NDP Data Challenges

Onboarding Tutorial

Log in / Register

NATIONAL DATA PLATFORM
v1.0.0
Release Notes

Catalog About Events and Press

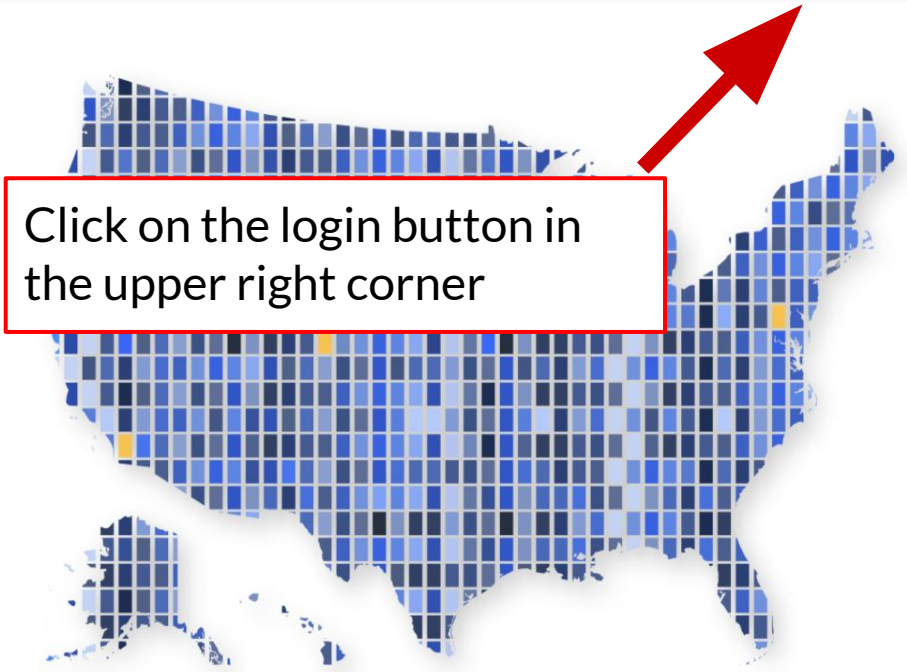
Get Started with NDP | Log in/Register

Open Data, Equitable Access and AI Services for All

Building the nation's federated data ecosystem.
Explore data. Run analyses. Transform AI education.

Explore our catalog of datasets >

SAN DIEGO SUPERCOMPUTER CENTER
SCI www.sci.utah.edu
EarthScope Consortium
UC San Diego
THE UNIVERSITY OF UTAH
University of Colorado Boulder



Click on the login button in the upper right corner

5491
data collections and livestreams

5
data and AI services

348
registered users

Click on CILogon

Contact - ndp@sdsc.edu
The National Data Platform was funded by NSF 2335609 under CI, CISE Research Resources programs. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

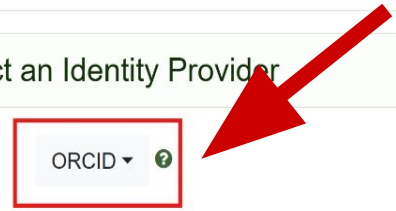
CILogon

Consent to Attribute Release

NSF National Data Platform (NDP) requests access to the following information. If you do not app

- Your CILogon user identifier

Click on the dropdown menu



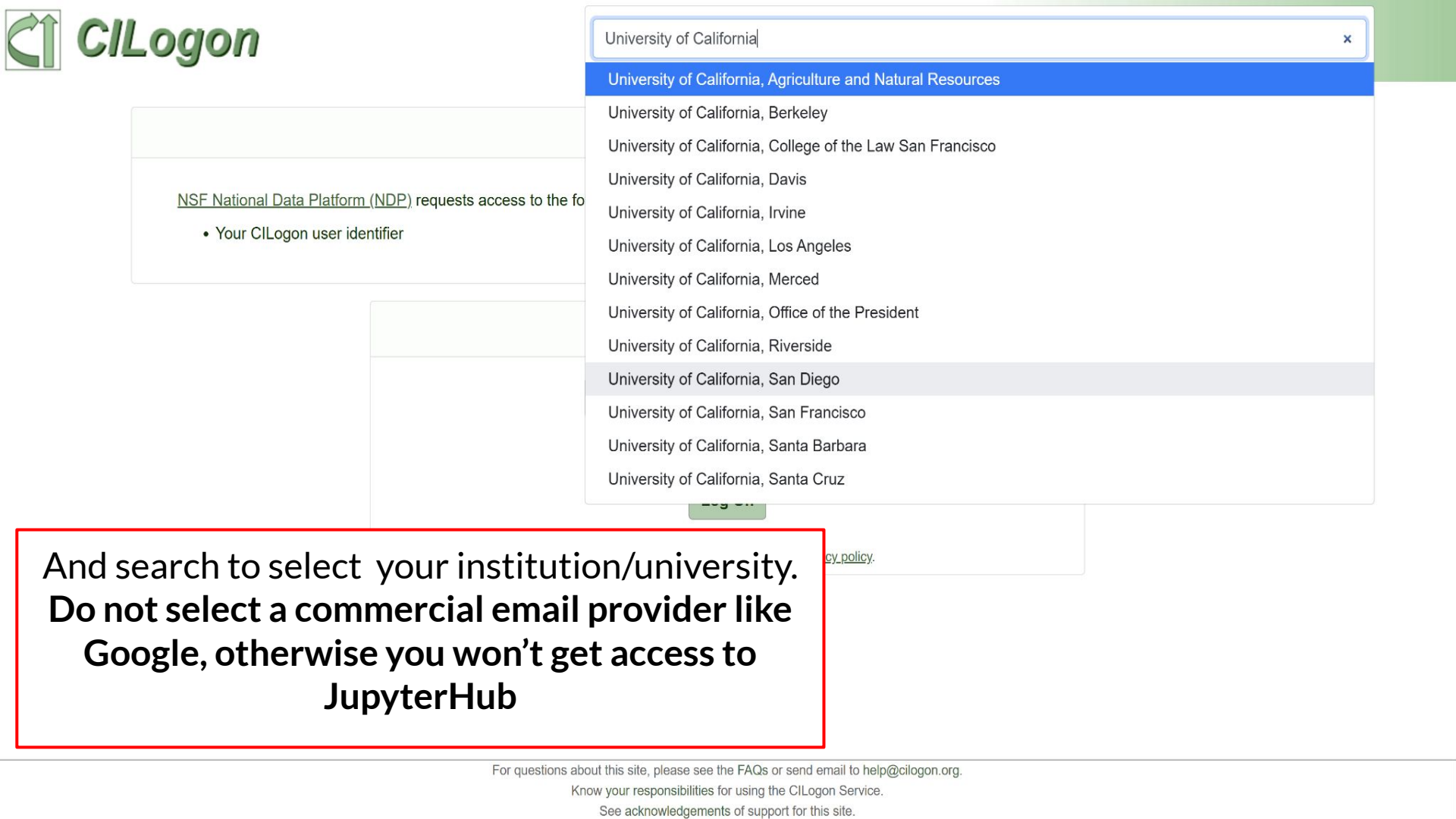
Select an Identity Provider

ORCID

Remember this selection

Log On

By selecting "Log On", you agree to the privacy policy.



And search to select your institution/university.
Do not select a commercial email provider like Google, otherwise you won't get access to JupyterHub

CILogon

Consent to Attribute Release

NSF National Data Platform (NDP) requests access to the following information. If you do not appr

- Your CILogon user identifier

Select an Identity Provid

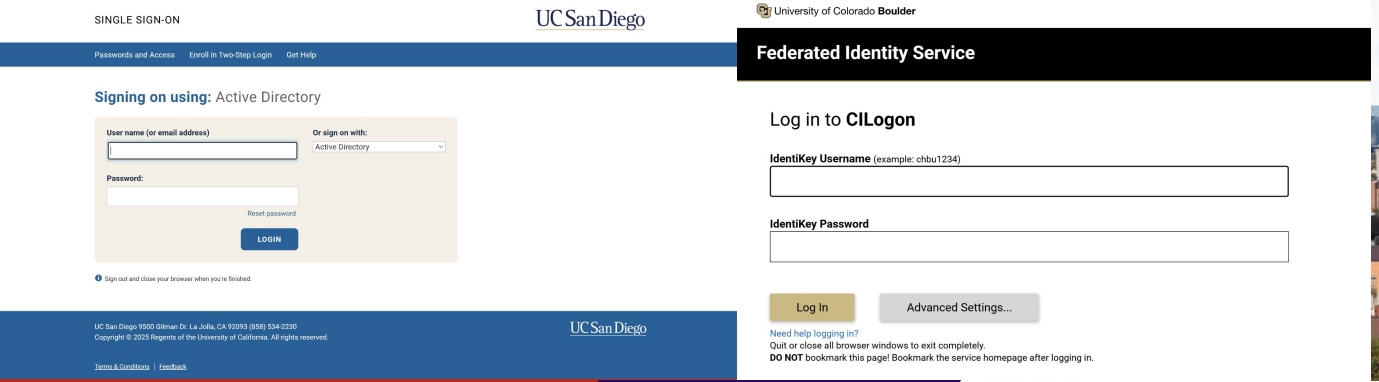
University of California, San Diego

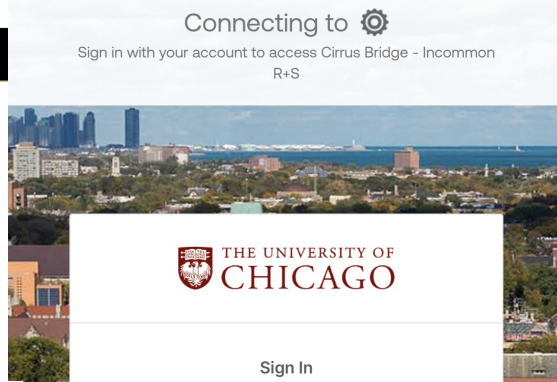
Remember this selection

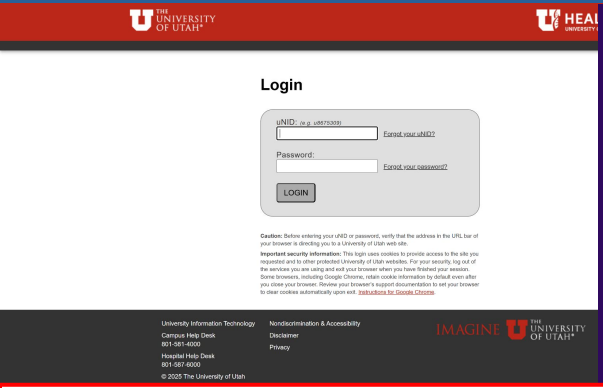
Log On

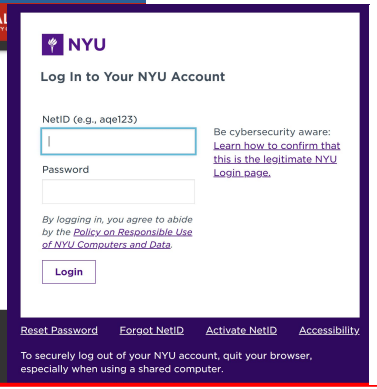
By selecting "Log On", you agree to the privacy policy.

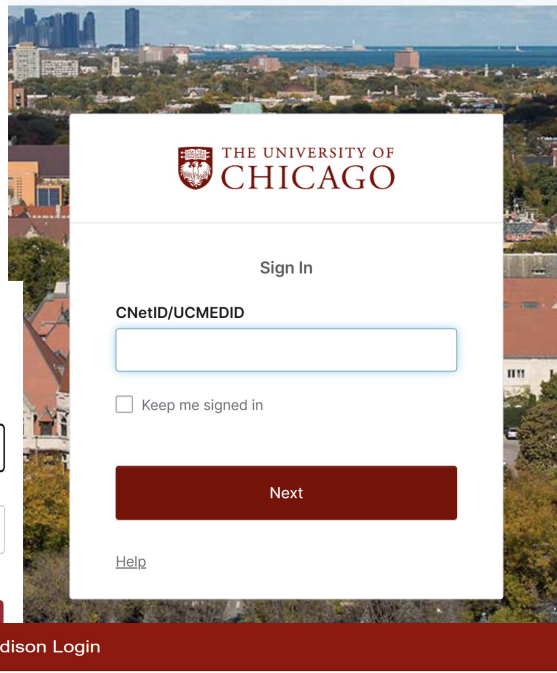
Click on *Remember this selection* so the system remembers your institution the next time you login, followed by *Log On*

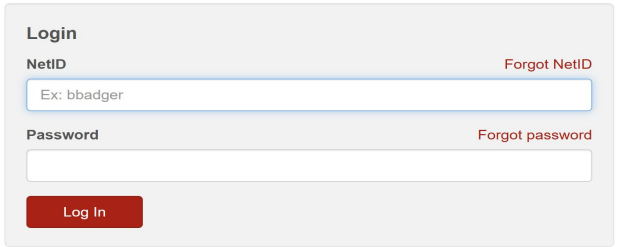












After selecting your university, you will be prompted to **your university's login system**. Login with your institutional credentials.

NOTE

If this is your first time logging in to NDP, you will be asked to confirm your name and your email as part of your registration.

Discover the Data Challenge

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard

My Uploads

Workspace

Catalogs

Education Hub

Educator Portal

Learner Portal

Explore

Open Learning

POPs

My Dashboard

2

PV

Welcome back, Pedro Antonio Ramonetti Vega!

LATEST ANNOUNCEMENTS

January 22-23, 2024	2-day all hands team meeting at the San Diego Supercomputer Center!
January 19, 2024	More details ironed out. Alpha version of NDP is live!
January 12, 2024	User dashboard drafted. JupyterHub integrated.
January 5, 2024	String search integrated.

Quick Explore

Upload

JupyterHub

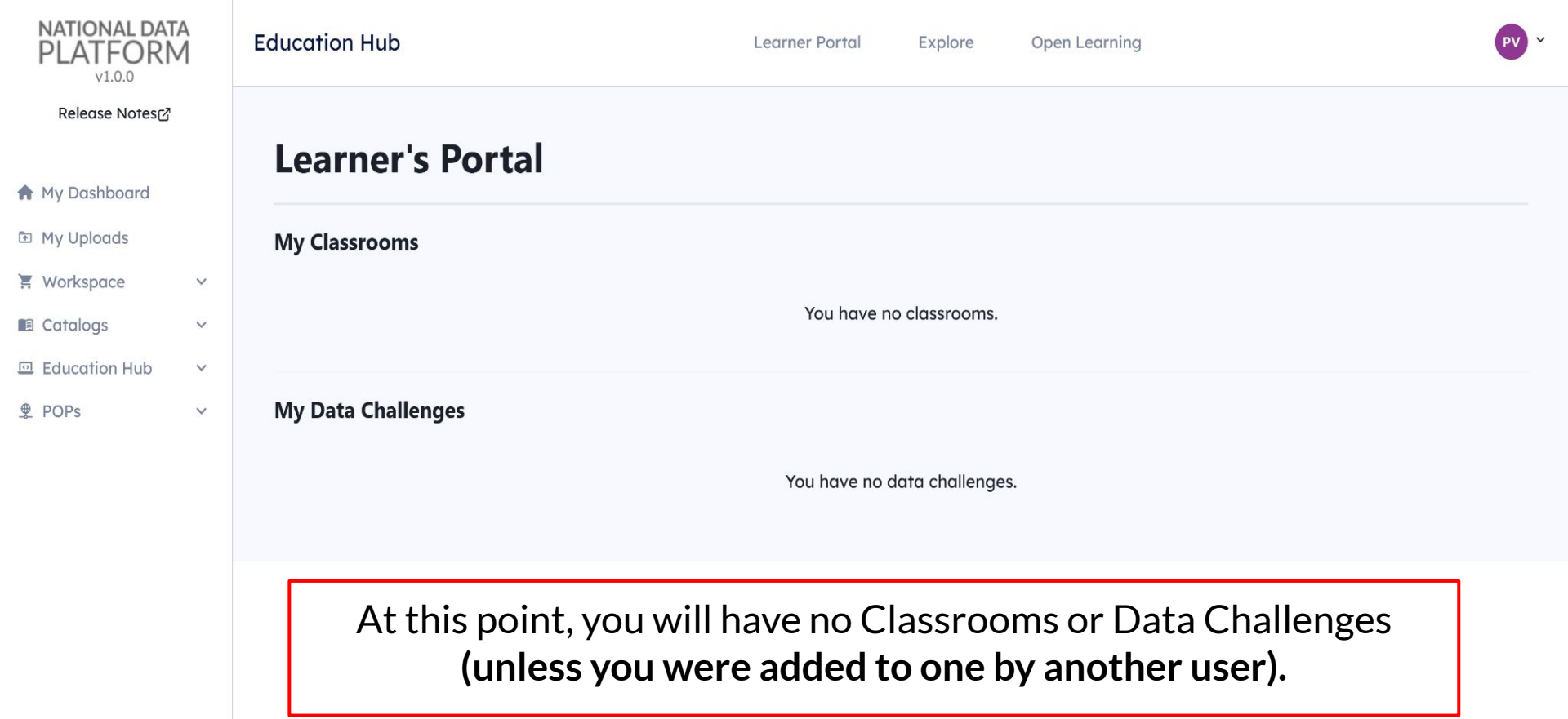
Education Gateway

Click on Education Hub, followed by Learner Portal

Your Profile

PV

Pedro Antonio Ramonetti Vega
pramonettivega@ucsd.edu



At this point, you will have no Classrooms or Data Challenges **(unless you were added to one by another user).**

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Education Hub

Learner Portal | Explore | Open Learning

PV

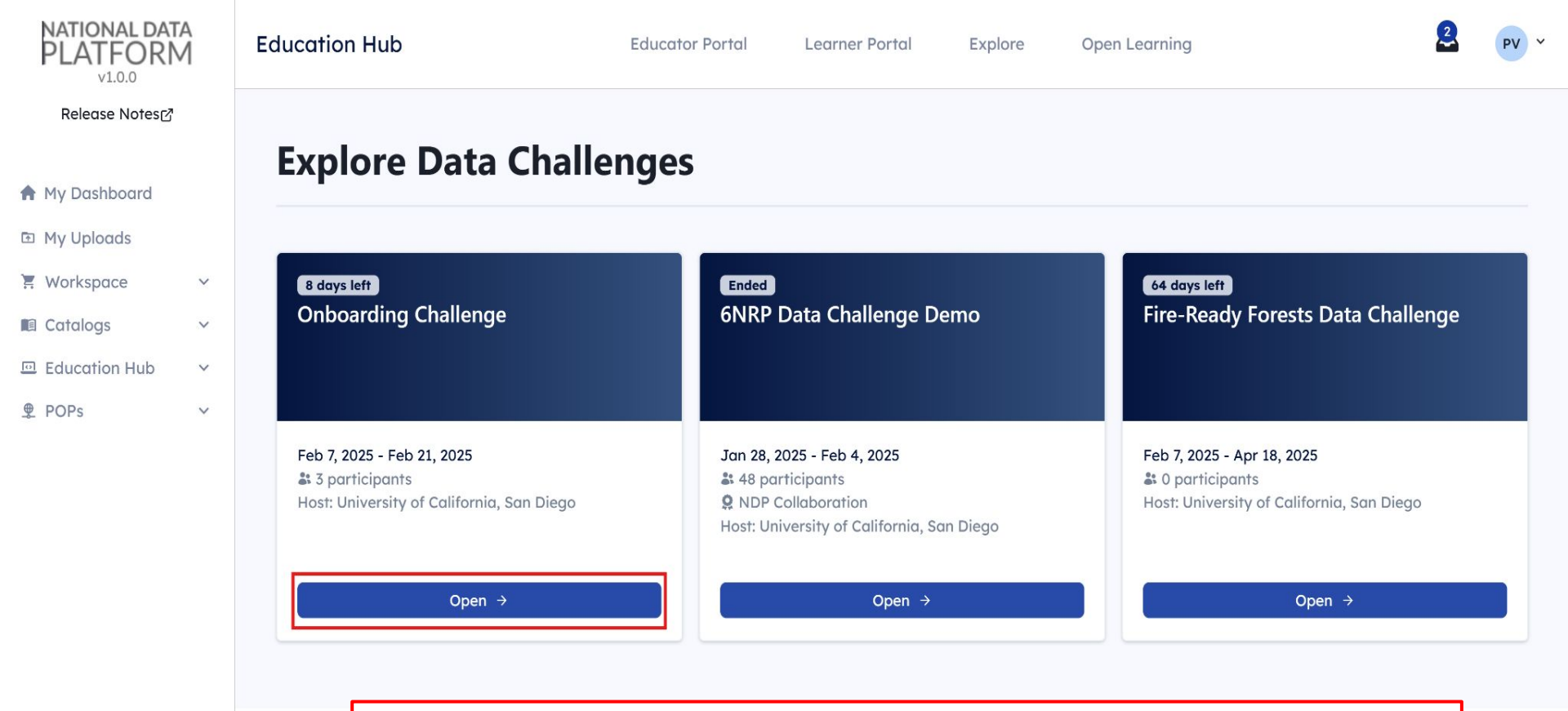
Learner's Portal

My Classrooms

Click on Explore

My Data Challenges

You have no data challenges.



Identify the Data Challenge of your interest and click on Open.
For this onboarding, we will join the *Onboarding Challenge*

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

← Go Back

Onboarding Challenge

University of California, San Diego

This Onboarding Challenge is designed to familiarize Data Challenge participants with the user experience of NDP's Education Hub.

Join Challenge

Overview | Rules and Eligibility | FAQ | Modules | Dataset Details

Background

The National Data Platform (NDP) has been developed as a federated and extensible data ecosystem to promote collaboration, innovation and equita

Explore all the information provided in the Data Challenge

The E ment of an AI-ready workforce. It enables instructors to deliver resource-intensive courses, facilitate data challenges, and share their expertise through hands-on, open learning resources

Challenge Description

This challenge is designed to be completed in teams, providing participants with a more engaging and collaborative experience.

Join the Data Challenge / Create a Team

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

← Go Back

Onboarding Challenge

University of California, San Diego

This Onboarding Challenge is designed to familiarize Data Challenge participants with the user experience of NDP's Education Hub.

Join Challenge

Overview | Rules and Eligibility | FAQ | Modules | set Details

Background

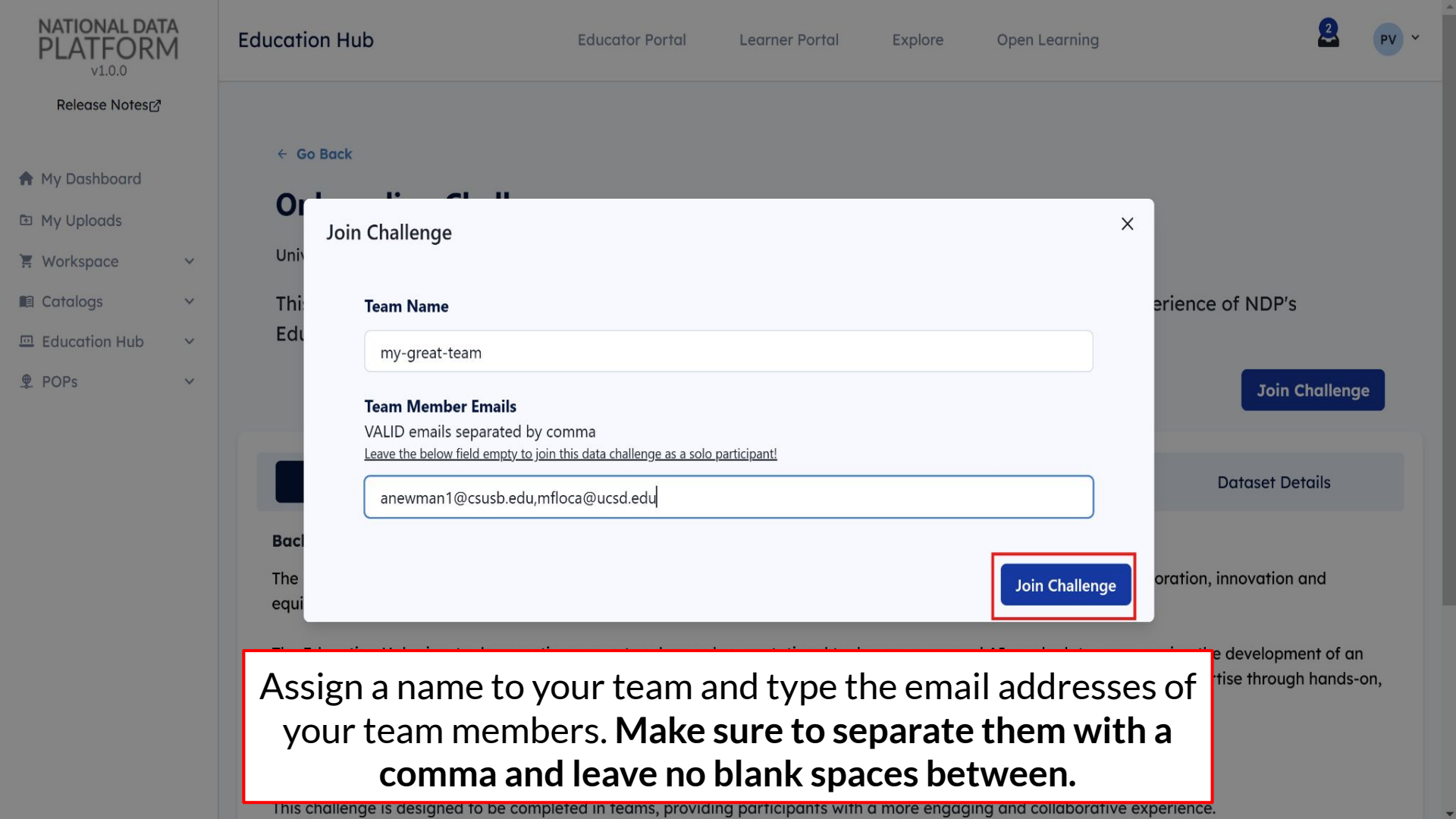
The National Data Platform (NDP) has been developed as a federated and extensible data ecosystem to promote collaboration, innovation and equitable use of data using existing and future national cyberinfrastructure (

The Education Hub aims to democratize access to advanced computational t velopment of an AI-ready workforce. It enables instructors to deliver resource-intensive courses, facilitate data challenges, and share their expertise through hands-on, open learning resources

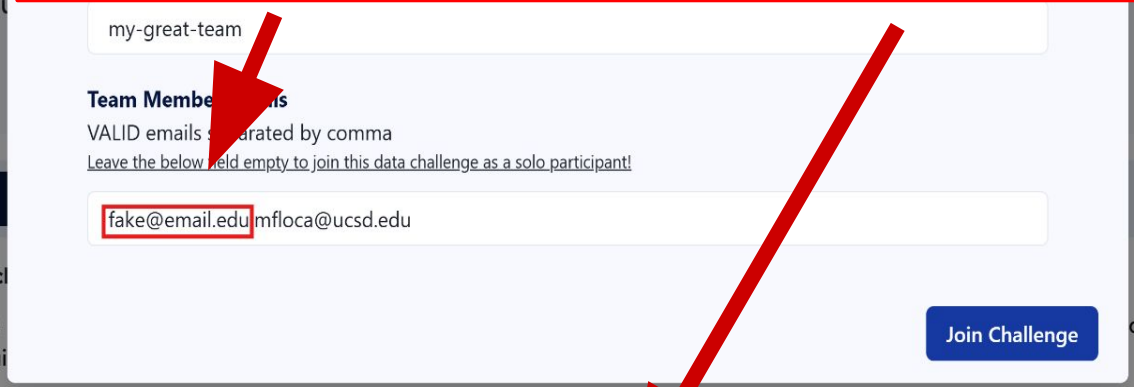
Challenge Description

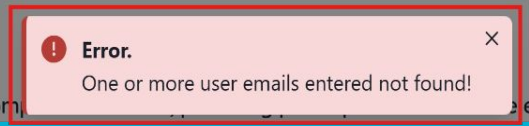
This challenge is designed to be completed in teams, providing participants with a more engaging and collaborative experience.

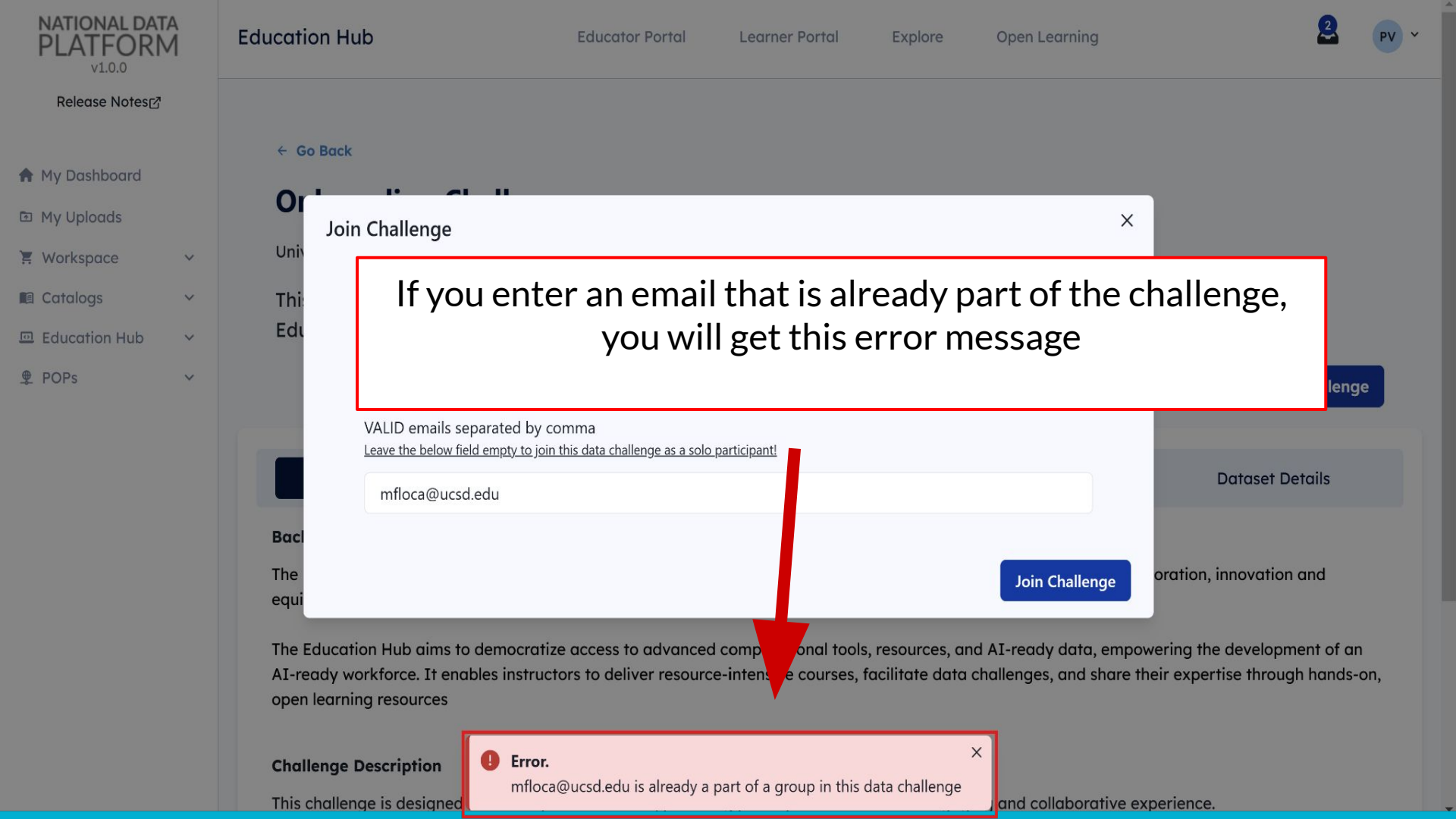
Click on Join Challenge


NATIONAL DATA PLATFORM
v1.0.0
Release Notes
Education Hub
Educator Portal
Learner Portal
Explore
Open Learning
PV
My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs
Go Back
Join Challenge
Dataset Details
Join Challenge
Team Name
my-great-team
Team Member Emails
VALID emails separated by comma
Leave the below field empty to join this data challenge as a solo participant!
anewman1@csusb.edu,mfloca@ucsd.edu
Join Challenge
Assign a name to your team and type the email addresses of your team members. **Make sure to separate them with a comma and leave no blank spaces between.**

If you enter a non-valid email, or an email that has not been registered yet (slides 3-9), you will get this error message






NATIONAL DATA PLATFORM
v1.0.0
Release Notes
Education Hub
Educator Portal
Learner Portal
Explore
Open Learning
PV
My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs
Go Back
Join Challenge
If you enter an email that is already part of the challenge, you will get this error message
VALID emails separated by comma
Leave the below field empty to join this data challenge as a solo participant!
mfloca@ucsd.edu
Join Challenge
Dataset Details
The Education Hub aims to democratize access to advanced comp onal tools, resources, and AI-ready data, empowering the development of an AI-ready workforce. It enables instructors to deliver resource-intens e courses, facilitate data challenges, and share their expertise through hands-on, open learning resources
Challenge Description
Error.
mfloca@ucsd.edu is already a part of a group in this data challenge

The onboarding module

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

Your Team | Overview | Rules and Eligibility | FAQ | Dataset Details

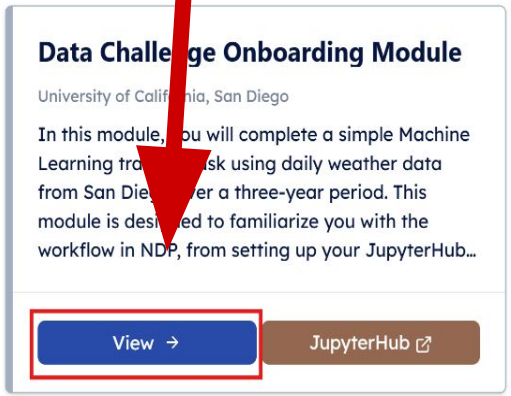
Deadline: February 20, 2025

8 DAYS LEFT

Once you join a Data Challenge, your modules will appear at the bottom of *Your Team* page. To view the content of the onboarding module (the single module of this onboarding challenge), click on the *View* button.

Edit

Modules



NOTE: Read Modules Carefully

Read carefully the instructions of each module prior to launching them on JupyterHub. Each module should come with specific instructions on how to:

- Reserve your JupyterHub server computing resources
- Clone the attached repository (if provided)
- Install dependencies
- Download your data to JupyterHub

Launching JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Datasets

HPWREN Weather Station Measurements

Models

Once you have read the module's instructions, you can go to JupyterHub by clicking on the JupyterHub button at the bottom of the module

Scripts

https://github.com/pramonettivega/eh-onboarding.git

Go to JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Your Team | Overview | Rules and Eligibility | FAQ | Dataset Details

Deadline: February 20, 2025

8 DAYS LEFT

Team Details

Edit

This button is also located in the cover of each module

Modules

Data Challenge Onboarding Modul

University of California, San Diego

In this module, you will complete a simple Mac ne Learning training task using daily weather dat from San Diego over a three-year period. module is designed to familiarize you with workflow in NDP, from setting up your Jupy Hub...

View →

JupyterHub

Setup resources as indicated in your module

NDP JupyterHub Server Options

Available resources page

Region

Any

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

/dev/shm for pytorch

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

Enter your custom image URL here, including the tag. For example: jupyter/r-notebook:latest

Timeout (in seconds): once a server has been successfully spawned, time to wait until it actually starts

1200

Reserving unauthorized resources (such as reserving GPU's without prior authorization) might ban you from the platform

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

☐ /dev/shm for pytorch

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

Enter your custom image URL here, including the tag.

Timeout (in seconds): once a server has been successfully spawned, time to wait until it actually starts

1200

Architecture

amd64

Note: *Please stop your server after it is no longer needed, or in case you want to laun… …ent content image*
In order to stop the server from running Jupyter Lab, go to File > Hub Control Panel > … …erver

Note: */home/jovyan/work/_User-Persistent-Storage_CephBlock_ is the persistent volu… directory, make sure to save your work in it, otherwise it will be deleted*

Start

After setting up your resources, click on start to launch your server

NDP JupyterHub Server Options

Available resources page

Error: HTTP 401: Unauthorized (Your session has expired. Please log out and log in again.)

You may encounter this message. This means your access credentials within JupyterHub have expired.

Region

Any

Zone

Any

GPUs

16

GPU type

Any

☐ **/dev/shm for pytorch**

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

NATIONAL DATA PLATFORM Home Token pramonettivega@ucsd.edu Logout

NDP JupyterHub Server Options

Available resources page

Error: HTTP 401: Unauthorized (Your session has expired. Please log out and log in again.)

To solve this problem, simply logout using this button, and login again through CILogon. Once you do that, you will be able to start your server

Region

Any

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

☐ **/dev/shm for pytorch**

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

NATIONAL DATA PLATFORM Home Token pramonettivega@ucsd.edu Logout

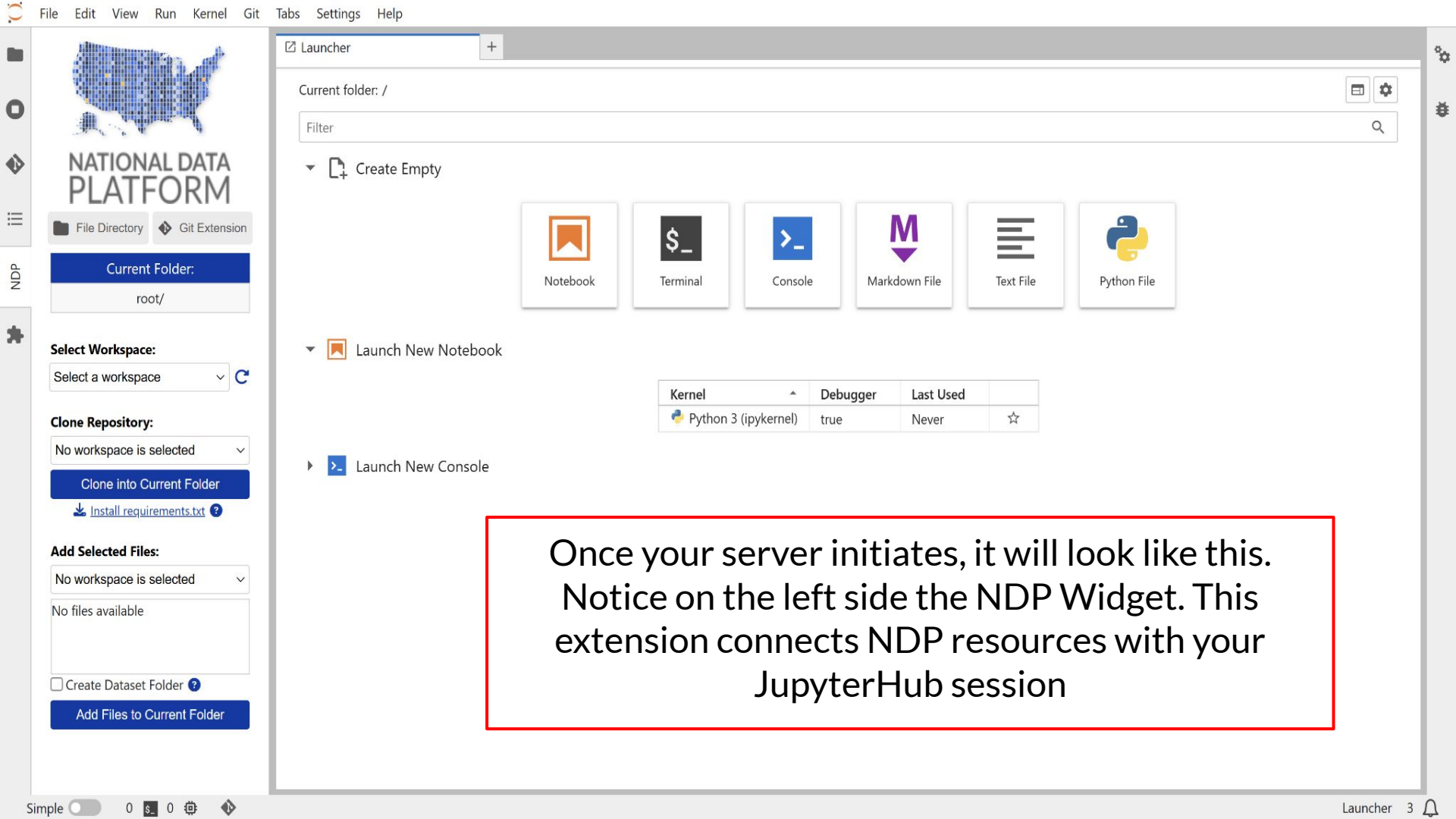
Your server is starting up.

You will be redirected automatically when it's ready for you.

2025-01-29T21:35:52Z [Normal] AttachVolume.Attach succeeded for volume "pvc-f54bb41b-2b4b-4efd-8c20-e43441ec0215"

Event log

Wait for your server to initiate


File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
NDP
Once your server initiates, it will look like this. Notice on the left side the NDP Widget. This extension connects NDP resources with your JupyterHub session
Simple
Launcher 3

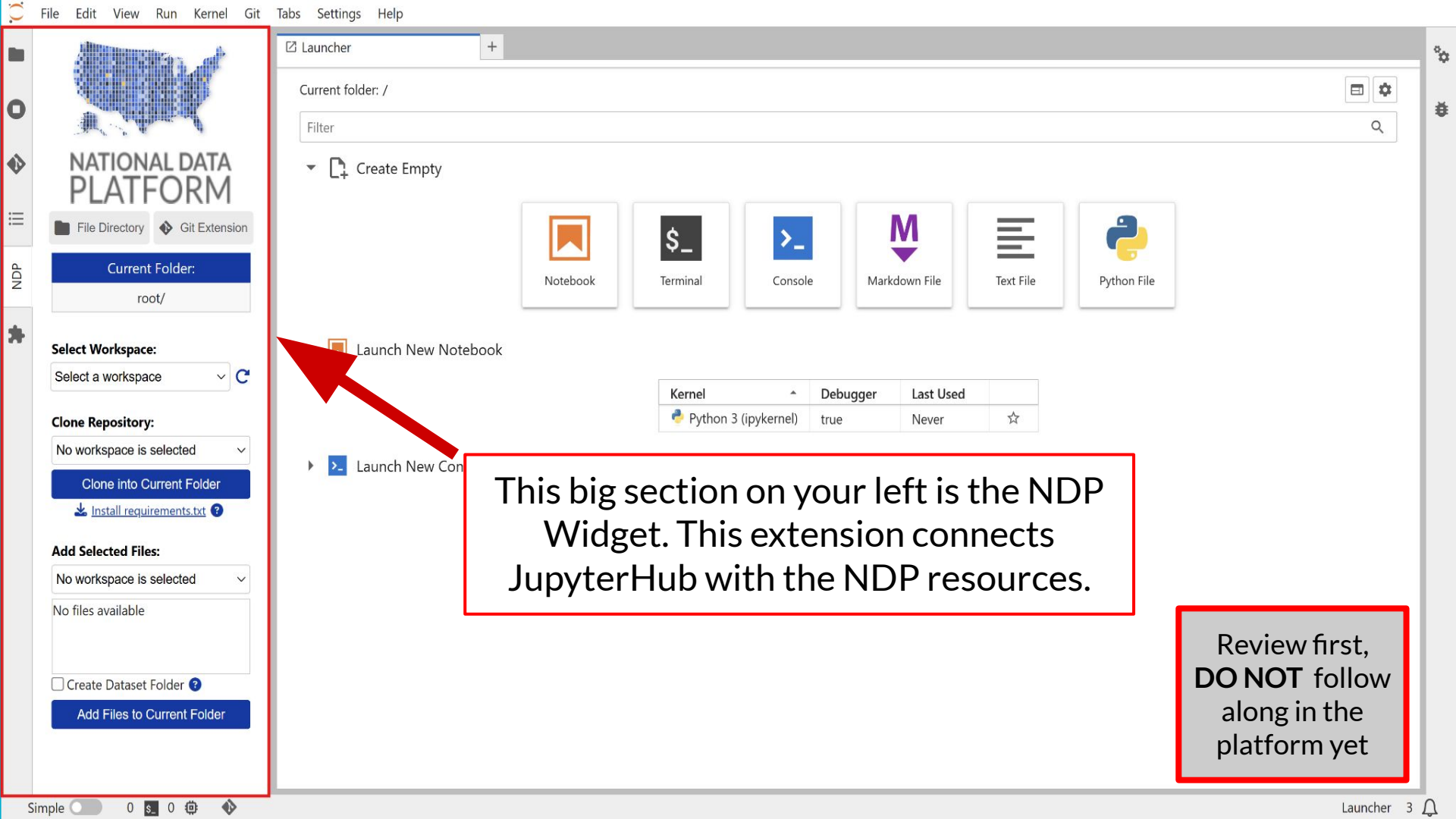
JupyterHub sections

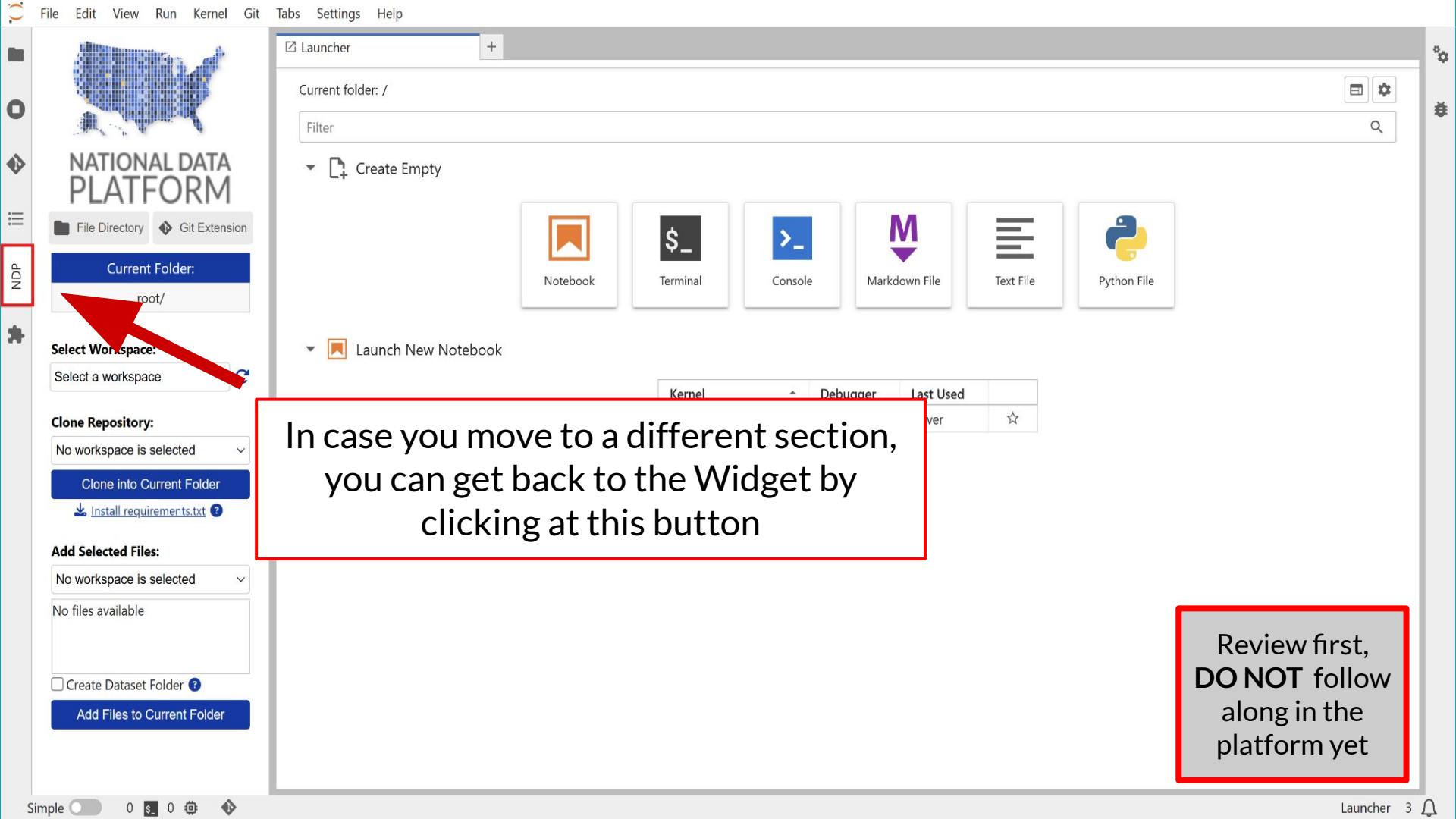
In the next slides, we will take a look at the different sections and pieces of your JupyterHub server. After the quick overview, we will proceed to follow the steps of the onboarding module.

The following slides are for reviewing first. Please **DO NOT FOLLOW ALONG IN THE PLATFORM YET.**

a. NDP Widget

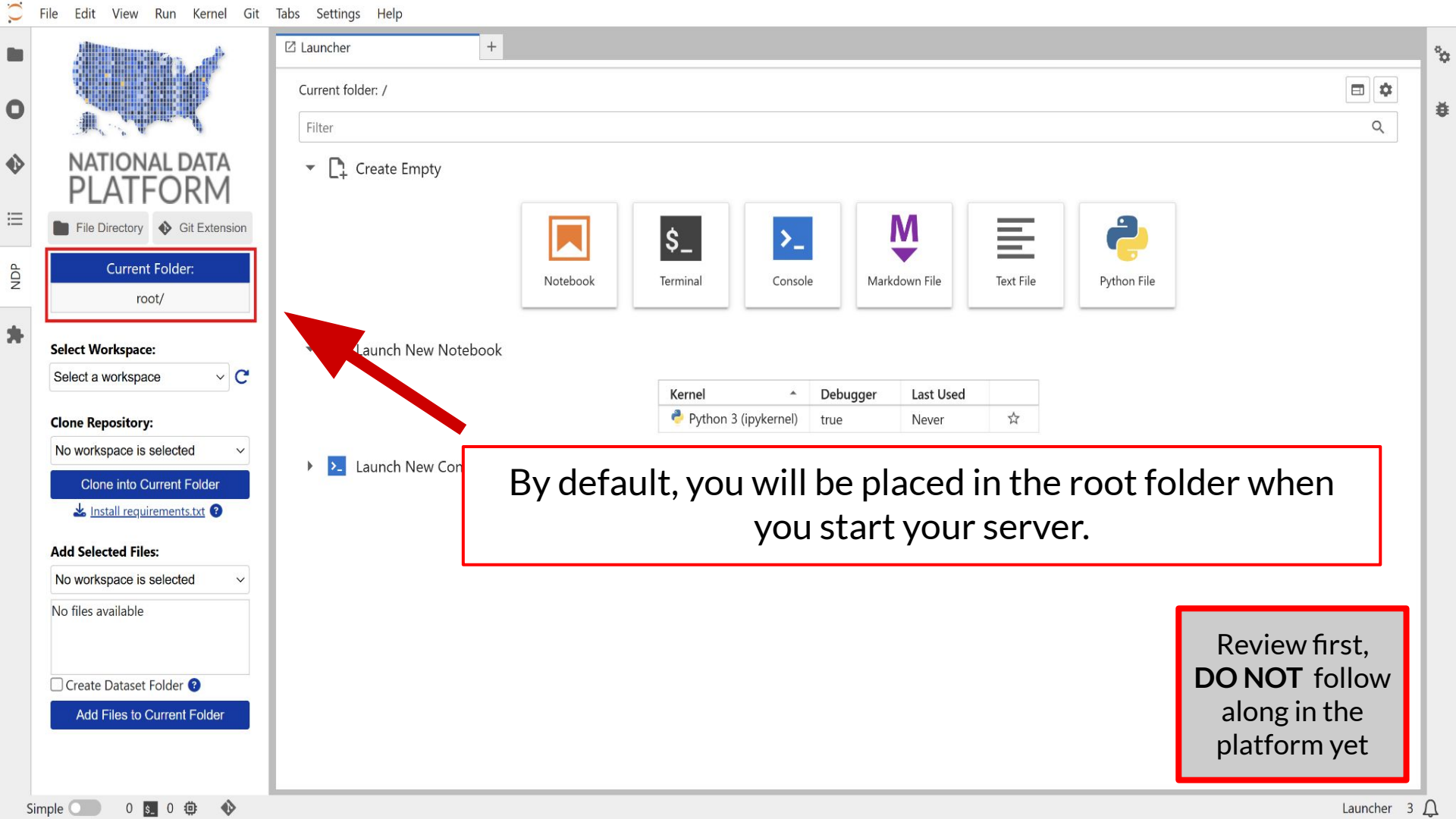
Review first, **DO NOT** follow along in the platform yet

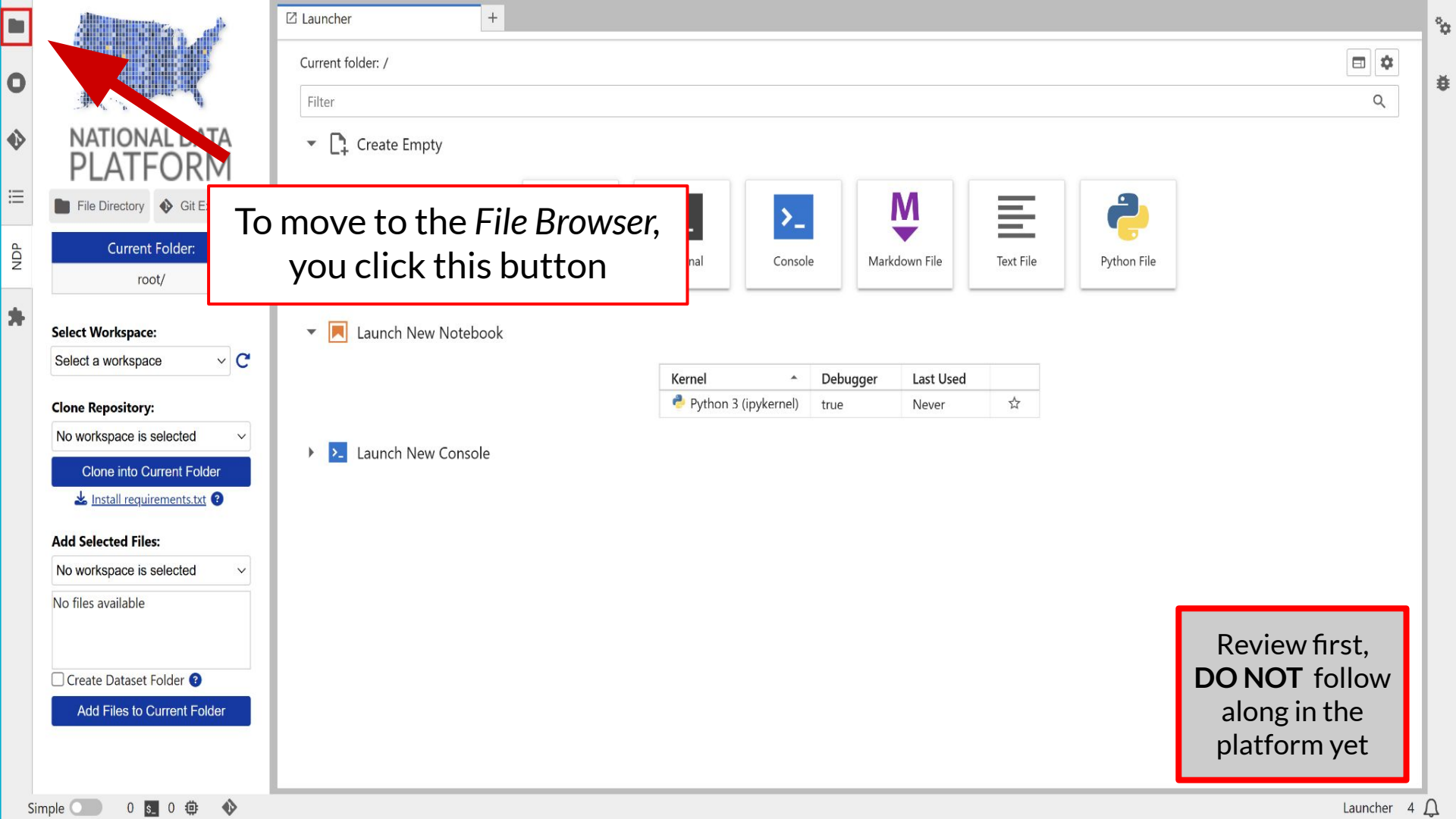

File Edit View Run Kernel Git Tabs Settings Help
NDP
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Con
This big section on your left is the NDP Widget. This extension connects JupyterHub with the NDP resources.
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
NDP
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
In case you move to a different section, you can get back to the Widget by clicking at this button
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3

b. Current Folder & File Browser

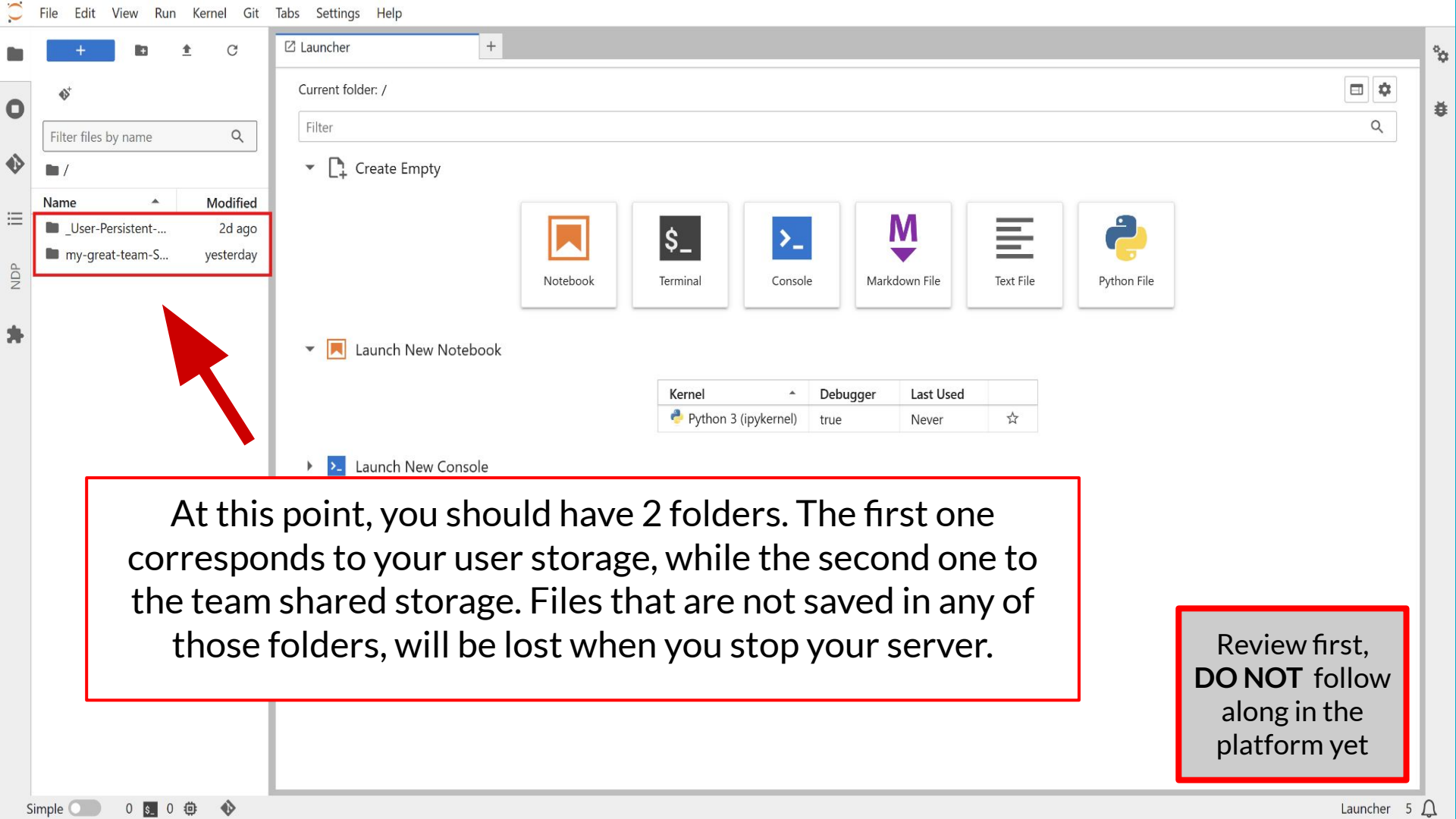
Review first, **DO NOT** follow along in the platform yet


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Con
By default, you will be placed in the root folder when you start your server.
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3


NATIONAL DATA PLATFORM
File Directory
Git E
Current Folder:
root/
To move to the *File Browser*, you click this button
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Review first, **DO NOT** follow along in the platform yet
Simple
Launcher 4

c. User Storage and Shared storage

Review first, **DO NOT** follow along in the platform yet


File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: /
Filter files by name
/
Name
Modified
_User-Persistent-...
2d ago
my-great-team-S...
yesterday
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
At this point, you should have 2 folders. The first one corresponds to your user storage, while the second one to the team shared storage. Files that are not saved in any of those folders, will be lost when you stop your server.
Review first, DO NOT follow along in the platform yet
Simple
Launcher 5

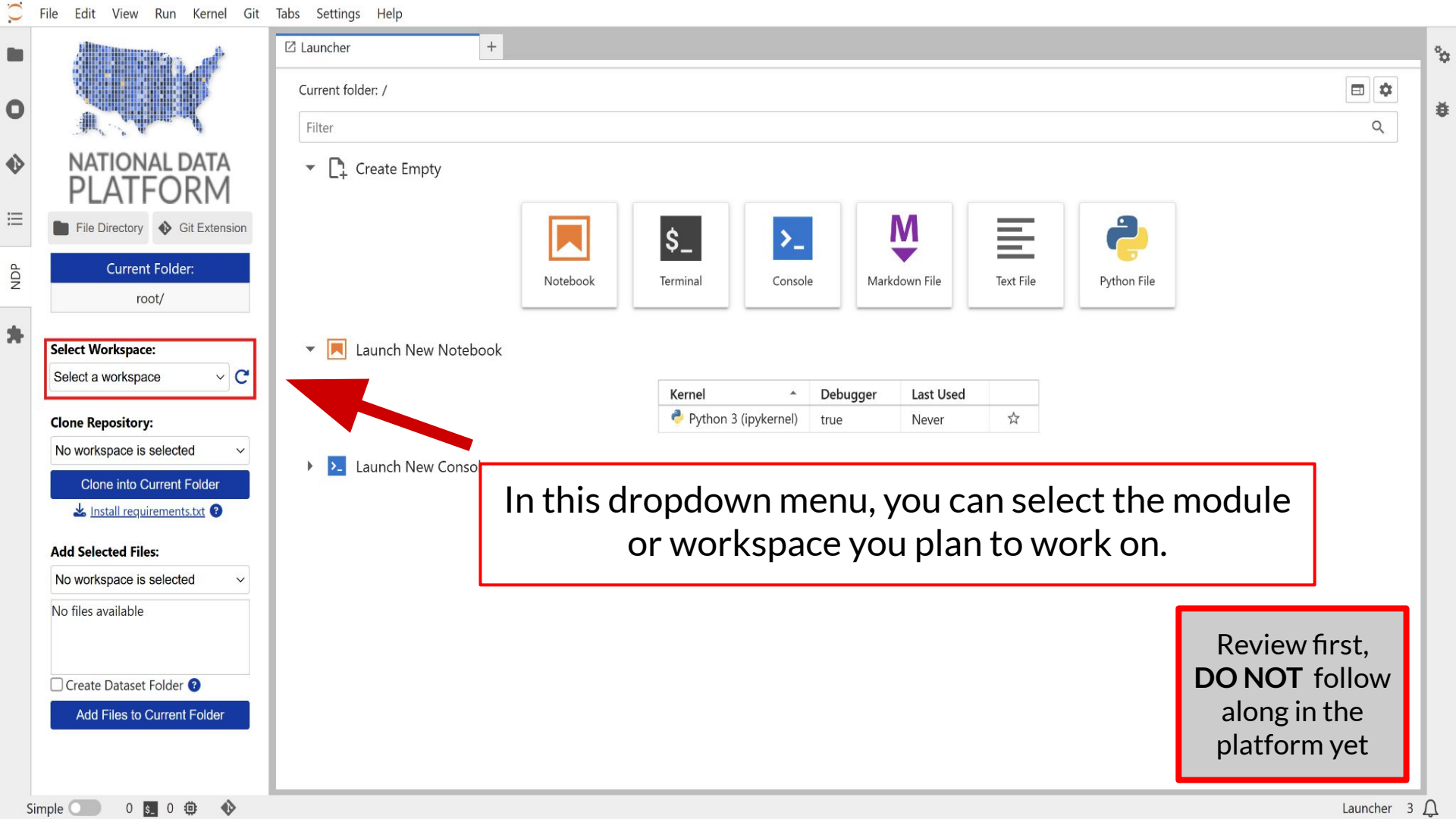
Good practices about shared storage

- Always communicate with your team members about the files they place/remove from this folder.
- Do not work on the same file at the same time or you will run into an overwrite conflict.
- Shared storage is limited, so use it to share frequently accessed files or data derived from your workflow.

Review first, **DO NOT** follow along in the platform yet

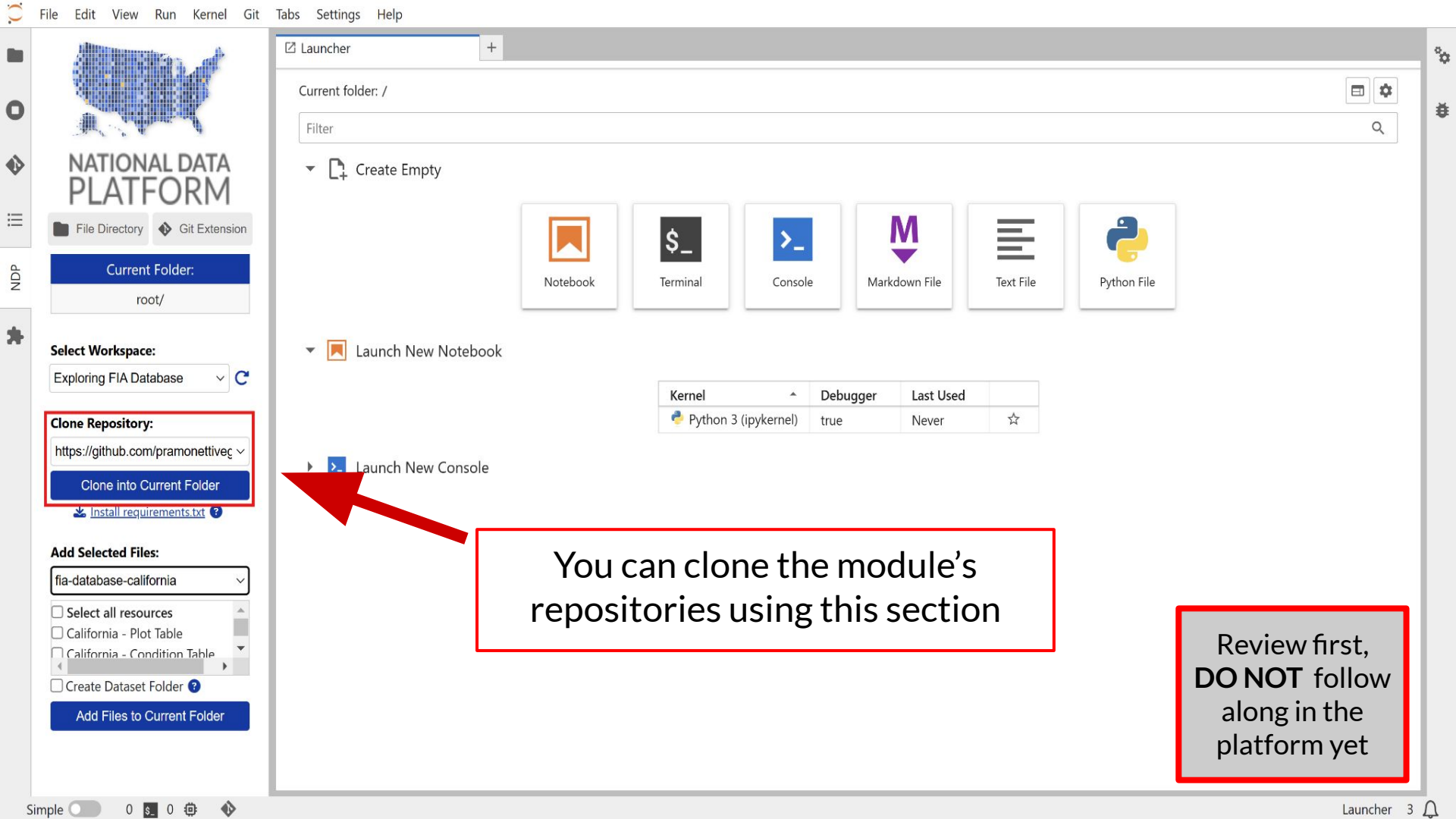
e. Modules/Workspaces List

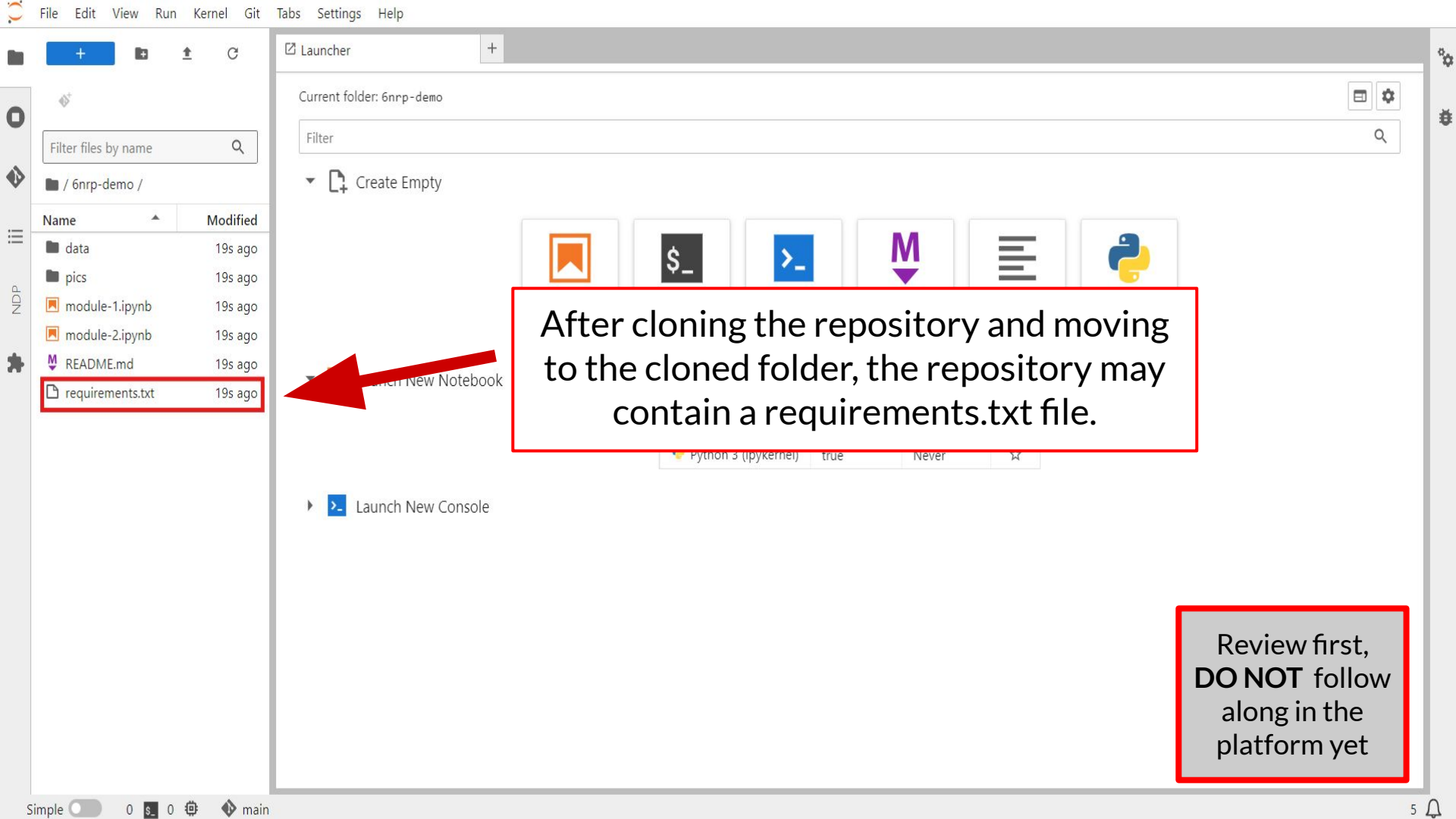
Review first, **DO NOT** follow along in the platform yet

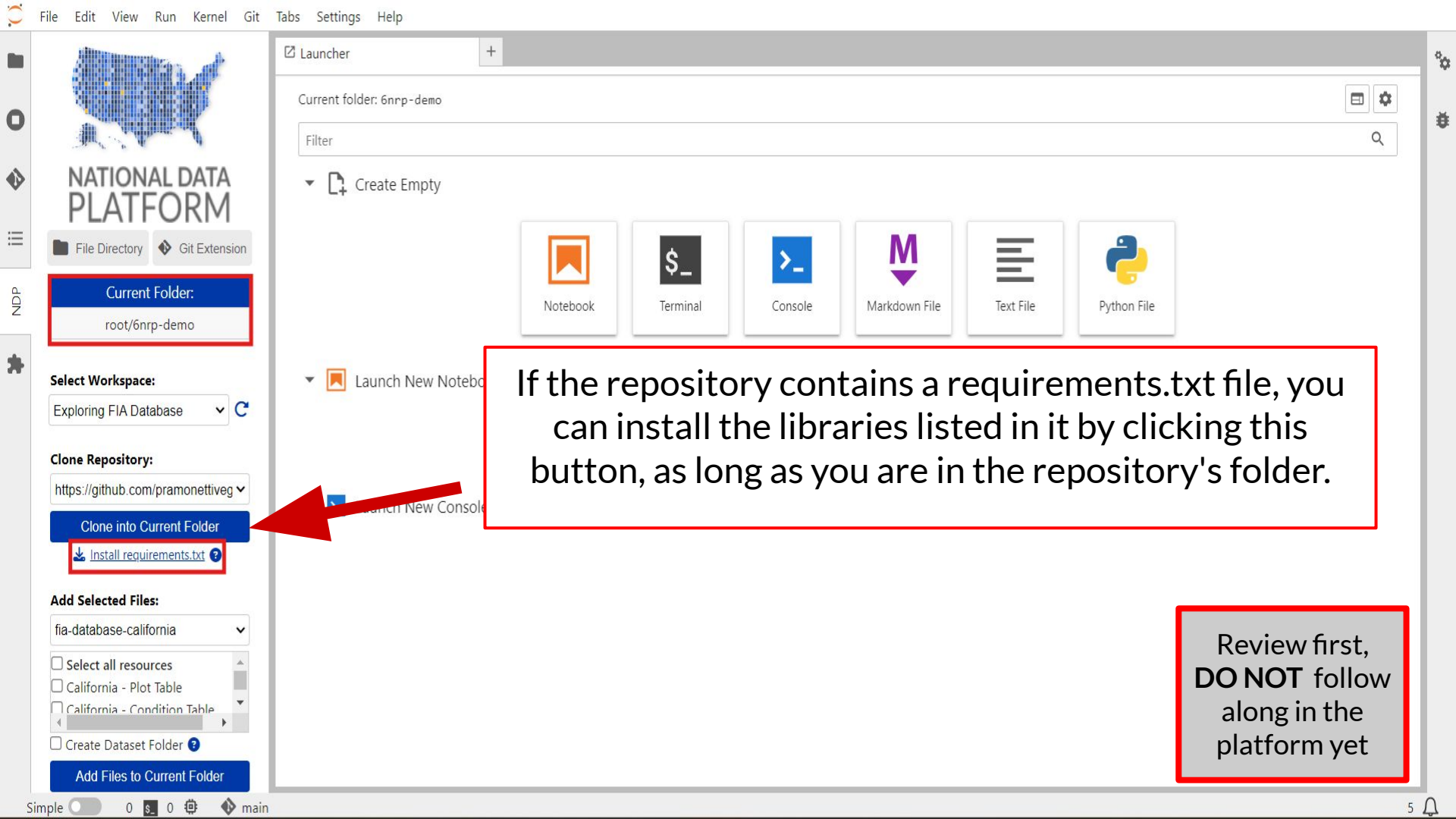

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Conso
In this dropdown menu, you can select the module or workspace you plan to work on.
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3

f. Repositories List

Review first, **DO NOT** follow along in the platform yet

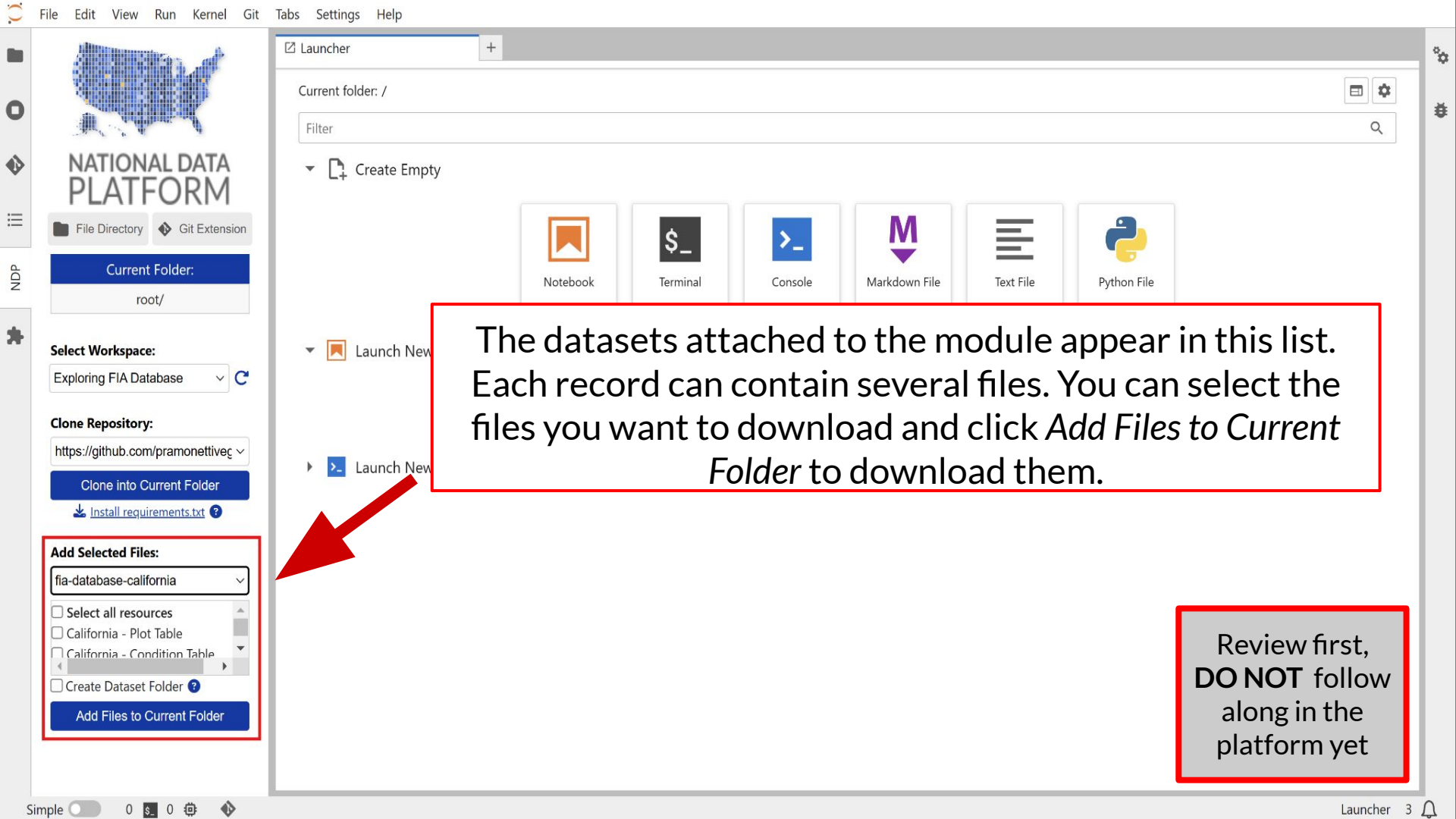

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
You can clone the module's repositories using this section
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3

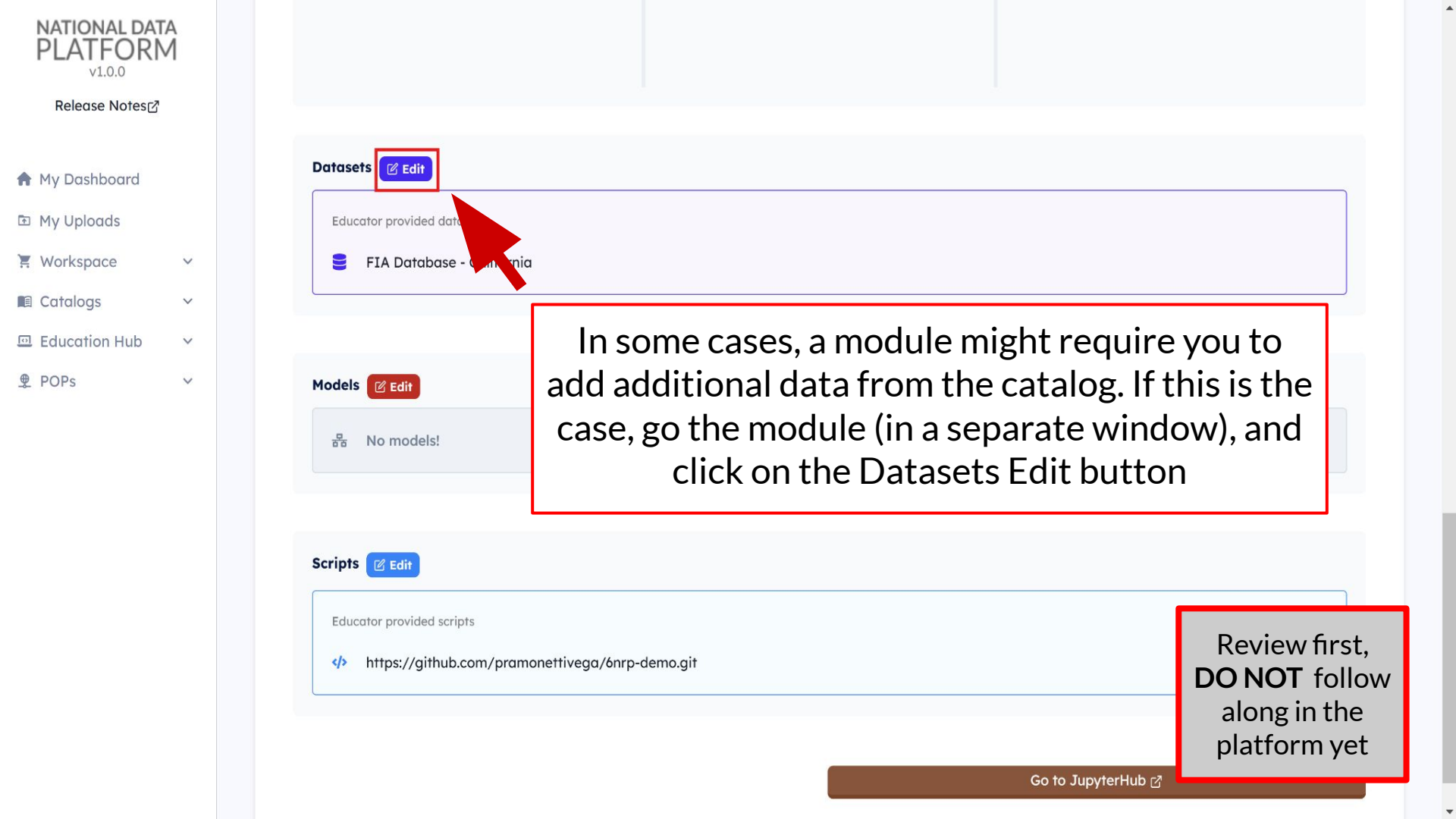

File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: 6nrp-demo
Filter files by name
/ 6nrp-demo /
Name
Modified
data 19s ago
pics 19s ago
module-1.ipynb 19s ago
module-2.ipynb 19s ago
README.md 19s ago
requirements.txt 19s ago
Filter
Create Empty
Launch New Notebook
Launch New Console
After cloning the repository and moving to the cloned folder, the repository may contain a requirements.txt file.
Review first, DO NOT follow along in the platform yet
Simple
main


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/6nrp-demo
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: 6nrp-demo
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebo
If the repository contains a requirements.txt file, you can install the libraries listed in it by clicking this button, as long as you are in the repository's folder.
Review first, DO NOT follow along in the platform yet
Simple
main

g. Datasets List

Review first, **DO NOT** follow along in the platform yet


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New
Launch New
The datasets attached to the module appear in this list. Each record can contain several files. You can select the files you want to download and click *Add Files to Current Folder* to download them.
Review first, **DO NOT** follow along in the platform yet
Simple
Launcher 3


NATIONAL DATA PLATFORM
v1.0.0
Release Notes
My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs
Datasets
Edit
Educator provided dat
FIA Database -
nia
In some cases, a module might require you to add additional data from the catalog. If this is the case, go the module (in a separate window), and click on the Datasets Edit button
Models
Edit
No models!
Scripts
Edit
Educator provided scripts
https://github.com/pramonettivega/6nrp-demo.git
Review first, DO NOT follow along in the platform yet
Go to JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

Edit Datasets

Module datasets

FIA Database - California

You have not added any datasets.

Review first, do not follow along in the platform yet

Search data catalog

sierra

Search | Clear

1 - 3 of 97 Data Collections and Streams

1 2 ... 33

Sierra Nevada Sawtimber Volume

This metric expresses the amount of total existing, aboveground, live tree stem biomass measured in dry weight tons per acre for the Sierra Nevada region. This metric can be used to assess the sawtimber volume present at t...

View More | + Add

Sierra Nevada Biomass Volume

...ry weight tons per acre) from ...for the Sierra Nevada region....

View More | + Add

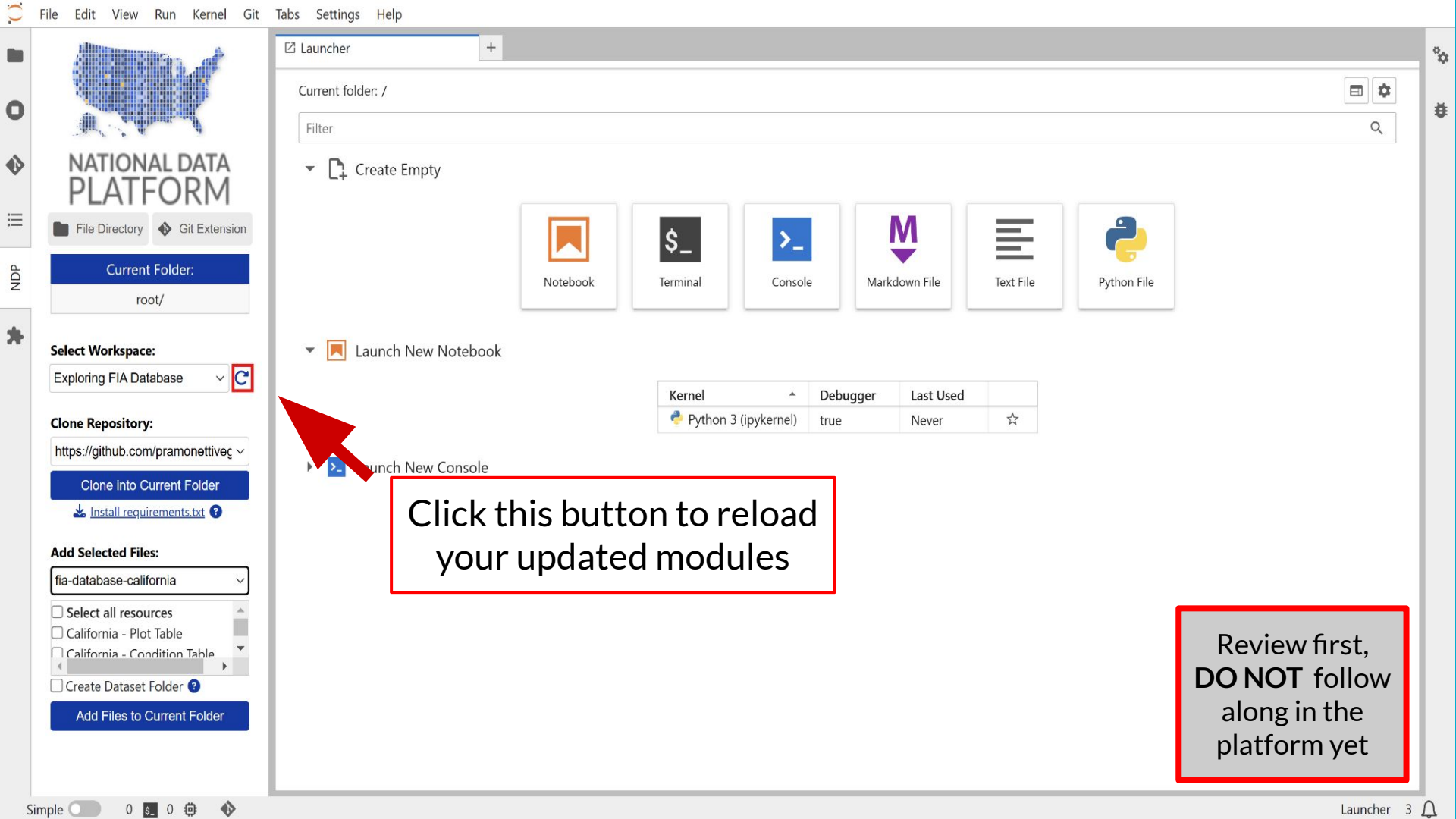
This metric is intended to be used to identify areas where reforestation may be necessary if stakeholders want to reestablish coniferous forests following fire. Conifers in our region generally lack the capacity to resprout after fi...

View Mor

Search for the dataset of your interest and add it to your module

Review first, **DO NOT** follow along in the platform yet

Go to JupyterHub

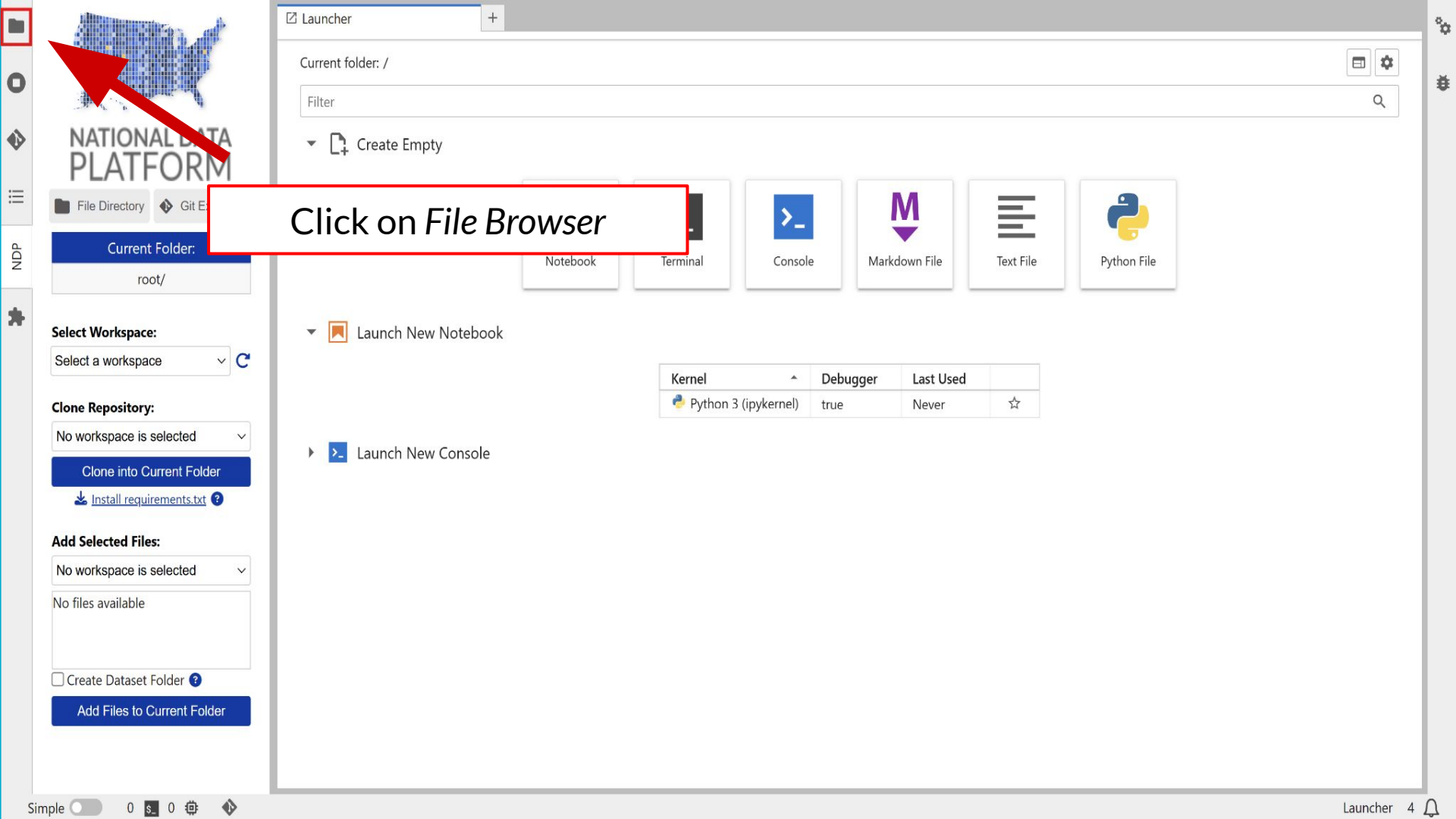

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Click this button to reload your updated modules
Review first, DO NOT follow along in the platform yet
Simple
Launcher 3

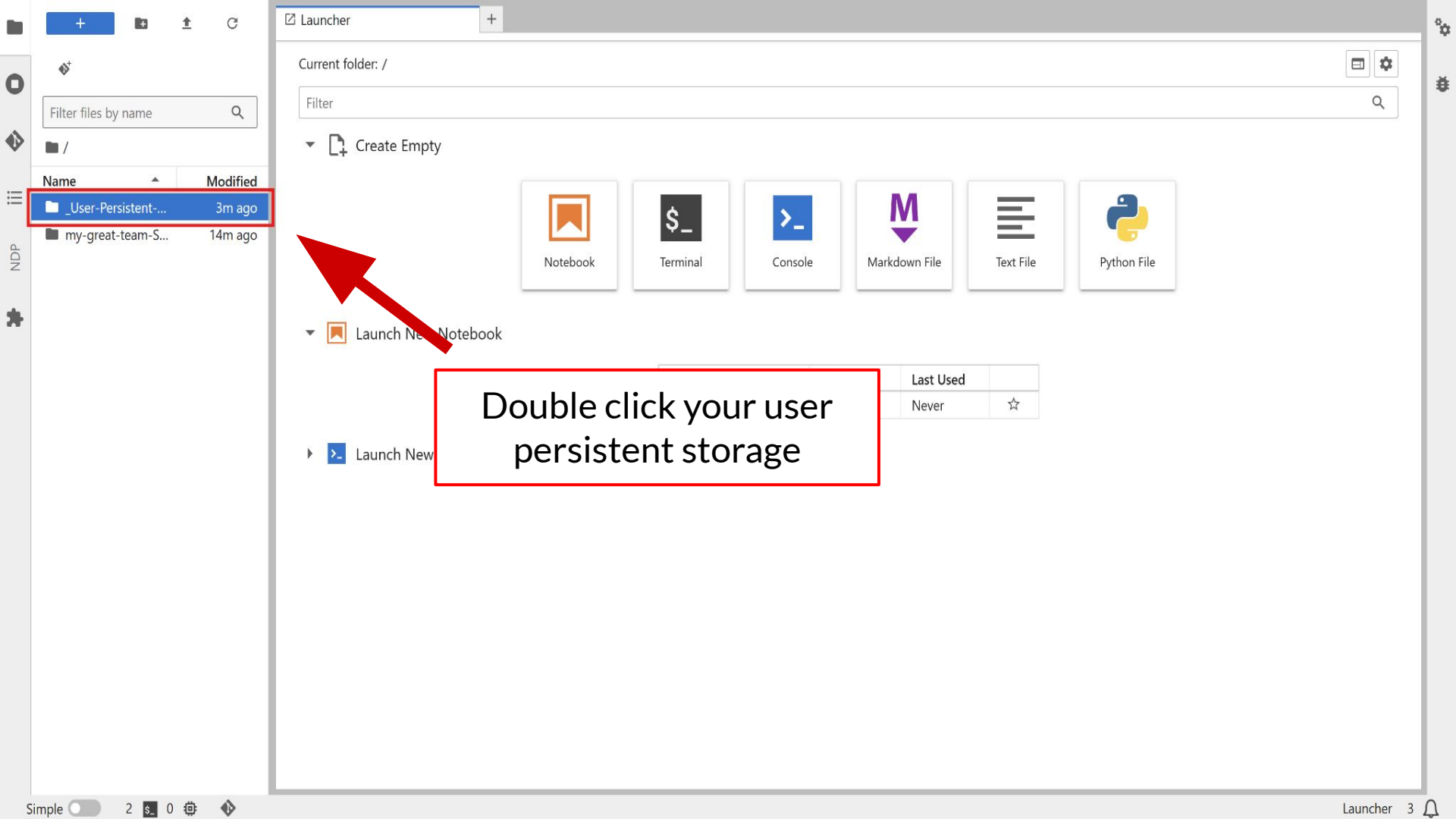
The onboarding module

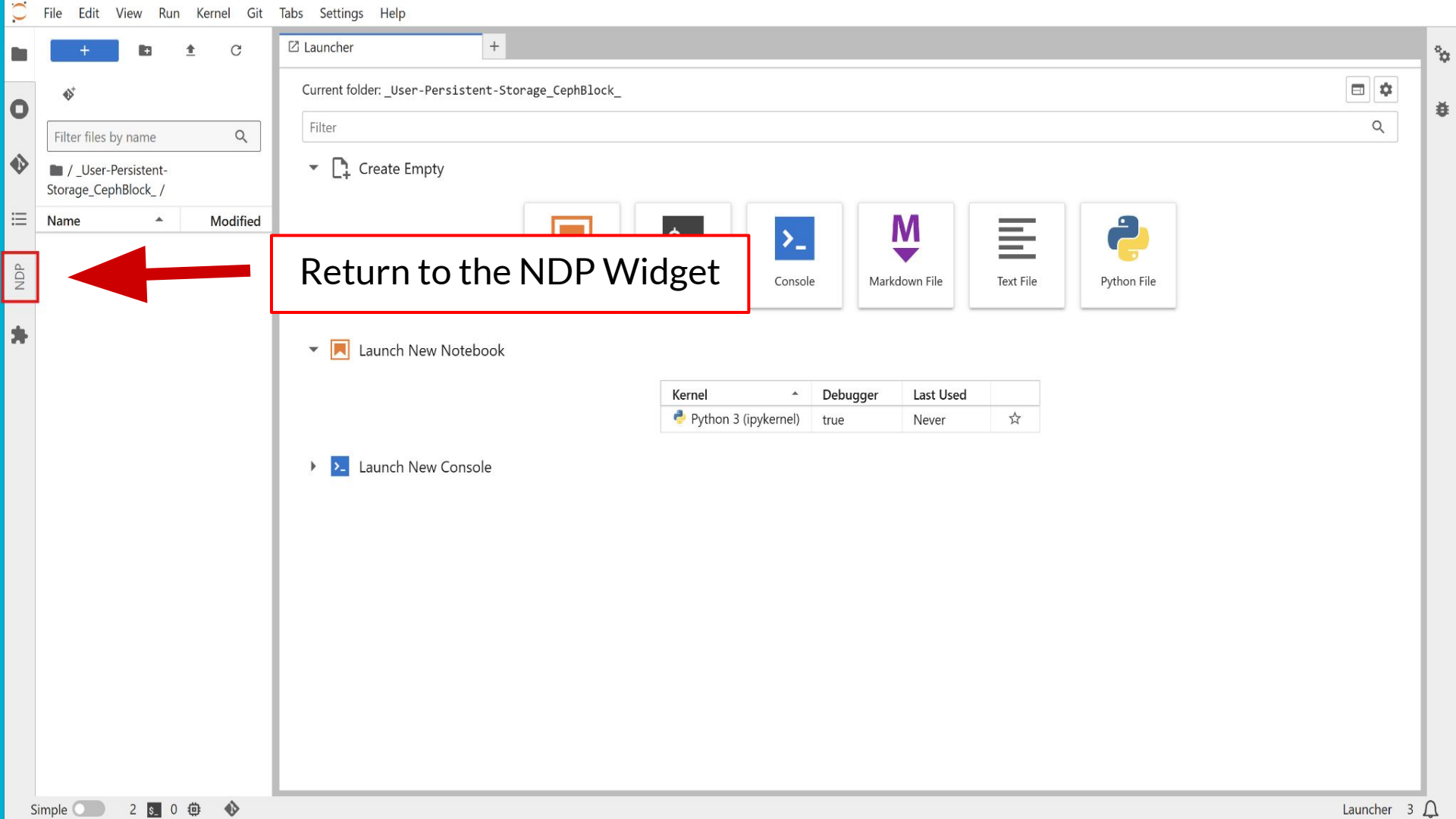
In the next slides, we will give you a visual guidance for the instructions 4-10 of the onboarding module.

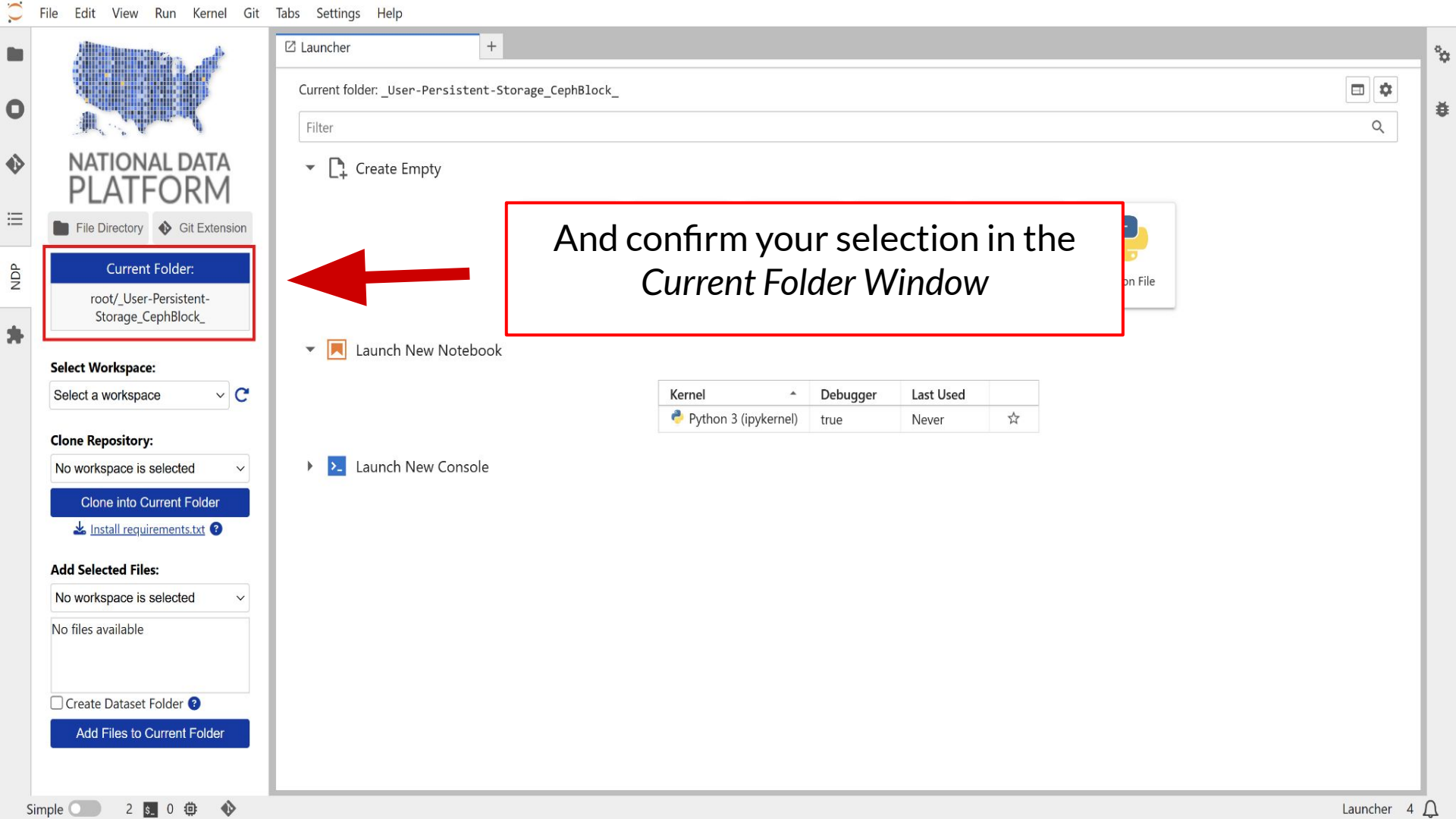
You may now follow along.

4. Set your storage

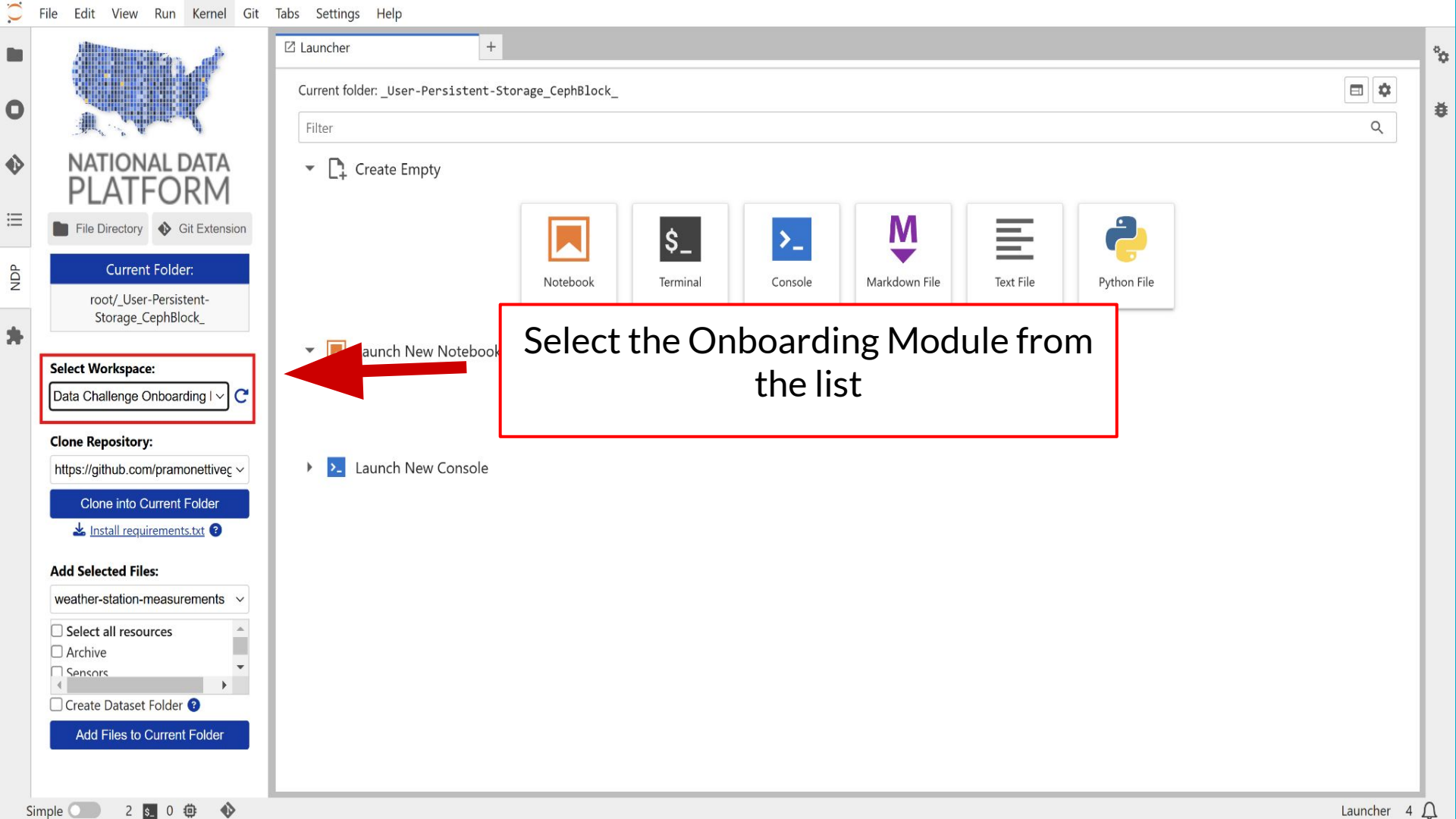

Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
NATIONAL DATA PLATFORM
File Directory
Git
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
NDP
Click on File Browser
Simple
Launcher 4


Launcher
Current folder: /
Filter files by name
/
Name
Modified
_User-Persistent-...
3m ago
my-great-team-S...
14m ago
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Last Used
Never
Launch New
Double click your user persistent storage
Simple
2
0
Launcher
3

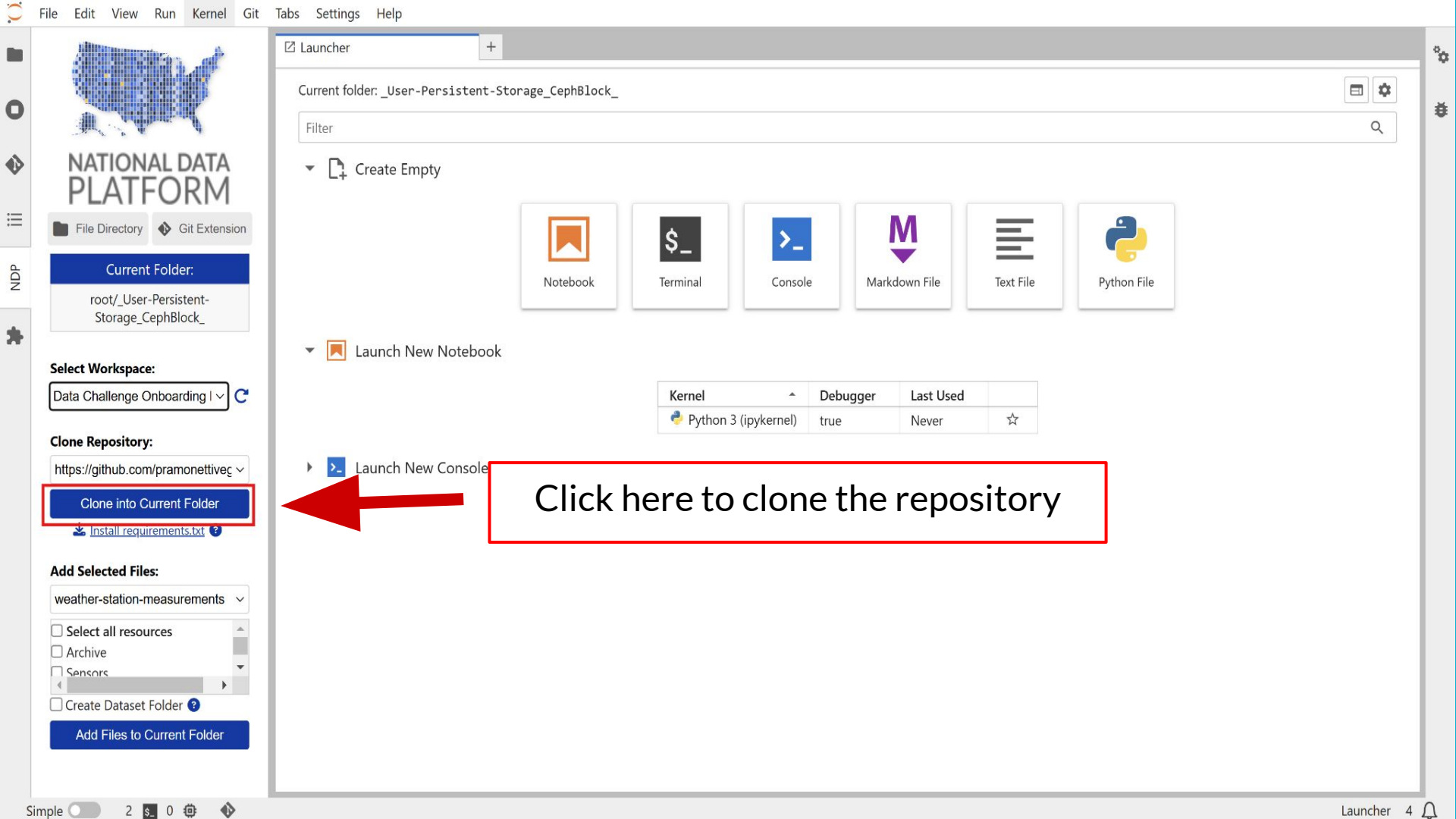

File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter files by name
/ _User-Persistent-Storage_CephBlock_ /
Name
Modified
NDP
Filter
Create Empty
Console
Markdown File
Text File
Python File
Return to the NDP Widget
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Simple
Launcher 3

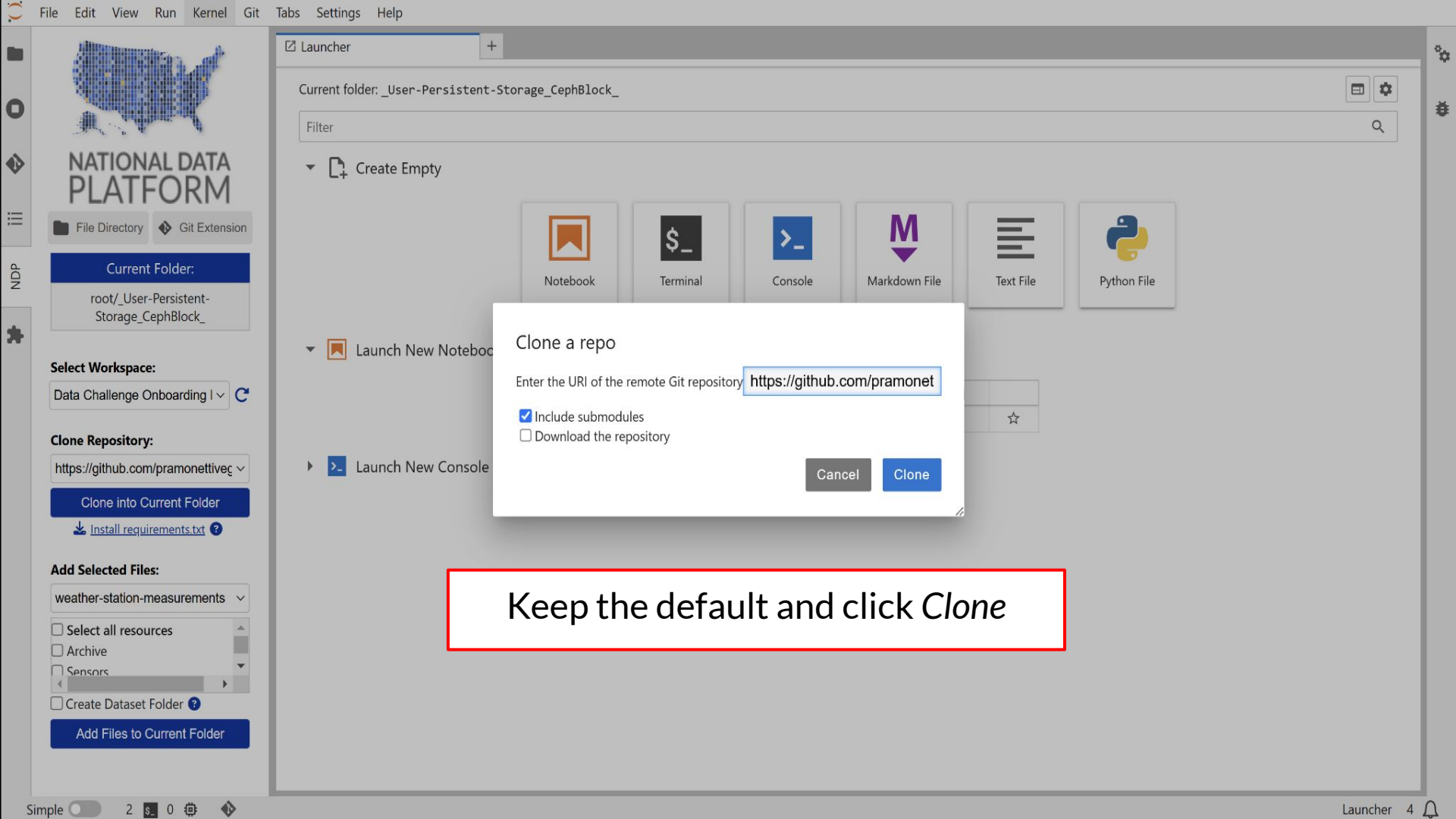

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter
Create Empty
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
And confirm your selection in the Current Folder Window
Simple
Launcher 4

5. Select workspace

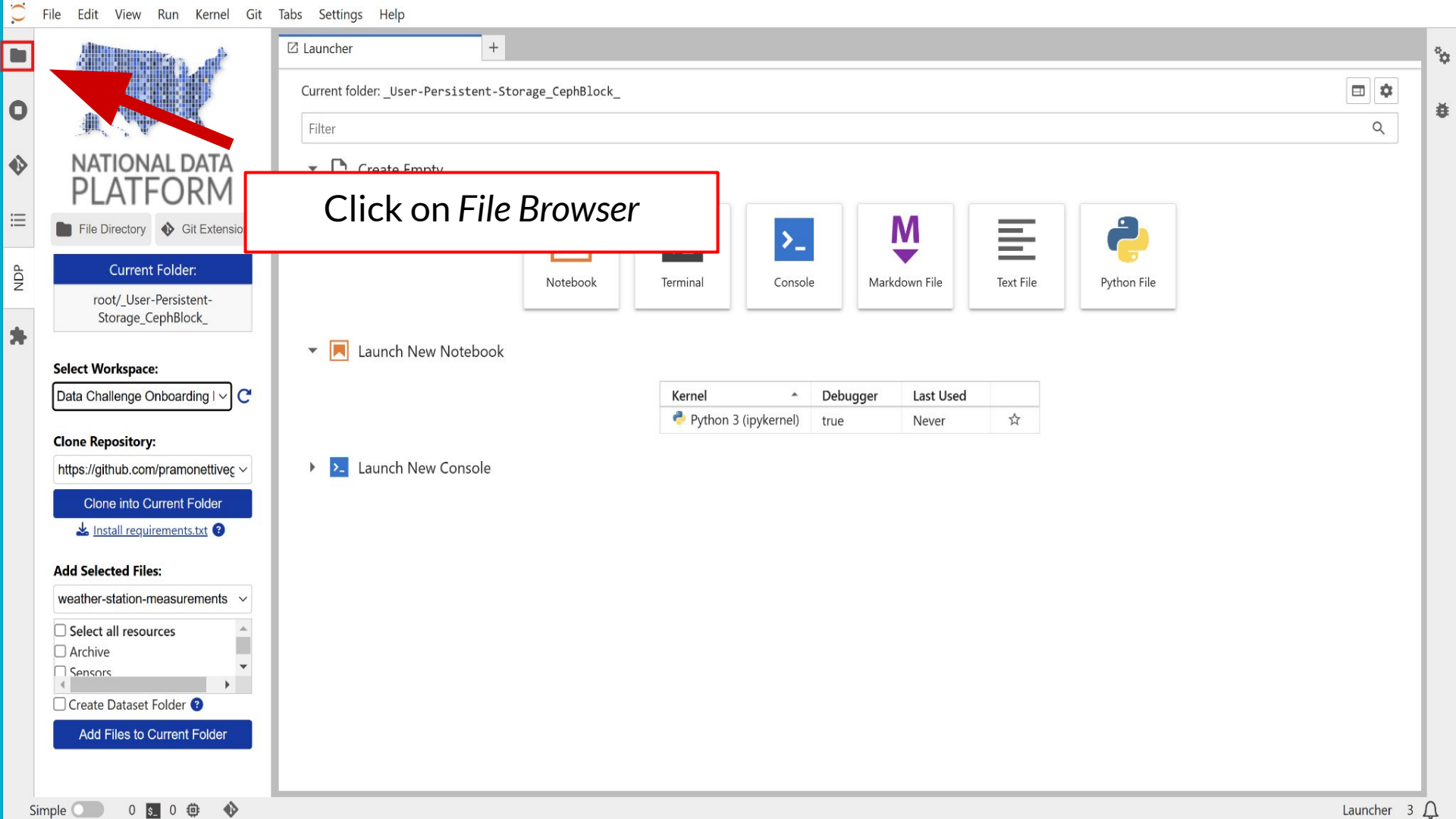

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_
Select Workspace:
Data Challenge Onboarding
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Select all resources
Archive
Sensors
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Launch New Console
Select the Onboarding Module from the list
Simple
Launcher 4

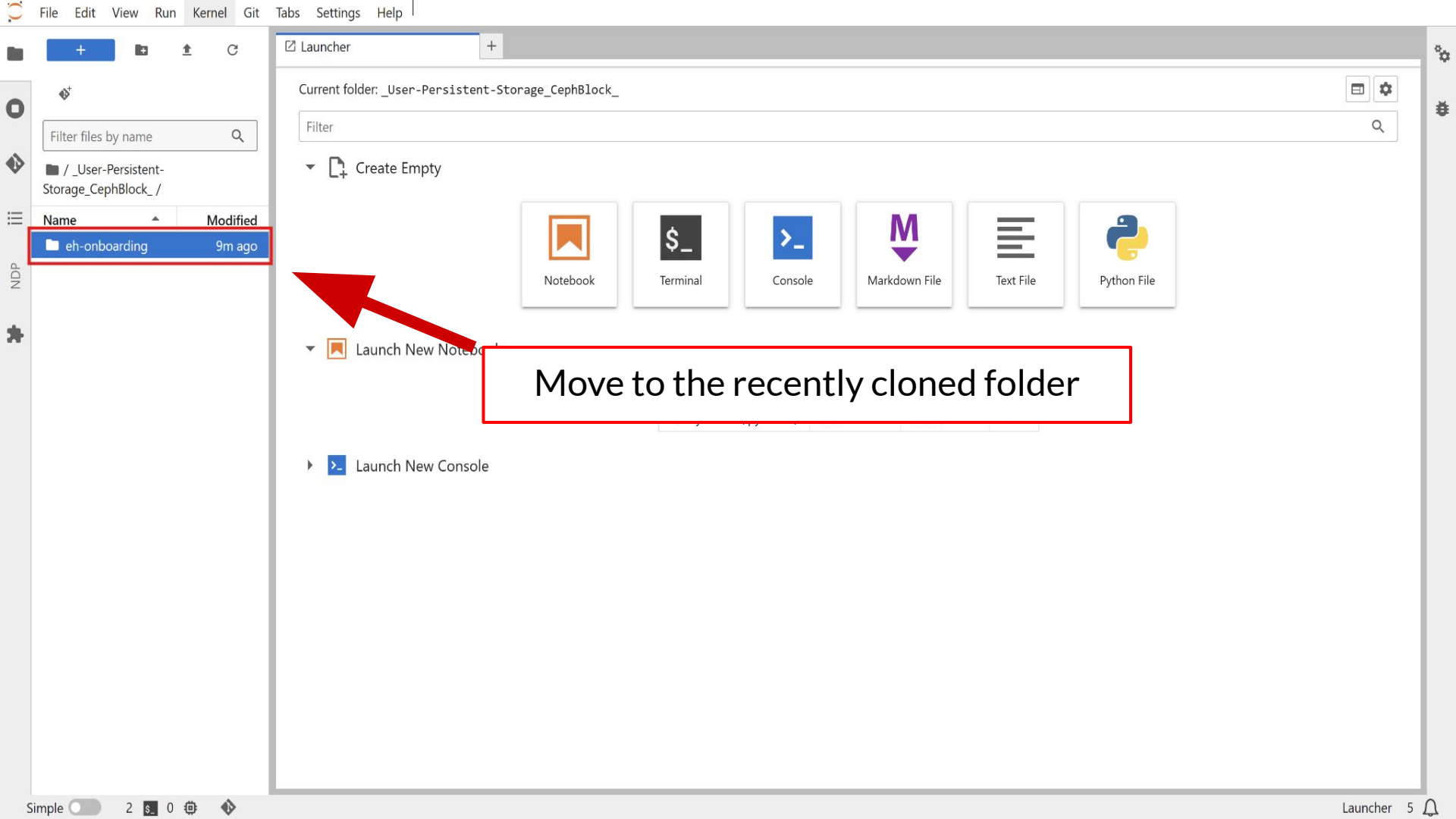
6. Clone repository

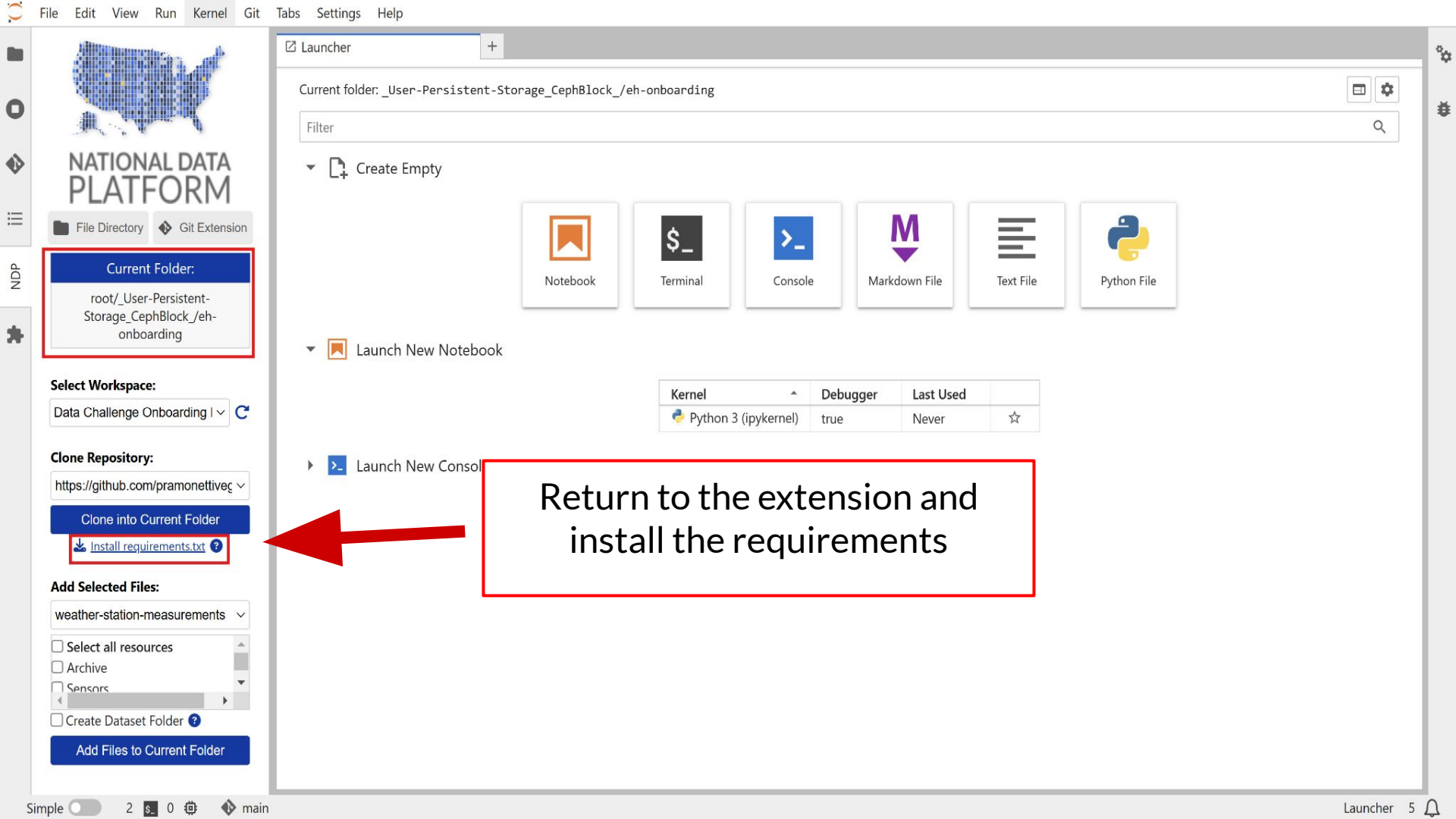

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_
Select Workspace:
Data Challenge Onboarding
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Select all resources
Archive
Sensors
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Click here to clone the repository
Simple
Launcher 4


File Edit View Run Kernel Git Tabs Settings Help
Launcher
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_
Select Workspace:
Data Challenge Onboarding I
Clone Repository:
https://github.com/pramonettiveç
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Select all resources
Archive
Sensors
Create Dataset Folder
Add Files to Current Folder
Current folder: _User-Persistent-Storage_CephBlock_
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Launch New Console
Clone a repo
Enter the URI of the remote Git repository
https://github.com/pramonet
Include submodules
Download the repository
Cancel
Clone
Keep the default and click *Clone*
Simple
Launcher 4

7. Install requirements


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
Click on *File Browser*
File Directory
Git Extensio
Current Folder:
root/_User-Persistent-Storage_CephBlock_
Select Workspace:
Data Challenge Onboarding
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Select all resources
Archive
Sensors
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Simple
Launcher 3


File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: _User-Persistent-Storage_CephBlock_
Filter files by name
/ _User-Persistent-Storage_CephBlock_ /
Name
Modified
eh-onboarding
9m ago
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Launch New Console
Move to the recently cloned folder
Simple
Launcher 5


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_/eh-onboarding
Select Workspace:
Data Challenge Onboarding
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Select all resources
Archive
Sensors
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_/eh-onboarding
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Consol
Return to the extension and install the requirements
Simple
2
0
main
Launcher
5

File Edit View Run Kernel Git Tabs Settings Help

NATIONAL DATA PLATFORM

File Directory | Git Extension

Current Folder:

root/_User-Persistent-Storage_CephBlock_/eh-onboarding

Select Workspace:

Data Challenge Onboarding I

Clone Repository:

https://github.com/pramonettiveg

Clone into Current Folder

Install requirements.txt

Add Selected Files:

weather-station-measurements

- [ ] Archive
- [ ] Sensors
- [x] San Diego Weather Sample

[ ] Create Dataset Folder

Add Files to Current Folder

Launcher | pip_install.log

```
Collecting pandas (from -r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1))
  Downloading pandas-2.2.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (89 kB)
Collecting matplotlib (from -r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 2))
  Downloading matplotlib-3.10.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (11 kB)
Collecting seaborn (from -r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 3))
  Downloading seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Collecting scikit-learn (from -r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 4))
  Downloading scikit_learn-1.6.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (18 kB)
Collecting numpy>=1.23.2 (from pandas->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1))
  Downloading numpy-2.2.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (62 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.11/site-packages (from pandas->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1)) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-packages (from pandas->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1)) (2024.1)
Collecting tzdata>=2022.7 (from pandas->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1))
  Downloading tzdata-2025.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting contourpy>=1.0.1 (from matplotlib->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 2))
  Downloading contourpy-1.3.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (5.4 kB)
Collecting cycler>=0.10 (from matplotlib->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 2))
  Downloadi
Collecting
  Downloadi
Collecting
  Downloadi
Collecting
  Downloadi
Requirement                                                                    CephBlock_/eh-onboarding/
Collecting
  Downloadi
Collecting
  Downloadi
Collecting
  Downloadi
Collecting joblib>=1.2.0 (from scikit-learn->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 4))
  Downloading joblib-1.4.2-py3-none-any.whl.metadata (5.4 kB)
Collecting threadpoolctl>=3.1.0 (from scikit-learn->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 4))
  Downloading threadpoolctl-3.5.0-py3-none-any.whl.metadata (13 kB)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-packages (from python-dateutil>=2.8.2->pandas->-r _User-Persistent-Storage_CephBlock_/eh-onboarding/requirements.txt (line 1)) (1.16.0)
```

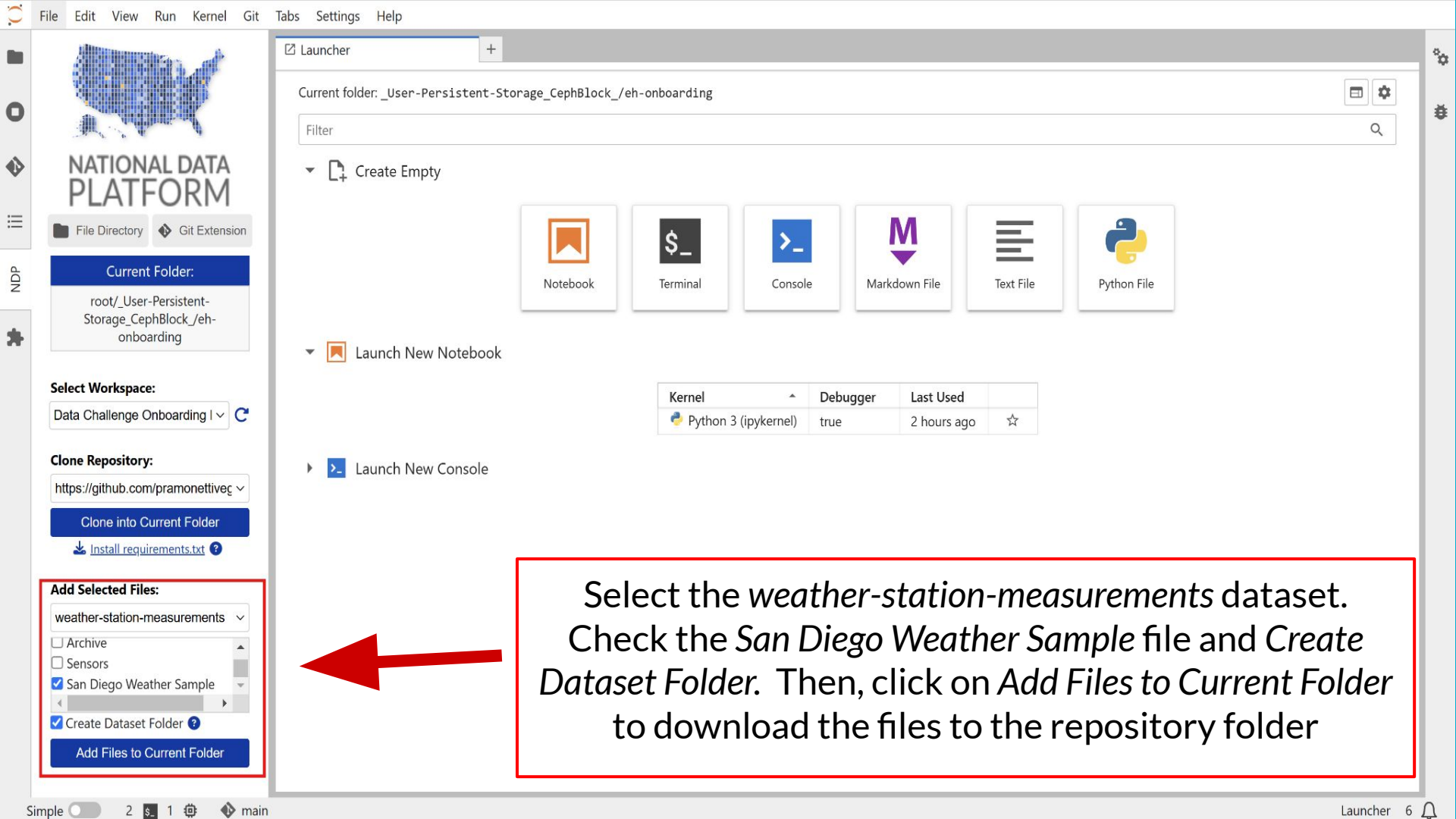
Wait for this log file to be displayed as a confirmation that your libraries installed successfully. In case it is not automatically generated, click *Install requirements.txt* one more time

Simple 2 1 none main Ln 1, Col 1 Spaces: 4 pip_install.log 6

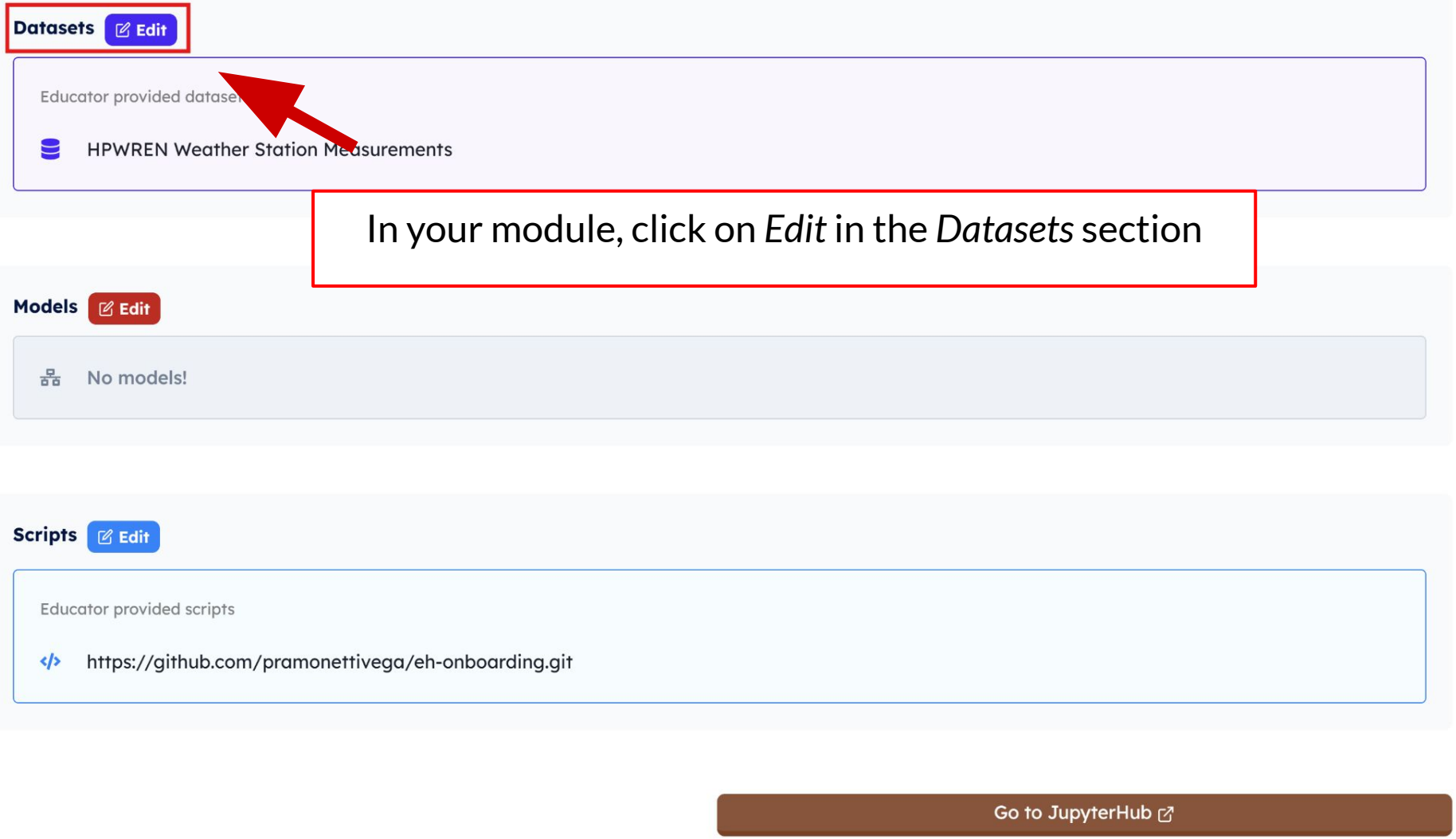
NOTE: Libraries are not persistent

Installing libraries will be effective for your current JupyterHub session. However, once you stop your server, your libraries will be lost. Therefore, each time you come back to work on a particular module, you need to reinstall the libraries.

8. Add selected files


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_/eh-onboarding
Select Workspace:
Data Challenge Onboarding I
Clone Repository:
https://github.com/pramonettiveç
Clone into Current Folder
Install requirements.txt
Add Selected Files:
weather-station-measurements
Archive
Sensors
San Diego Weather Sample
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_/eh-onboarding
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
2 hours ago
Launch New Console
Select the weather-station-measurements dataset. Check the San Diego Weather Sample file and Create Dataset Folder. Then, click on Add Files to Current Folder to download the files to the repository folder
Simple
main
Launcher 6

9. Add a dataset


Datasets
Edit
Educator provided dataset
HPWREN Weather Station Measurements
In your module, click on *Edit* in the *Datasets* section
Models
Edit
No models!
Scripts
Edit
Educator provided scripts
https://github.com/pramonettivega/eh-onboarding.git
Go to JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

Edit Datasets

Module datasets

HPWREN Weather Station...
TLS Fuel-Size Classification...

Search data catalog

lidar

Search | Clear

1 - 3 of 3243 Data Collections and Streams

1 2 ... 1081

TLS Fuel-Size Classification 2023

A collection of 34 Terrestrial Laser Scans (TLS) labelled with fuel size classes (1, 10, 100, 1000 hour fuels). Point clouds were initially preprocessed using the IntELiMon pipeline (https://dmsdata.cr.usgs.gov/lidar-monitoring/...

View More | Add

iDAR Survey of Dune Fields (Sept

NCALM Project. PI: Gary Kocurek, University of Texas at Austin. The survey area is a rectangular polygon on the White Sands National Monument, lying west of Holloman Air Force Base NM and enclosing 59 square...

View More | + Add

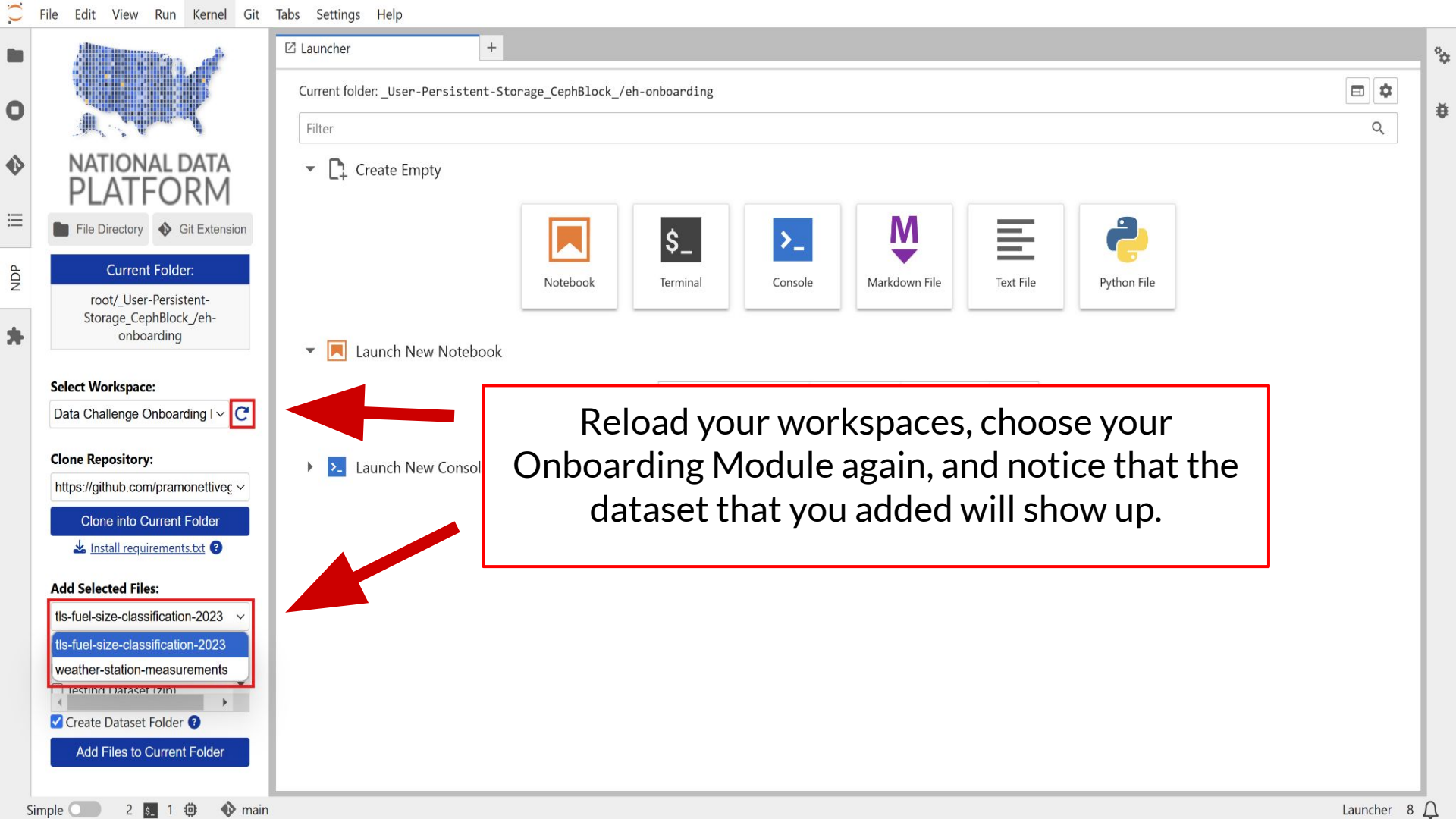
John Day Watershed 2006 - US Bureau of Reclamation

These Lidar data were collected on October 5-7, 2006 by Watershed Sciences Incorporated for the Puget Sound Lidar Consortium, with funding from the US Bureau of Reclamation and Oregon Trout in collaboration with the

View More

Search for the term *lidar* and add a dataset of your choice

https://github.com/pramonettivega/eh-onboarding.git


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/_User-Persistent-Storage_CephBlock_/eh-onboarding
Select Workspace:
Data Challenge Onboarding
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
tls-fuel-size-classification-2023
tls-fuel-size-classification-2023
weather-station-measurements
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: _User-Persistent-Storage_CephBlock_/eh-onboarding
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Launch New Consol
Reload your workspaces, choose your Onboarding Module again, and notice that the dataset that you added will show up.
Simple
2
1
main
Launcher
8

10. Complete your notebook

File Edit View Run Kernel Git Tabs Settings Help

Launcher | onboarding.ipynb

Markdown | git | Open in... | Python 3 (ipykernel)

Filter files by name

/ _User-Persistent-Storage_CephBlock_ / eh-onboarding /

Name	Modified
pics	5h ago
weather-station-...	51m ago
onboarding.ipynb	1m ago
README.md	5h ago
requirements.txt	5h ago

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

NATIONAL DATA PLATFORM
Bridging the Data Gaps for AI

Open the onboarding notebook and complete it

This module is designed to provide an onboarding experience and introduce you to working with NDP Modules. The problem and dataset presented here align with research areas explored by multiple collaborators of the National Data Platform, including the WORDS team at the San Diego Supercomputer Center.

The problem and data used in this demo module were originally developed as part of the Big Data Specialization offered by UC San Diego on Coursera.

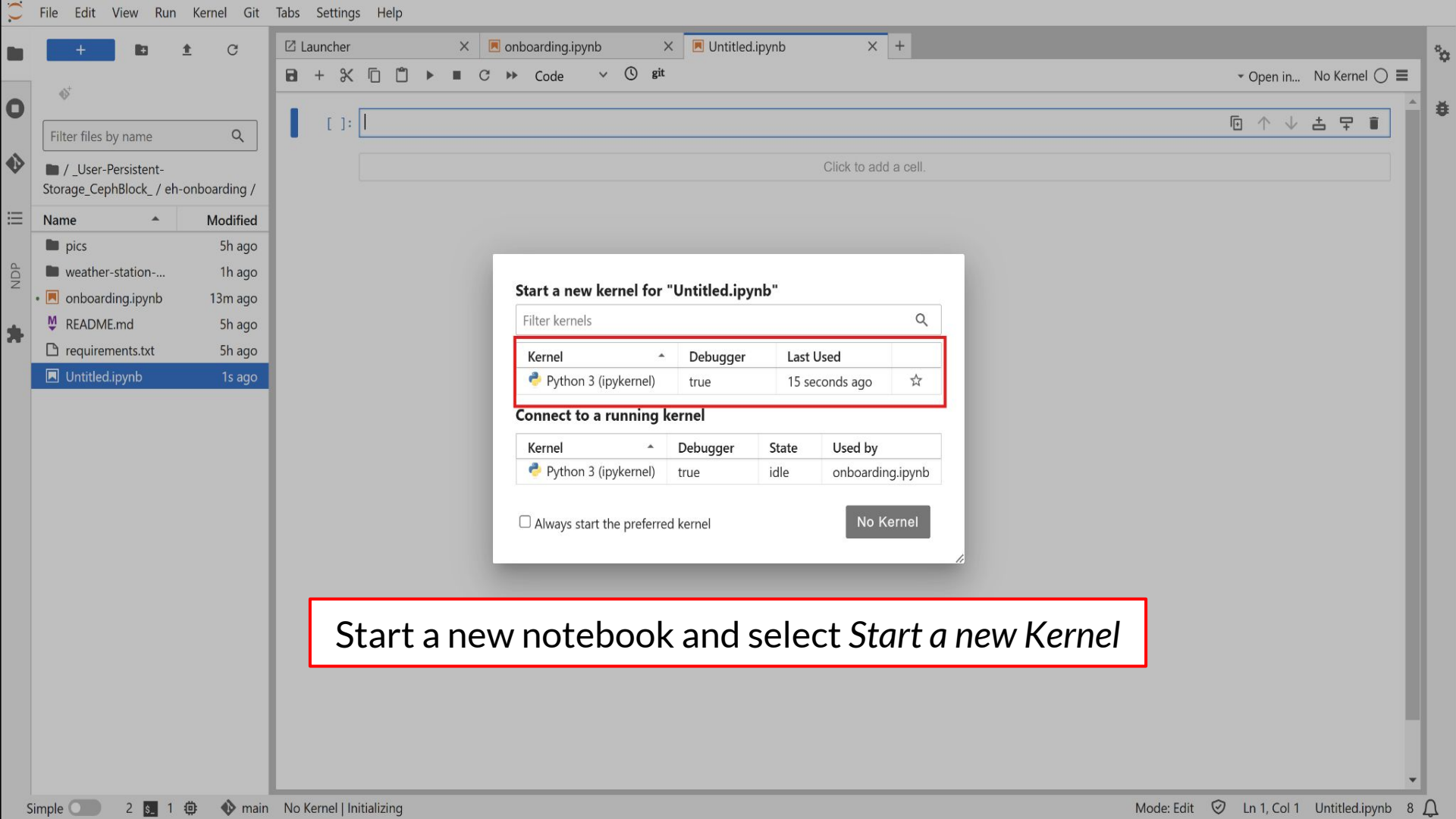
The Data

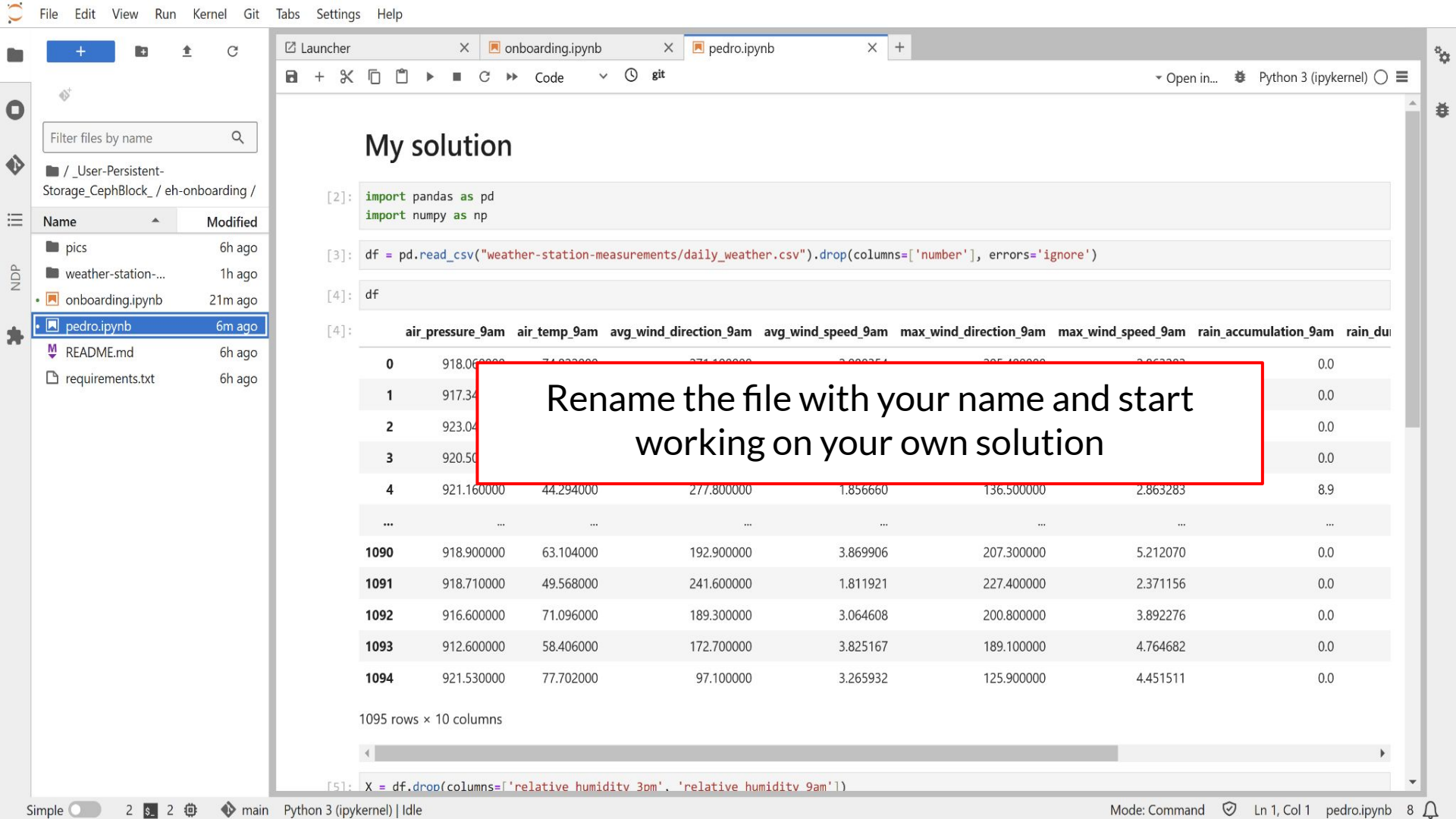
The file `daily_weather.csv` is a comma-separated file that contains weather data. This data comes from a weather station located in San Diego, California. The weather station is equipped with sensors that capture weather-related measurements such as air temperature, air pressure, and relative humidity. Data was collected for a period of three years, from September 2011 to September 2014, to ensure that sufficient data for different seasons and weather conditions is captured.

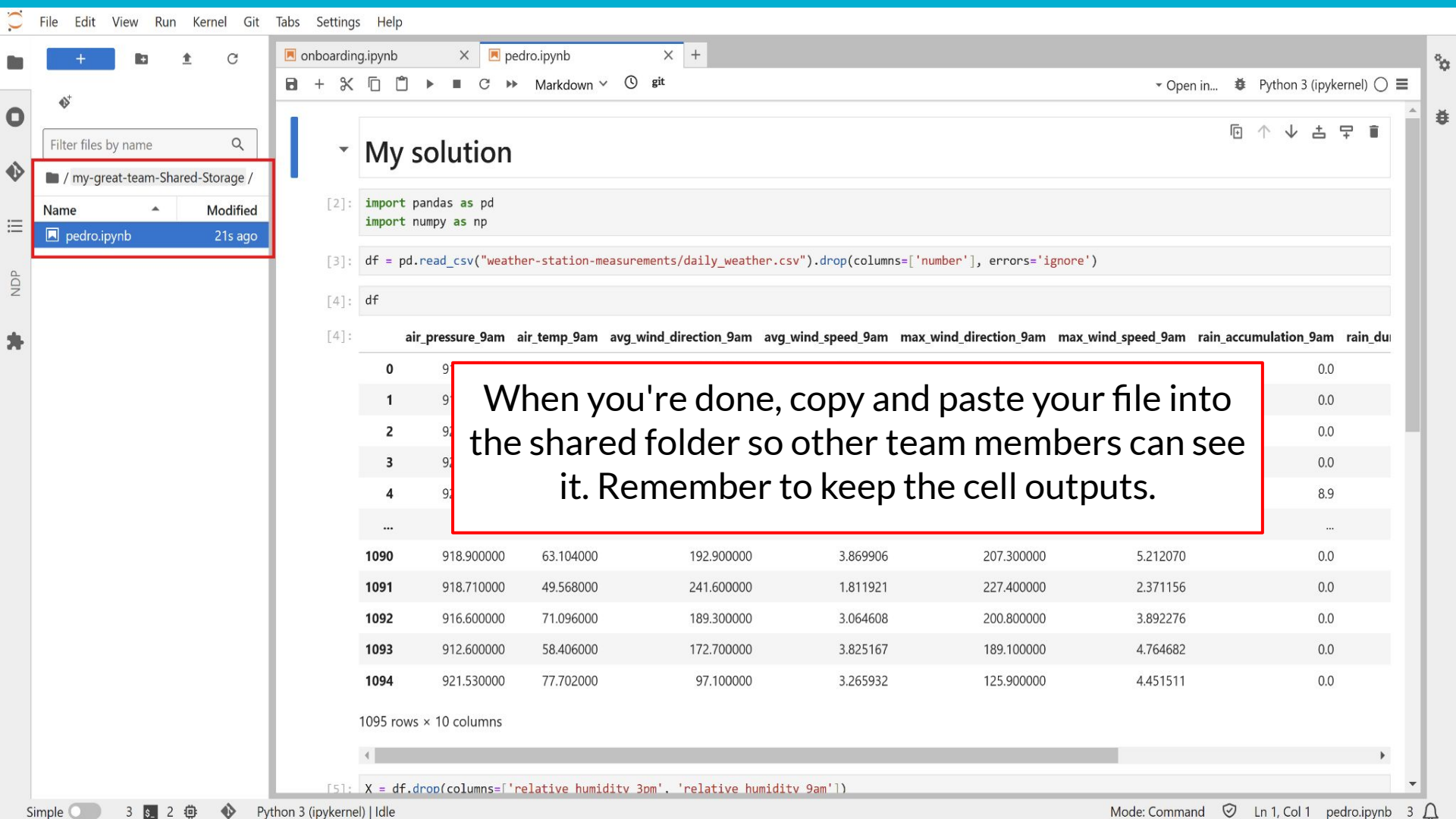
Sensor measurements from the weather station were captured at one-minute intervals. These measurements were then processed to generate values to describe daily weather. Since this dataset was created to classify low-humidity days vs. non-low-humidity days (that is, days with normal or high humidity), the variables included are weather measurements in the morning, with one measurement, namely relatively humidity, in the afternoon.

Simple | 2 | 1 | main | Python 3 (ipykernel) | Idle | Mode: Command | Ln 1, Col 1 | onboarding.ipynb | 8

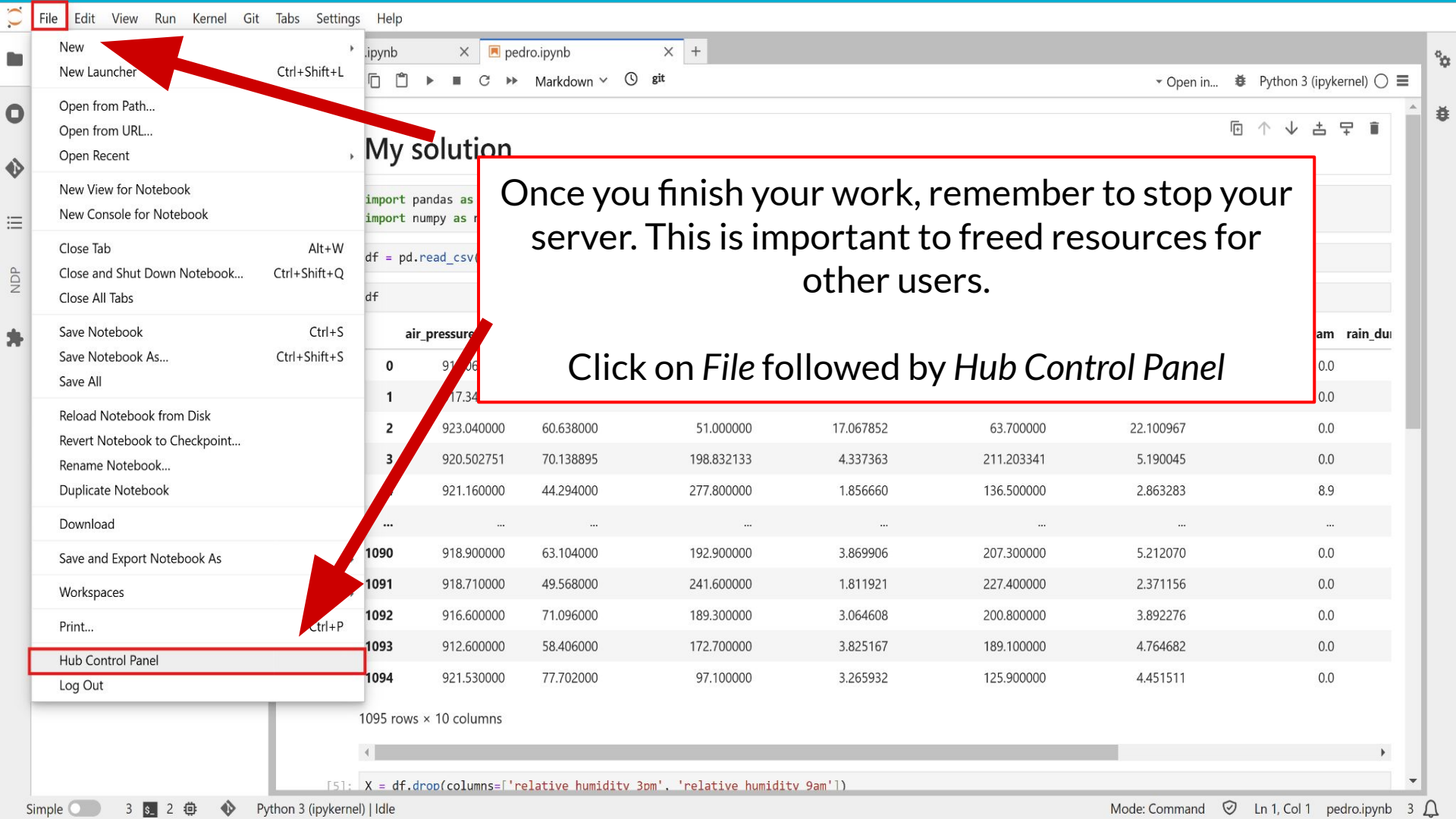
11. Work on your own solution


File Edit View Run Kernel Git Tabs Settings Help
Launcher
onboarding.ipynb
Untitled.ipynb
Code
git
Open in...
No Kernel
[ ]:
Click to add a cell.
Filter files by name
/ _User-Persistent-Storage_CephBlock_ / eh-onboarding /
Name
Modified
pics 5h ago
weather-station-... 1h ago
onboarding.ipynb 13m ago
README.md 5h ago
requirements.txt 5h ago
Untitled.ipynb 1s ago
Start a new kernel for "Untitled.ipynb"
Filter kernels
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
15 seconds ago
Connect to a running kernel
Kernel
Debugger
State
Used by
Python 3 (ipykernel)
true
idle
onboarding.ipynb
Always start the preferred kernel
No Kernel
Start a new notebook and select *Start a new Kernel*
Simple
main
No Kernel | Initializing
Mode: Edit
Ln 1, Col 1
Untitled.ipynb

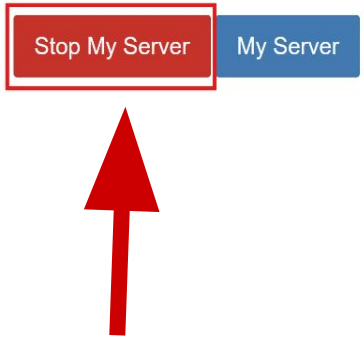

File Edit View Run Kernel Git Tabs Settings Help
Launcher
onboarding.ipynb
pedro.ipynb
Code
git
Open in...
Python 3 (ipykernel)
Filter files by name
/ _User-Persistent-Storage_CephBlock_ / eh-onboarding /
Name
Modified
pics 6h ago
weather-station-... 1h ago
onboarding.ipynb 21m ago
pedro.ipynb 6m ago
README.md 6h ago
requirements.txt 6h ago
My solution
[2]: import pandas as pd
import numpy as np
[3]: df = pd.read_csv("weather-station-measurements/daily_weather.csv").drop(columns=['number'], errors='ignore')
[4]: df
[4]:
air_pressure_9am air_temp_9am avg_wind_direction_9am avg_wind_speed_9am max_wind_direction_9am max_wind_speed_9am rain_accumulation_9am rain_du
0 0.0
1 0.0
2 0.0
3 0.0
4 921.160000 44.294000 277.800000 1.856660 136.500000 2.863283 8.9
... ...
1090 918.900000 63.104000 192.900000 3.869906 207.300000 5.212070 0.0
1091 918.710000 49.568000 241.600000 1.811921 227.400000 2.371156 0.0
1092 916.600000 71.096000 189.300000 3.064608 200.800000 3.892276 0.0
1093 912.600000 58.406000 172.700000 3.825167 189.100000 4.764682 0.0
1094 921.530000 77.702000 97.100000 3.265932 125.900000 4.451511 0.0
1095 rows × 10 columns
Rename the file with your name and start working on your own solution
[5]: X = df.drop(columns=['relative_humidity_3pm', 'relative_humidity_9am'])
Simple
main
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
pedro.ipynb


File Edit View Run Kernel Git Tabs Settings Help
onboarding.ipynb
pedro.ipynb
Markdown
git
Open in...
Python 3 (ipykernel)
Filter files by name
/ my-great-team-Shared-Storage /
Name
Modified
pedro.ipynb
21s ago
My solution
[2]:
import pandas as pd
import numpy as np
[3]:
df = pd.read_csv("weather-station-measurements/daily_weather.csv").drop(columns=['number'], errors='ignore')
[4]:
df
[4]:
air_pressure_9am
air_temp_9am
avg_wind_direction_9am
avg_wind_speed_9am
max_wind_direction_9am
max_wind_speed_9am
rain_accumulation_9am
1090 918.900000 63.104000 192.900000 3.869906 207.300000 5.212070 0.0
1091 918.710000 49.568000 241.600000 1.811921 227.400000 2.371156 0.0
1092 916.600000 71.096000 189.300000 3.064608 200.800000 3.892276 0.0
1093 912.600000 58.406000 172.700000 3.825167 189.100000 4.764682 0.0
1094 921.530000 77.702000 97.100000 3.265932 125.900000 4.451511 0.0
1095 rows × 10 columns
When you're done, copy and paste your file into the shared folder so other team members can see it. Remember to keep the cell outputs.
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
pedro.ipynb

12. Stop your server



	air_pressure						rain_du
0	91						0.0
1	17.34						0.0
2	923.040000	60.638000	51.000000	17.067852	63.700000	22.100967	0.0
3	920.502751	70.138895	198.832133	4.337363	211.203341	5.190045	0.0
	921.160000	44.294000	277.800000	1.856660	136.500000	2.863283	8.9
...	...	...	...	...	...	...	...
1090	918.900000	63.104000	192.900000	3.869906	207.300000	5.212070	0.0
1091	918.710000	49.568000	241.600000	1.811921	227.400000	2.371156	0.0
1092	916.600000	71.096000	189.300000	3.064608	200.800000	3.892276	0.0
1093	912.600000	58.406000	172.700000	3.825167	189.100000	4.764682	0.0
1094	921.530000	77.702000	97.100000	3.265932	125.900000	4.451511	0.0



And finally click on *Stop Server*. Wait for the red button to disappear as a confirmation.

FINAL NOTE

If you return to this module at some point, you will only need to complete steps 2, 3, 4, and 7 of the setup because your files will be stored in your shared folder, so you will not need to clone the repository or download the data files again.