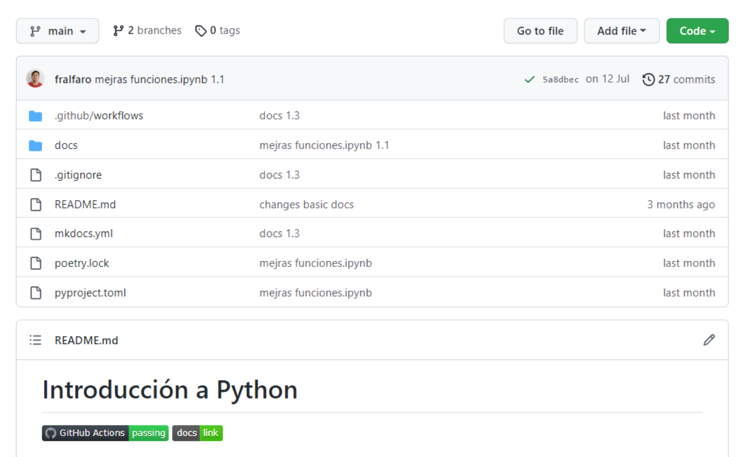
Realizar el curso de **Introducción a Python** mediante notebooks.

- Código este en Github
- Documentación ordenada del curso
- Replicabilidad del código.

Ocuparemos el repositorio: fralfaro/python_intro

Caso de Estudio

`https://github.com/fralfaro/python_intro`



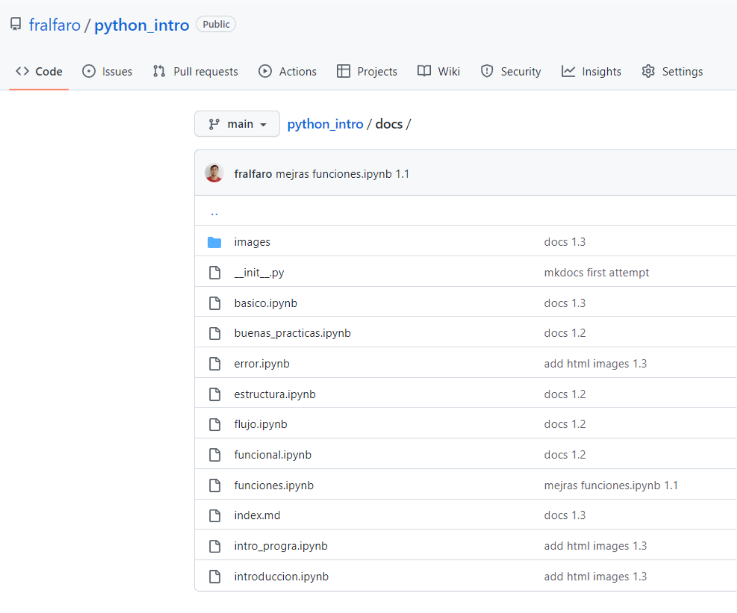
fralfaro mejras funciones.ipynb 1.1 ✓ 5a8dbec on 12 Jul 27 commits

.github/workflows	docs 1.3	last month
docs	mejras funciones.ipynb 1.1	last month
.gitignore	docs 1.3	last month
README.md	changes basic docs	3 months ago
mkdocs.yml	docs 1.3	last month
poetry.lock	mejras funciones.ipynb	last month
pyproject.toml	mejras funciones.ipynb	last month

README.md

Introducción a Python

Caso de Estudio



fralfaro / python_intro Public

python_intro / docs /

fralfaro mejras funciones.ipynb 1.1

..

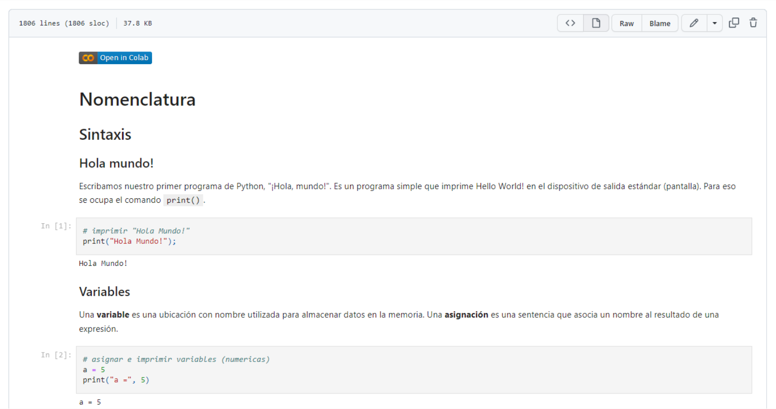
images	docs 1.3
__init__.py	mkdocs first attempt
basico.ipynb	docs 1.3
buenas_practicas.ipynb	docs 1.2
error.ipynb	add html images 1.3
estructura.ipynb	docs 1.2
flujo.ipynb	docs 1.2
funcional.ipynb	docs 1.2
funciones.ipynb	mejras funciones.ipynb 1.1
index.md	docs 1.3
intro_progra.ipynb	add html images 1.3
introduccion.ipynb	add html images 1.3

Combinación

Github + Google Colab: Google Colab está diseñado para integrarse directamente con GitHub (situación que no se puede hacer con Gitlab o Bitbucket).

Combinación

https://github.com/fralfaro/python_intro/blob/main/docs/basico.ipynb



Nomenclatura

Sintaxis

Hola mundo!

Escribamos nuestro primer programa de Python, "¡Hola, mundo!". Es un programa simple que imprime Hello World! en el dispositivo de salida estándar (pantalla). Para eso se ocupa el comando `print()`.

In [1]:

```
# imprimir "Hola Mundo!"
print("Hola Mundo!");
```

```
Hola Mundo!
```

Variables

Una **variable** es una ubicación con nombre utilizada para almacenar datos en la memoria. Una **asignación** es una sentencia que asocia un nombre al resultado de una expresión.

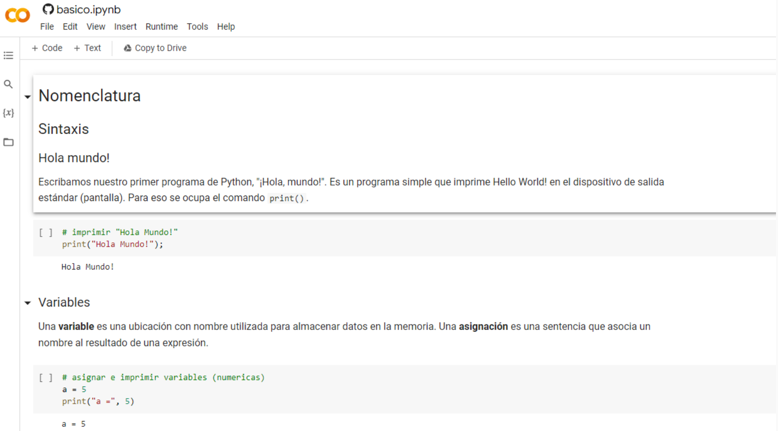
In [2]:

```
# asignar e imprimir variables (numericas)
a = 5
print("a =", 5)
```

```
a = 5
```

Combinación

https://colab.research.google.com/github/fralfaro/python_intro/blob/main/docs/basico.ipynb



Nomenclatura

Sintaxis

Hola mundo!

Escribamos nuestro primer programa de Python, "¡Hola, mundo!". Es un programa simple que imprime Hello World! en el dispositivo de salida estándar (pantalla). Para eso se ocupa el comando `print()`.

```
[ ] # imprimir "Hola Mundo!"
    print("Hola Mundo!");
```

```
Hola Mundo!
```

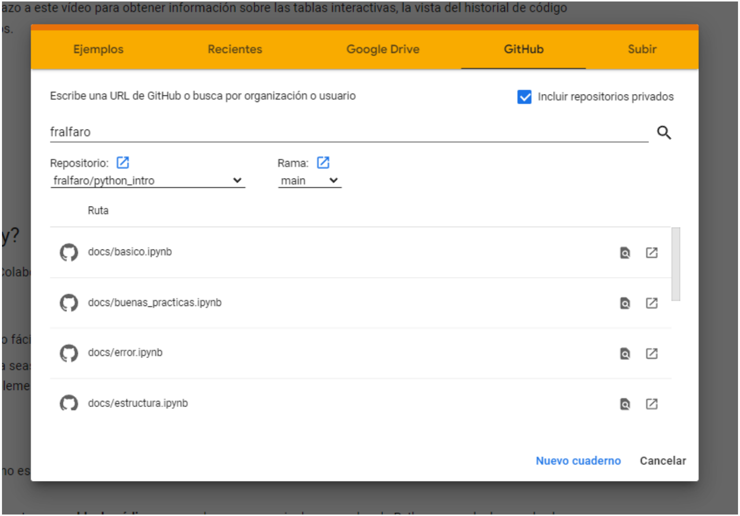
Variables

Una **variable** es una ubicación con nombre utilizada para almacenar datos en la memoria. Una **asignación** es una sentencia que asocia un nombre al resultado de una expresión.

```
[ ] # asignar e imprimir variables (numericas)
    a = 5
    print("a =", 5)
```

```
a = 5
```

Combinación



azo a este vídeo para obtener información sobre las tablas interactivas, la vista del historial de código

s.

y?

Colab

o fác

a sea

leme

no es

Escribe una URL de GitHub o busca por organización o usuario

☑ Incluir repositorios privados

fralfaro

Repositorio: fralfaro/python_intro

Rama: main

Ruta

docs/basico.ipynb

docs/buenas_practicas.ipynb

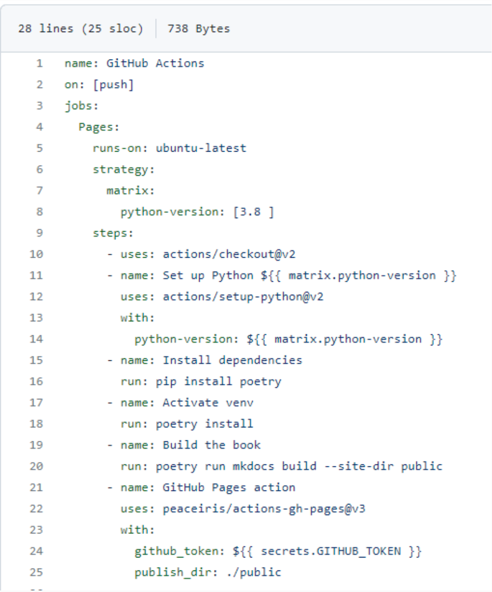
docs/error.ipynb

docs/estructura.ipynb

Combinación

Github + mkdocs: Mediante Github Pages se genera un sitio estático mediante mkdocs, el cual mucho más flexible, customizable y rápido de generar en comparación a utilizar Jupyterbook.

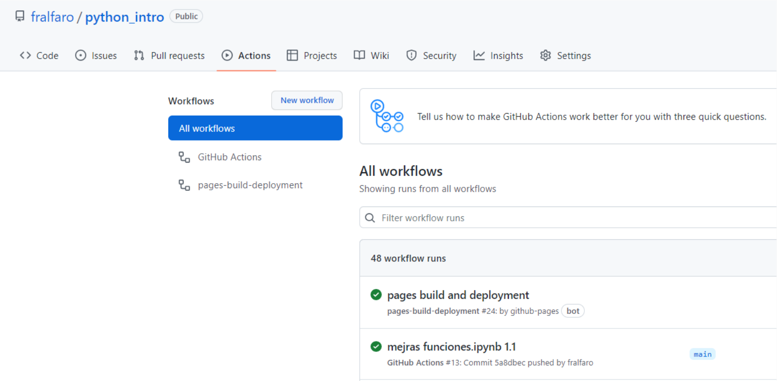
Combinación



28 lines (25 sloc) | 738 Bytes

```yaml
name: GitHub Actions
on: [push]
jobs:
  Pages:
    runs-on: ubuntu-latest
    strategy:
      matrix:
        python-version: [3.8 ]
    steps:
      - uses: actions/checkout@v2
      - name: Set up Python ${{ matrix.python-version }}
        uses: actions/setup-python@v2
        with:
          python-version: ${{ matrix.python-version }}
      - name: Install dependencies
        run: pip install poetry
      - name: Activate venv
        run: poetry install
      - name: Build the book
        run: poetry run mkdocs build --site-dir public
      - name: GitHub Pages action
        uses: peaceiris/actions-gh-pages@v3
        with:
          github_token: ${{ secrets.GITHUB_TOKEN }}
          publish_dir: ./public
```

Combinación



fralfaro / python_intro Public

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Workflows New workflow

All workflows

GitHub Actions

pages-build-deployment

Tell us how to make GitHub Actions work better for you with three quick questions.

All workflows

Showing runs from all workflows

Filter workflow runs

48 workflow runs

✓ pages build and deployment

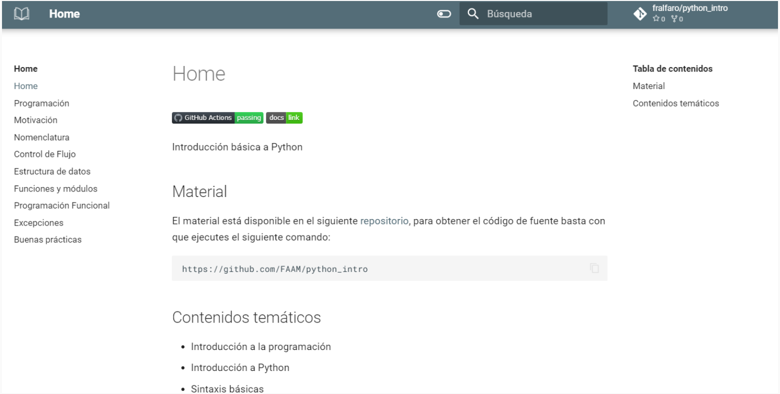
pages-build-deployment #24: by github-pages bot

✓ mejras funciones.ipynb 1.1 main

GitHub Actions #13: Commit 5a8dbec pushed by fralfaro

Combinación

https://fralfaro.github.io/python_intro/



Home

Introducción básica a Python

Material

El material está disponible en el siguiente repositorio, para obtener el código de fuente basta con que ejecutes el siguiente comando:

```
https://github.com/FAAM/python_intro
```

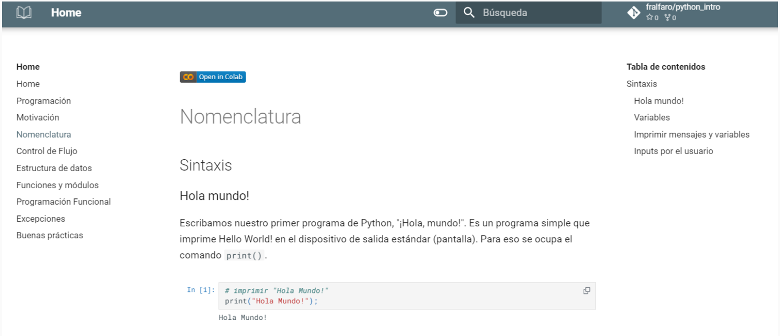
Contenidos temáticos

- Introducción a la programación
- Introducción a Python
- Sintaxis básicas

Combinación

Github + Google Colab + mkdocs: Poder generar un sitio estático flexible y customizable para alberargar jupyter notebooks replicables en Google Colab.

Combinación



Nomenclatura

Sintaxis

Hola mundo!

Escribamos nuestro primer programa de Python, "¡Hola, mundo!". Es un programa simple que imprime Hello World! en el dispositivo de salida estándar (pantalla). Para eso se ocupa el comando `print()`.

```
In [1]:  # imprimir "Hola Mundo!"
         print("Hola Mundo!");
```

```
Hola Mundo!
```

Ejemplos

Algunos Proyectos importantes.

- python4ds-book
  `github.com/fralfaro/PythonDataScienceHandbook`
- r4ds-book
  `github.com/fralfaro/r4ds-book`

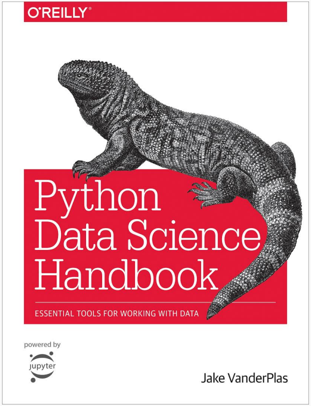
python4ds-book

Pros

- Documentación con Google Colab

Contras

- Documentación desprolija.
- No renderiza panda dataframes.
- Documentación en otra rama.



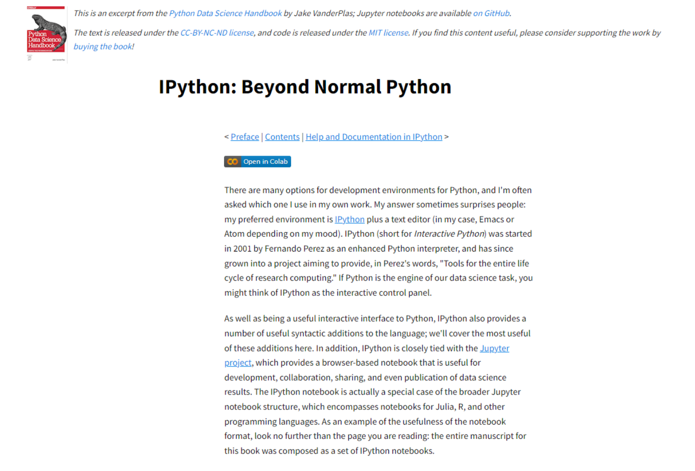
O'REILLY

Python Data Science Handbook

ESSENTIAL TOOLS FOR WORKING WITH DATA

powered by jupyter

Jake VanderPlas



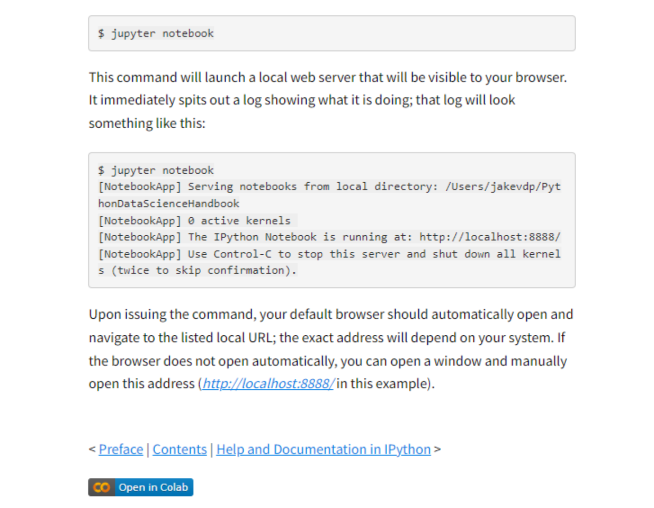
This is an excerpt from the Python Data Science Handbook by Jake VanderPlas; Jupyter notebooks are available on GitHub.

IPython: Beyond Normal Python

< Preface | Contents | Help and Documentation in IPython >

There are many options for development environments for Python, and I'm often asked which one I use in my own work. My answer sometimes surprises people: my preferred environment is IPython plus a text editor (in my case, Emacs or Atom depending on my mood). IPython (short for *Interactive Python*) was started in 2001 by Fernando Perez as an enhanced Python interpreter, and has since grown into a project aiming to provide, in Perez's words, "Tools for the entire life cycle of research computing." If Python is the engine of our data science task, you might think of IPython as the interactive control panel.

As well as being a useful interactive interface to Python, IPython also provides a number of useful syntactic additions to the language; we'll cover the most useful of these additions here. In addition, IPython is closely tied with the Jupyter project, which provides a browser-based notebook that is useful for development, collaboration, sharing, and even publication of data science results. The IPython notebook is actually a special case of the broader Jupyter notebook structure, which encompasses notebooks for Julia, R, and other programming languages. As an example of the usefulness of the notebook format, look no further than the page you are reading: the entire manuscript for this book was composed as a set of IPython notebooks.



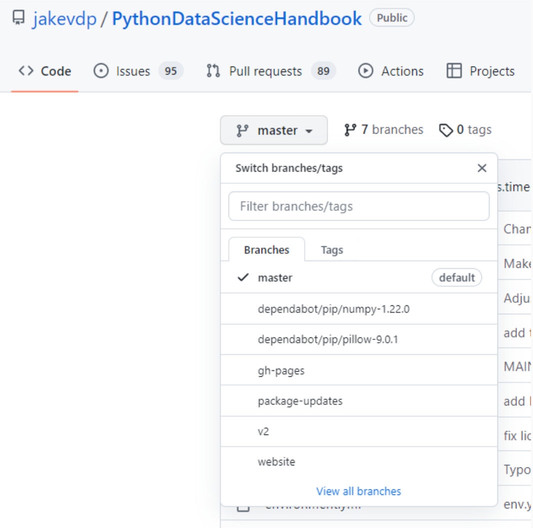
```
$ jupyter notebook
```

This command will launch a local web server that will be visible to your browser. It immediately spits out a log showing what it is doing; that log will look something like this:

```
$ jupyter notebook
[NotebookApp] Serving notebooks from local directory: /Users/jakevdp/Pyt
honDataScienceHandbook
[NotebookApp] 0 active kernels
[NotebookApp] The IPython Notebook is running at: http://localhost:8888/
[NotebookApp] Use Control-C to stop this server and shut down all kernel
s (twice to skip confirmation).
```

Upon issuing the command, your default browser should automatically open and navigate to the listed local URL; the exact address will depend on your system. If the browser does not open automatically, you can open a window and manually open this address (*http://localhost:8888/* in this example).

< Preface | Contents | Help and Documentation in IPython >



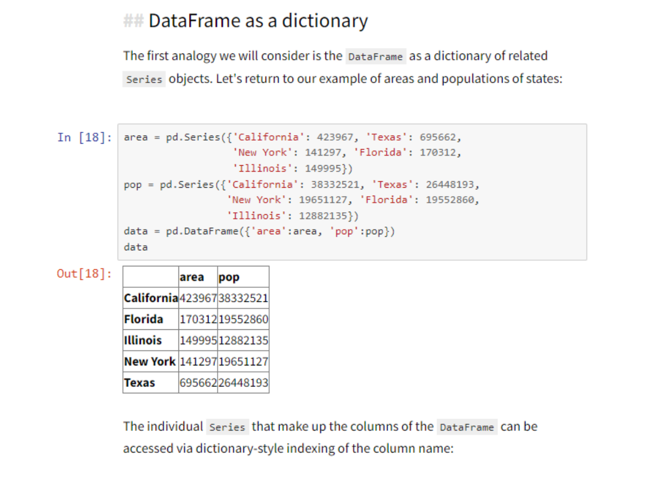
jakevdp / PythonDataScienceHandbook Public

Switch branches/tags

Filter branches/tags

- ✓ master default
- dependabot/pip/numpy-1.22.0
- dependabot/pip/pillow-9.0.1
- gh-pages
- package-updates
- v2
- website

View all branches



DataFrame as a dictionary

The first analogy we will consider is the `DataFrame` as a dictionary of related `Series` objects. Let's return to our example of areas and populations of states:

In [18]:

```
area = pd.Series({'California': 423967, 'Texas': 695662,
                  'New York': 141297, 'Florida': 170312,
                  'Illinois': 149995})
pop = pd.Series({'California': 38332521, 'Texas': 26448193,
                 'New York': 19651127, 'Florida': 19552860,
                 'Illinois': 12882135})
data = pd.DataFrame({'area':area, 'pop':pop})
data
```

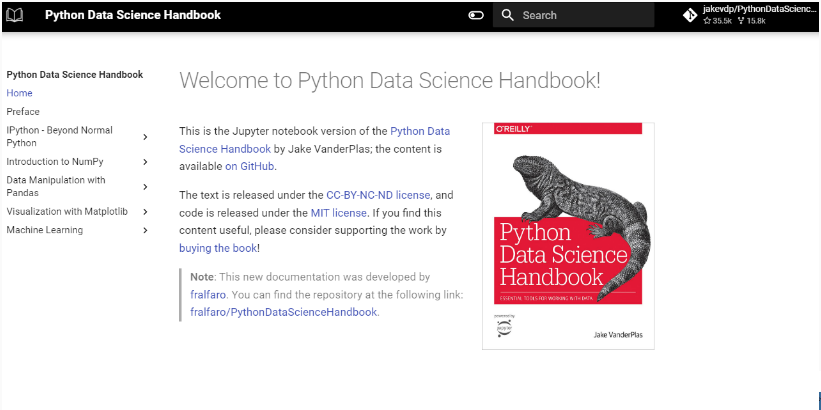
Out[18]:

	area	pop
California	423967	38332521
Florida	170312	19552860
Illinois	149995	12882135
New York	141297	19651127
Texas	695662	26448193

The individual `Series` that make up the columns of the `DataFrame` can be accessed via dictionary-style indexing of the column name:

python4ds-book

fralfaro.github.io/PythonDataScienceHandbook/

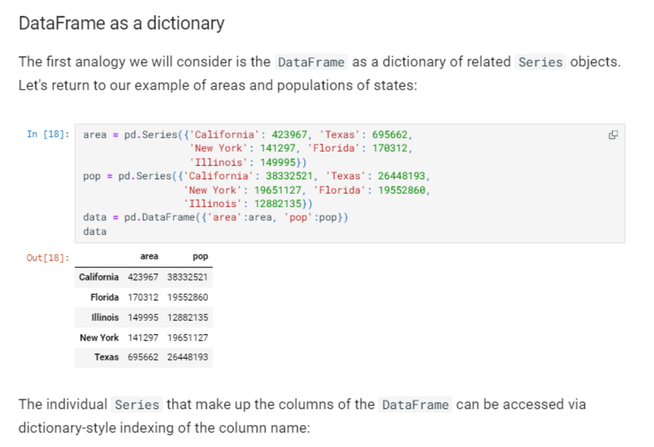

Python Data Science Handbook
Search
jakevdp/PythonDataScienc...
35.5k
15.8k
Python Data Science Handbook
Home
Preface
IPython - Beyond Normal Python
Introduction to NumPy
Data Manipulation with Pandas
Visualization with Matplotlib
Machine Learning
O'REILLY
Python Data Science Handbook
ESSENTIAL TOOLS FOR WORKING WITH DATA
powered by
jupyter
Jake VanderPlas

Welcome to Python Data Science Handbook!

This is the Jupyter notebook version of the Python Data Science Handbook by Jake VanderPlas; the content is available on GitHub.

The text is released under the CC-BY-NC-ND license, and code is released under the MIT license. If you find this content useful, please consider supporting the work by buying the book!

> **Note**: This new documentation was developed by fralfaro. You can find the repository at the following link: fralfaro/PythonDataScienceHandbook.



DataFrame as a dictionary

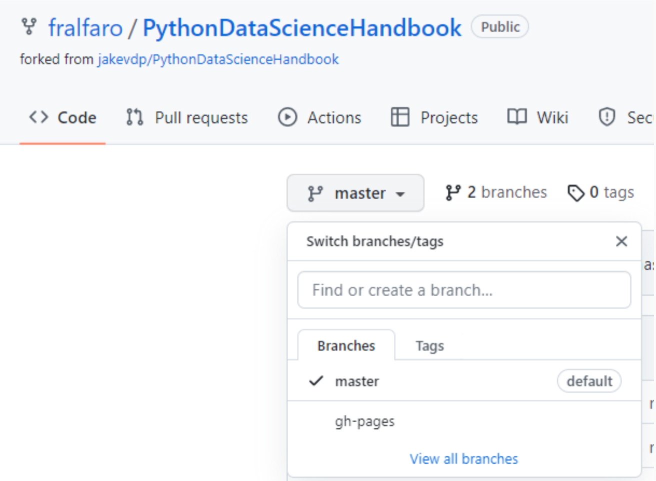
The first analogy we will consider is the `DataFrame` as a dictionary of related `Series` objects. Let's return to our example of areas and populations of states:

In [18]:
```
area = pd.Series({'California': 423967, 'Texas': 695662,
                  'New York': 141297, 'Florida': 170312,
                  'Illinois': 149995})
pop = pd.Series({'California': 38332521, 'Texas': 26448193,
                 'New York': 19651127, 'Florida': 19552860,
                 'Illinois': 12882135})
data = pd.DataFrame({'area':area, 'pop':pop})
data
```

Out[18]:

	area	pop
California	423967	38332521
Florida	170312	19552860
Illinois	149995	12882135
New York	141297	19651127
Texas	695662	26448193

The individual `Series` that make up the columns of the `DataFrame` can be accessed via dictionary-style indexing of the column name:



fralfaro / PythonDataScienceHandbook Public

forked from jakevdp/PythonDataScienceHandbook

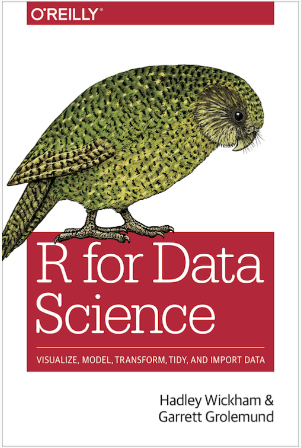
r4ds-book

Pros

- Docuementación con Rmarkdown.
- Todo en una rama.

Contras

- No hay replicabilidad google colab.
- No se imprimen dataframe output.



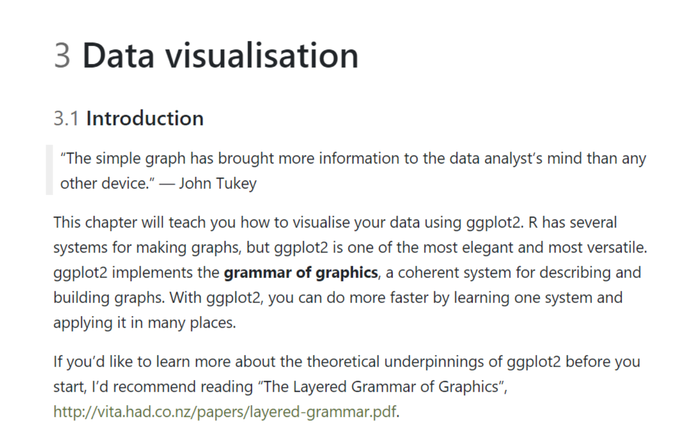
O'REILLY

R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham &
Garrett Grolemund

r4ds-book



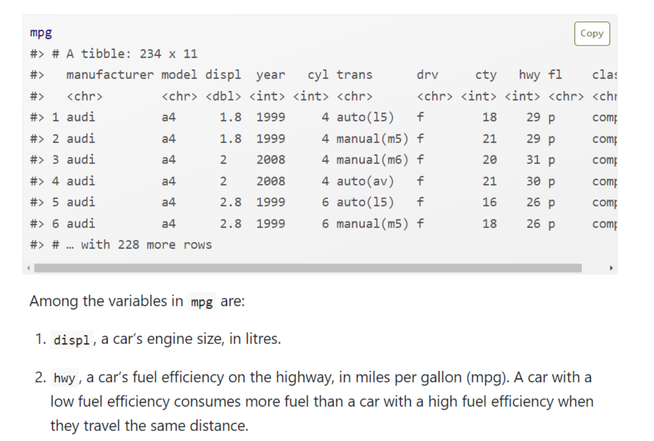
3 Data visualisation

3.1 Introduction

> “The simple graph has brought more information to the data analyst’s mind than any other device.” — John Tukey

This chapter will teach you how to visualise your data using ggplot2. R has several systems for making graphs, but ggplot2 is one of the most elegant and most versatile. ggplot2 implements the **grammar of graphics**, a coherent system for describing and building graphs. With ggplot2, you can do more faster by learning one system and applying it in many places.

If you’d like to learn more about the theoretical underpinnings of ggplot2 before you start, I’d recommend reading “The Layered Grammar of Graphics”, http://vita.had.co.nz/papers/layered-grammar.pdf.



Copy

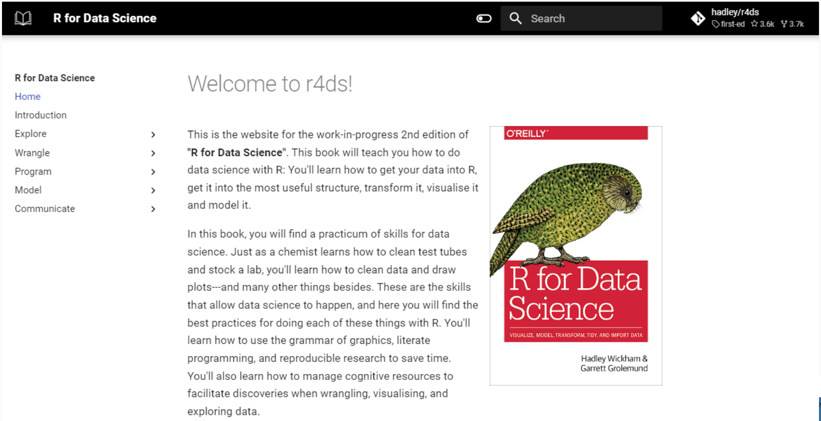
```
mpg
#> # A tibble: 234 x 11
#>   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    clas
#>   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr
#> 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     comp
#> 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     comp
#> 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     comp
#> 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     comp
#> 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     comp
#> 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     comp
#> # … with 228 more rows
```

Among the variables in `mpg` are:

1. `displ`, a car's engine size, in litres.
2. `hwy`, a car's fuel efficiency on the highway, in miles per gallon (mpg). A car with a low fuel efficiency consumes more fuel than a car with a high fuel efficiency when they travel the same distance.

r4ds-book

fralfaro.github.io/r4ds-book/



Welcome to r4ds!

This is the website for the work-in-progress 2nd edition of "**R for Data Science**". This book will teach you how to do data science with R: You'll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it.

In this book, you will find a practicum of skills for data science. Just as a chemist learns how to clean test tubes and stock a lab, you'll learn how to clean data and draw plots---and many other things besides. These are the skills that allow data science to happen, and here you will find the best practices for doing each of these things with R. You'll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. You'll also learn how to manage cognitive resources to facilitate discoveries when wrangling, visualising, and exploring data.

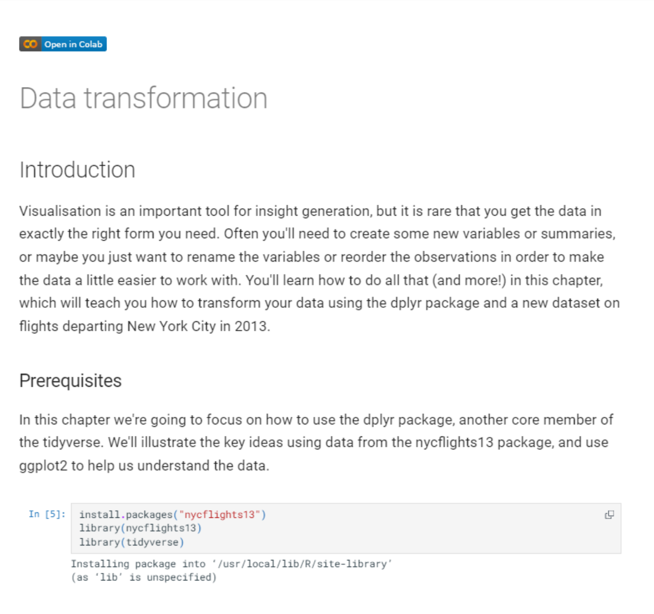
O'REILLY

R for Data Science

VISUALIZE, MODEL, TRANSFORM, TIDY, AND IMPORT DATA

Hadley Wickham & Garrett Grolemund

r4ds-book



Data transformation

Introduction

Visualisation is an important tool for insight generation, but it is rare that you get the data in exactly the right form you need. Often you'll need to create some new variables or summaries, or maybe you just want to rename the variables or reorder the observations in order to make the data a little easier to work with. You'll learn how to do all that (and more!) in this chapter, which will teach you how to transform your data using the dplyr package and a new dataset on flights departing New York City in 2013.

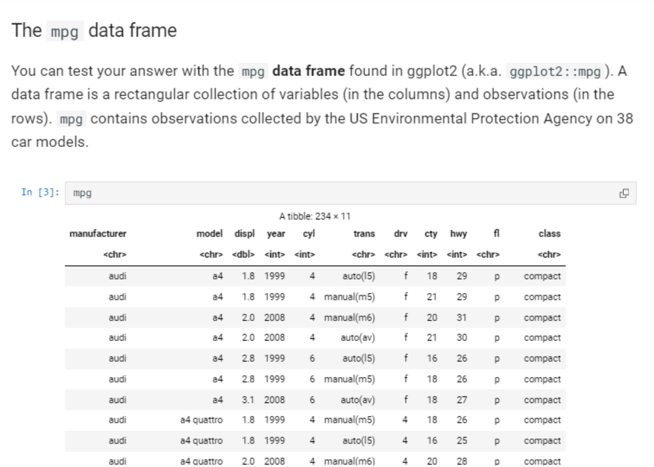
Prerequisites

In this chapter we're going to focus on how to use the dplyr package, another core member of the tidyverse. We'll illustrate the key ideas using data from the nycflights13 package, and use ggplot2 to help us understand the data.

In [5]:

```
install.packages("nycflights13")
library(nycflights13)
library(tidyverse)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```



The `mpg` data frame

You can test your answer with the `mpg` **data frame** found in ggplot2 (a.k.a. `ggplot2::mpg`). A data frame is a rectangular collection of variables (in the columns) and observations (in the rows). `mpg` contains observations collected by the US Environmental Protection Agency on 38 car models.

```
In [3]: mpg
```

A tibble: 234 × 11

manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
<chr>	<chr>	<dbl>	<int>	<int>	<chr>	<chr>	<int>	<int>	<chr>	<chr>
audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact

Tabla de Contenidos

Introducción

- Motivación

Herramientas

- Google Colab + Mkdocs + Github

Manos a la Obra

- Caso de Estudio + Ejemplos

Conclusiones

- Resultados

Resultados

- Versatilidad de Documentación.
- Replicabilidad del código.
- Relativamente sencillo de trabajar.

Resultados

- Versatilidad de Documentación.
- Replicabilidad del código.
- Relativamente sencillo de trabajar.

Resultados

- Versatilidad de Documentación.
- Replicabilidad del código.
- Relativamente sencillo de trabajar.

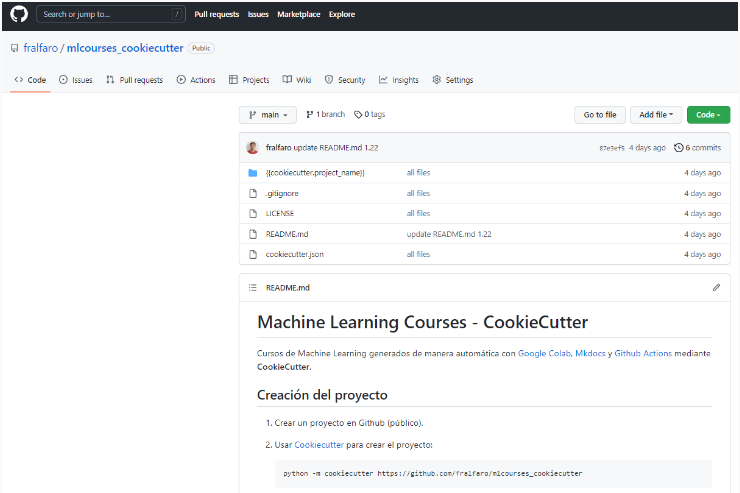
Comienza tus propios cursos!

Se dejan dos plantillas para que puedan prácticar con sus propios notebooks.

- Cookiecutter
  `github.com/fralfaro/mlcourses_cookiecutter`
- Github Template Use this template
  `github.com/fralfaro/mlcourses_gh_template`

Cookiecutter

github.com/fralfaro/mlcourses_cookiecutter



fralfaro update README.md 1.22 87e3ef5 4 days ago 6 commits

{{cookiecutter.project_name}}	all files	4 days ago
.gitignore	all files	4 days ago
LICENSE	all files	4 days ago
README.md	update README.md 1.22	4 days ago
cookiecutter.json	all files	4 days ago

README.md

Machine Learning Courses - CookieCutter

Cursos de Machine Learning generados de manera automática con Google Colab, Mkdocs y Github Actions mediante **CookieCutter**.

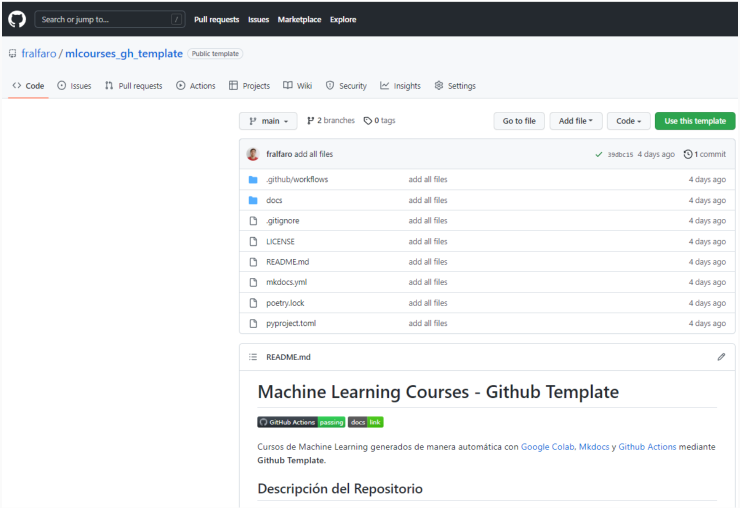
Creación del proyecto

1. Crear un proyecto en Github (público).
2. Usar Cookiecutter para crear el proyecto:

```
python -m cookiecutter https://github.com/fralfaro/mlcourses_cookiecutter
```

Github Template Use this template

github.com/fralfaro/mlcourses_gh_template



fralfaro / mlcourses_gh_template Public template

2 branches 0 tags

fralfaro add all files ✓ 39dbc15 4 days ago 1 commit

.github/workflows	add all files	4 days ago
docs	add all files	4 days ago
.gitignore	add all files	4 days ago
LICENSE	add all files	4 days ago
README.md	add all files	4 days ago
mkdocs.yml	add all files	4 days ago
poetry.lock	add all files	4 days ago
pyproject.toml	add all files	4 days ago

README.md

Machine Learning Courses - Github Template

GitHub Actions passing docs link

Cursos de Machine Learning generados de manera automática con Google Colab, Mkdocs y Github Actions mediante **Github Template**.

Descripción del Repositorio

Clases Interactivas con Google Colab, Mkdocs y Github Actions

Francisco Alfaro

26 de Agosto del 2022