Software Environment Setup

Machine Learning in Water and Environmental Modeling
Pre-Workshop Setup Meeting, May 1, 2025

Brad Tom and Peyman Namadi
DWR Delta Modeling Section

Part A: Local Computer

Prerequisites

- Laptop or Desktop with Windows*
- 6GB free space
- Administrative privileges

* The ML code will also run on Mac OS, Linux, and Google Colab

Overview

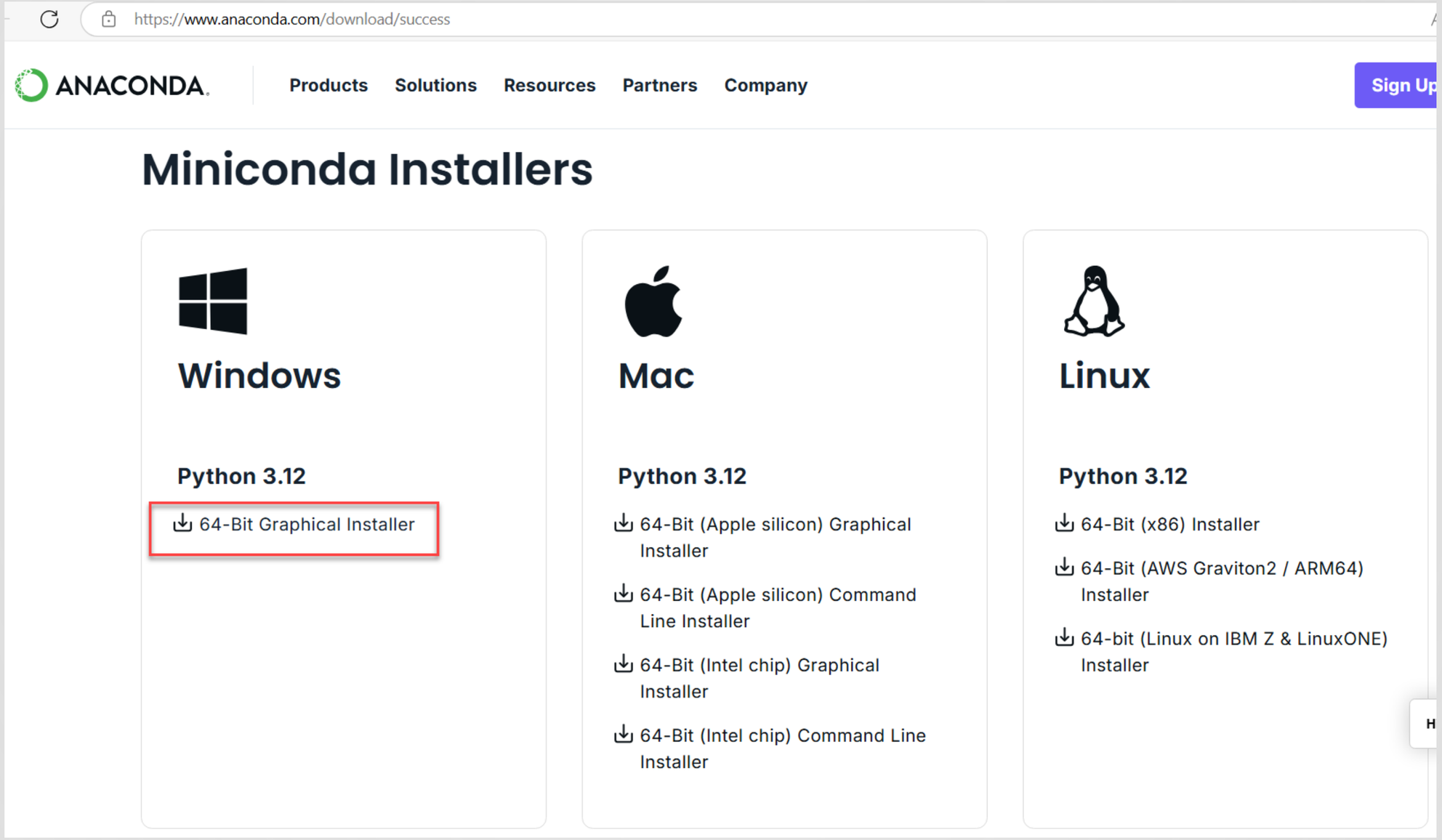
- Install Miniconda[1]
- Create a conda environment[2]
- Download ANN code and input data files
- Run the Jupyter Notebook[3] ANN code

1. Anaconda is a free, open-source platform that allows you to write and execute code in the Python programming language. Miniconda is a lightweight version of Anaconda. We recommend Miniconda because installation and setup take less time.
2. A conda environment is a folder containing a version of Python, and a specific set of python packages
3. Jupyter Notebook is a web application for creating and sharing computational documents

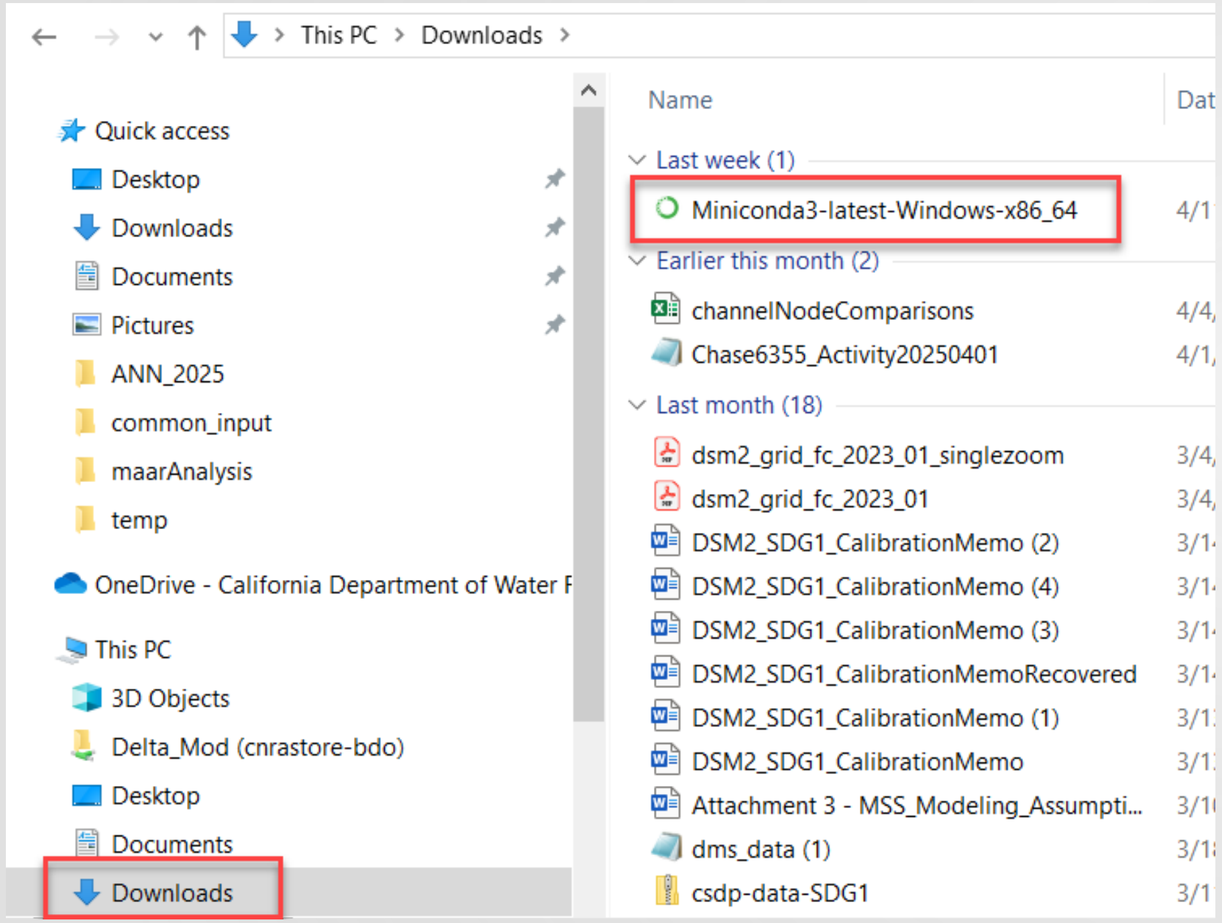
Install Miniconda/Anaconda

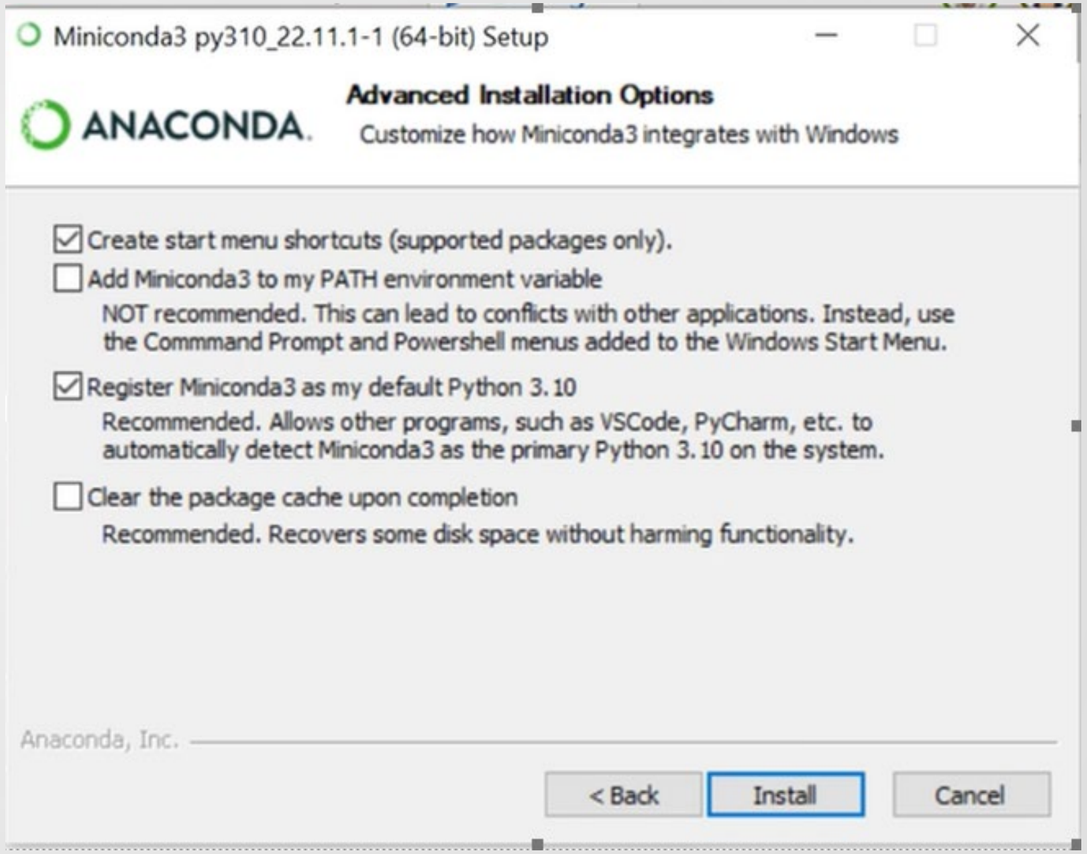
- Miniconda or Anaconda: https://www.anaconda.com/download/success

Downloading the Miniconda Installer



Install Miniconda





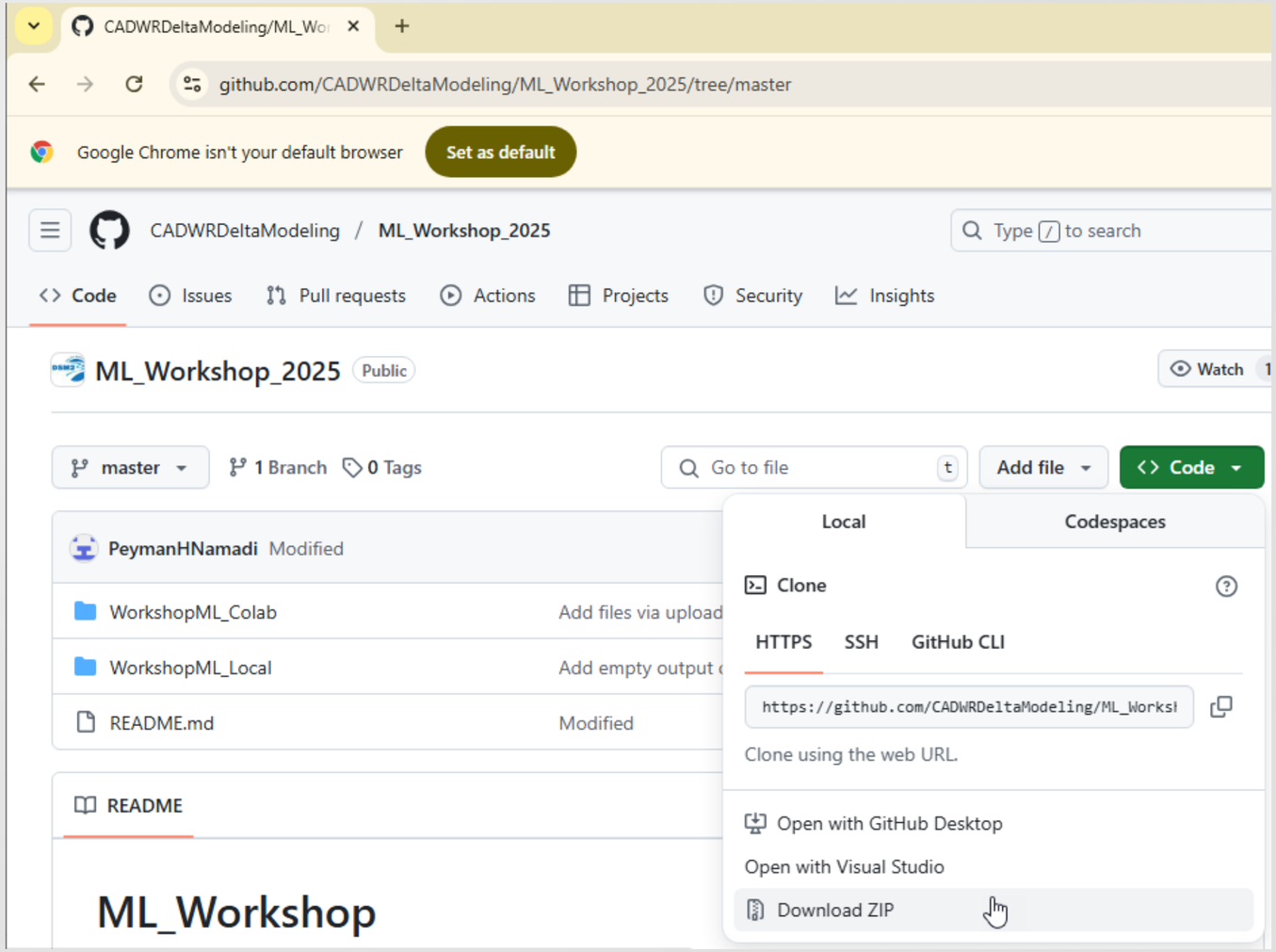
Download ML Code/Data

a) If you don't have Git* installed

- https://github.com/CADWRDeltaModeling/ML_Workshop_2025/

Click "Code", then "Download ZIP"

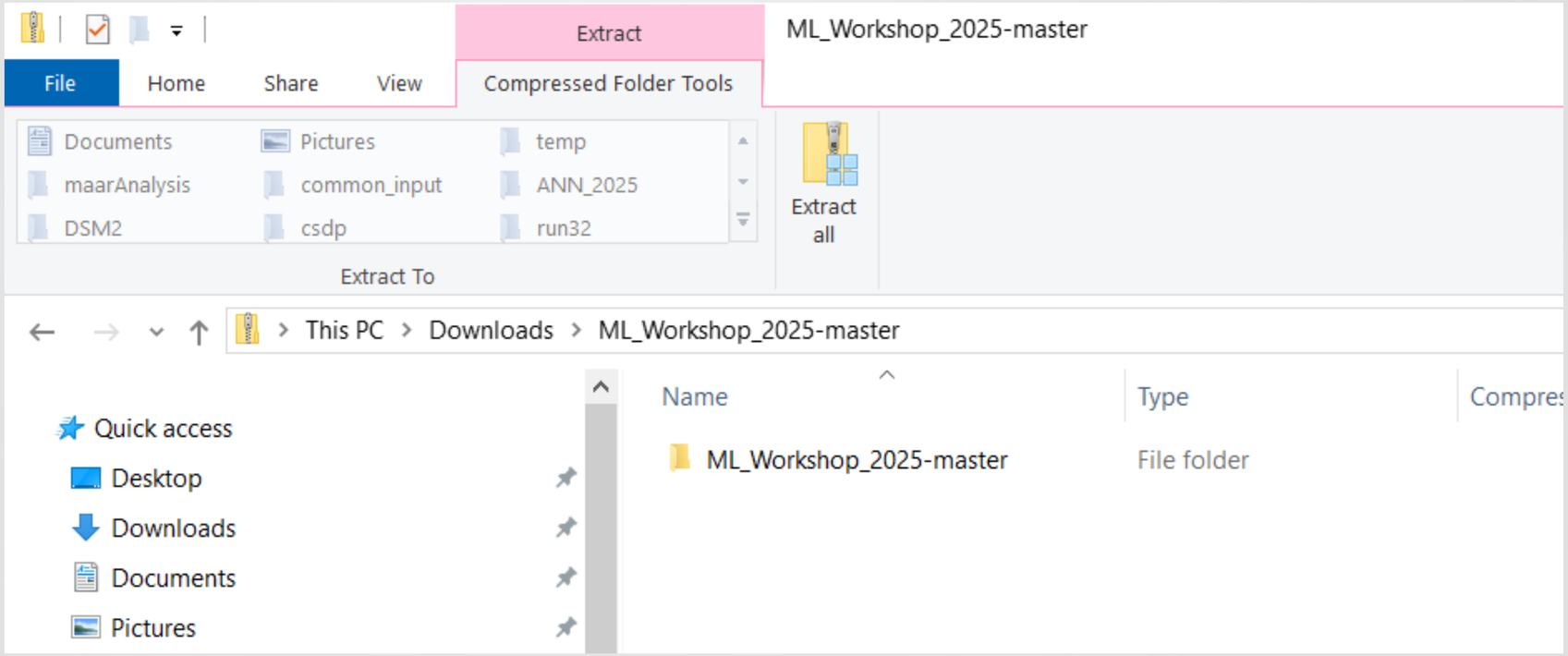
Git* is a free and open source distributed version control system



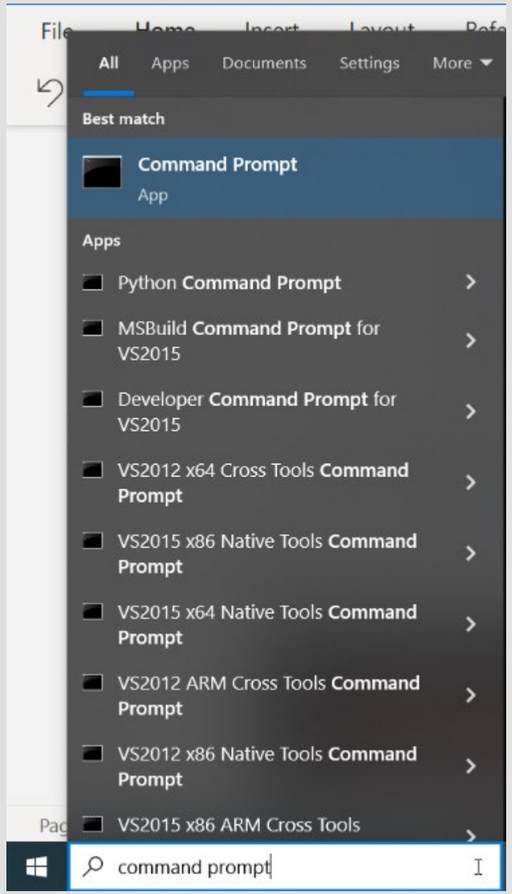
Download ML Code/Data

a) If you <u>don't have</u> Git installed (cont)

- Double click the zip file
- Copy and paste the folder inside to another location



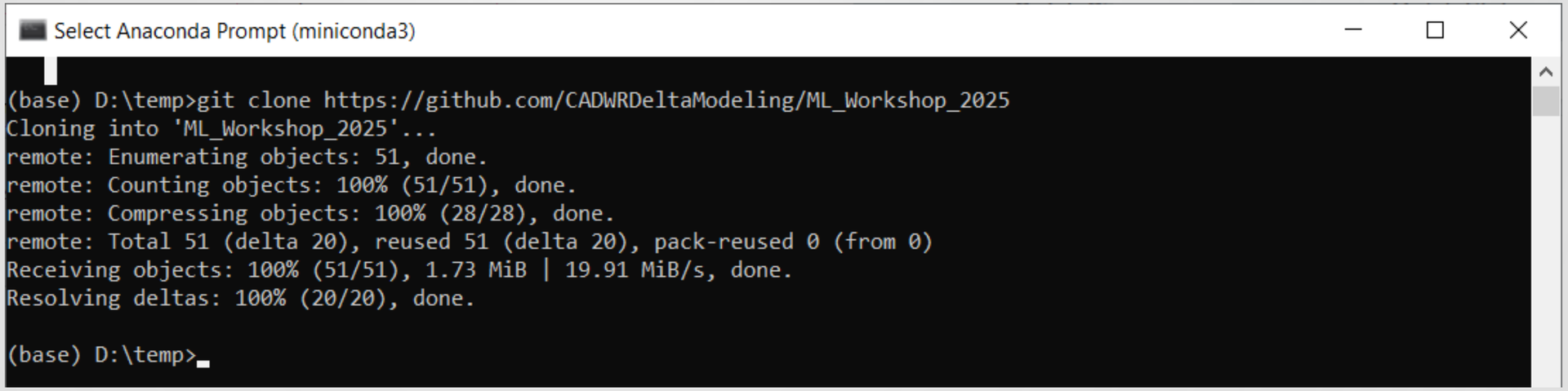
Download ML Code/Data



b) If you <u>have</u> Git installed

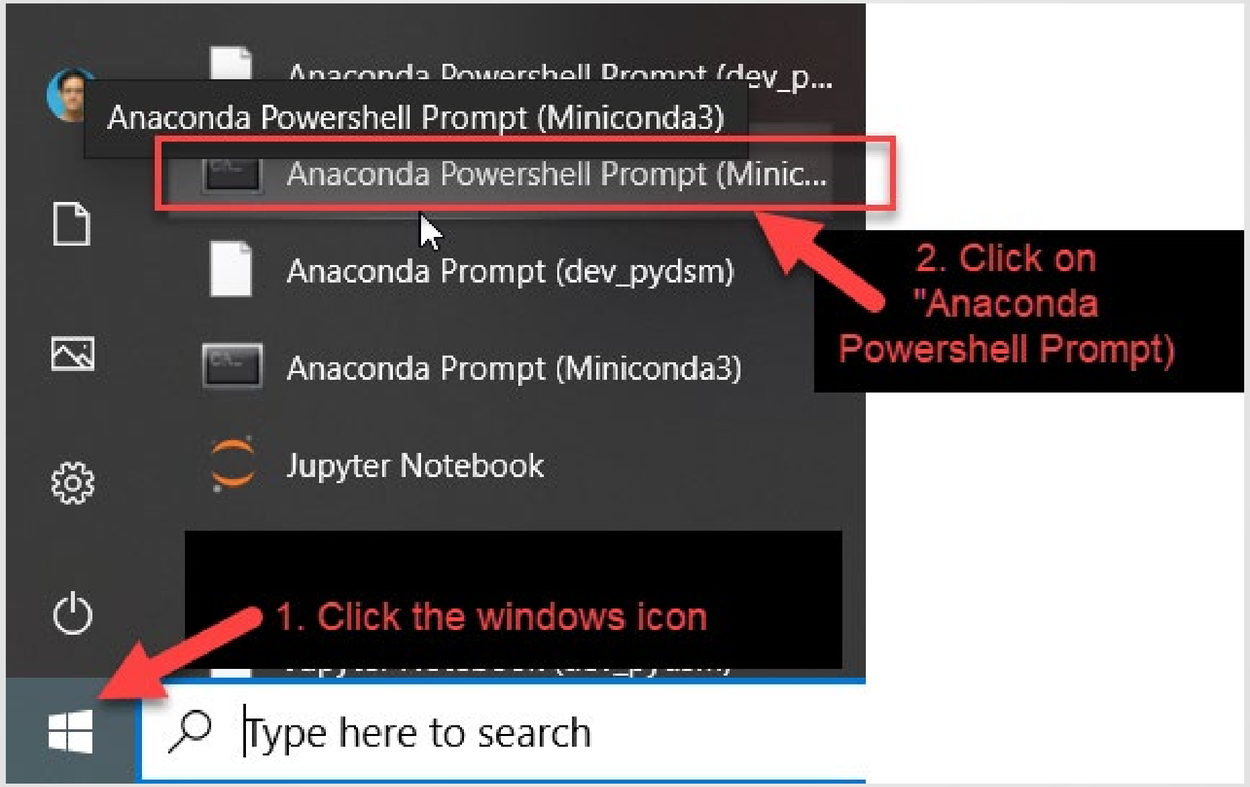
Open a command prompt window

Navigate to the location where you want to save files



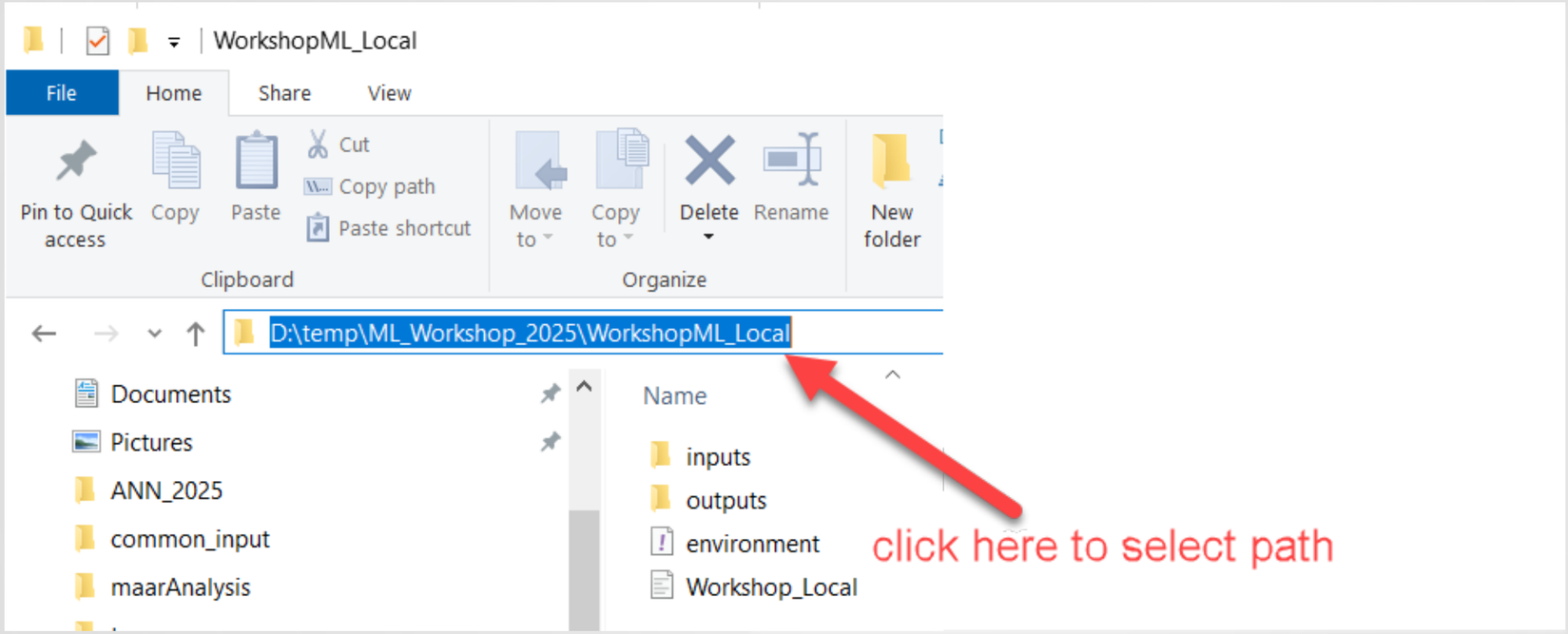
git clone https://github.com/CADWRDeltaModeling/ML_Workshop_2025

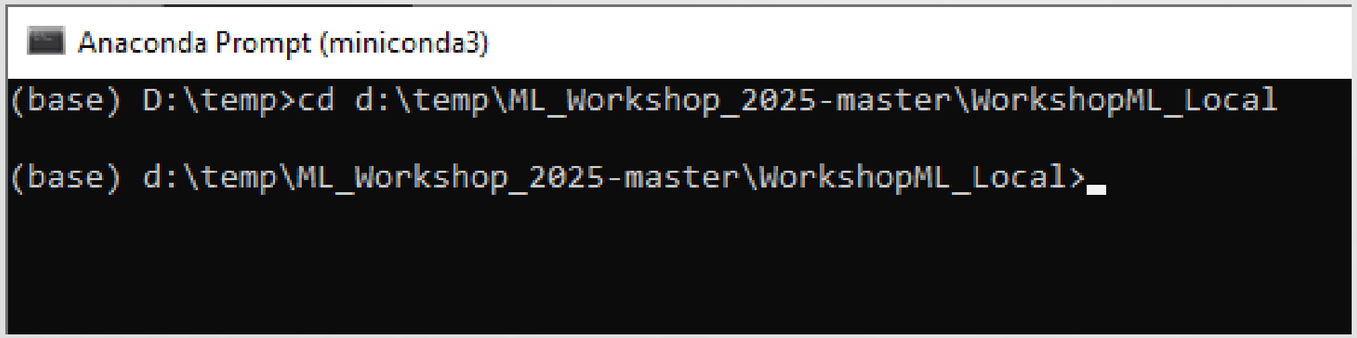
Open an Anaconda Powershell prompt



Create a conda environment

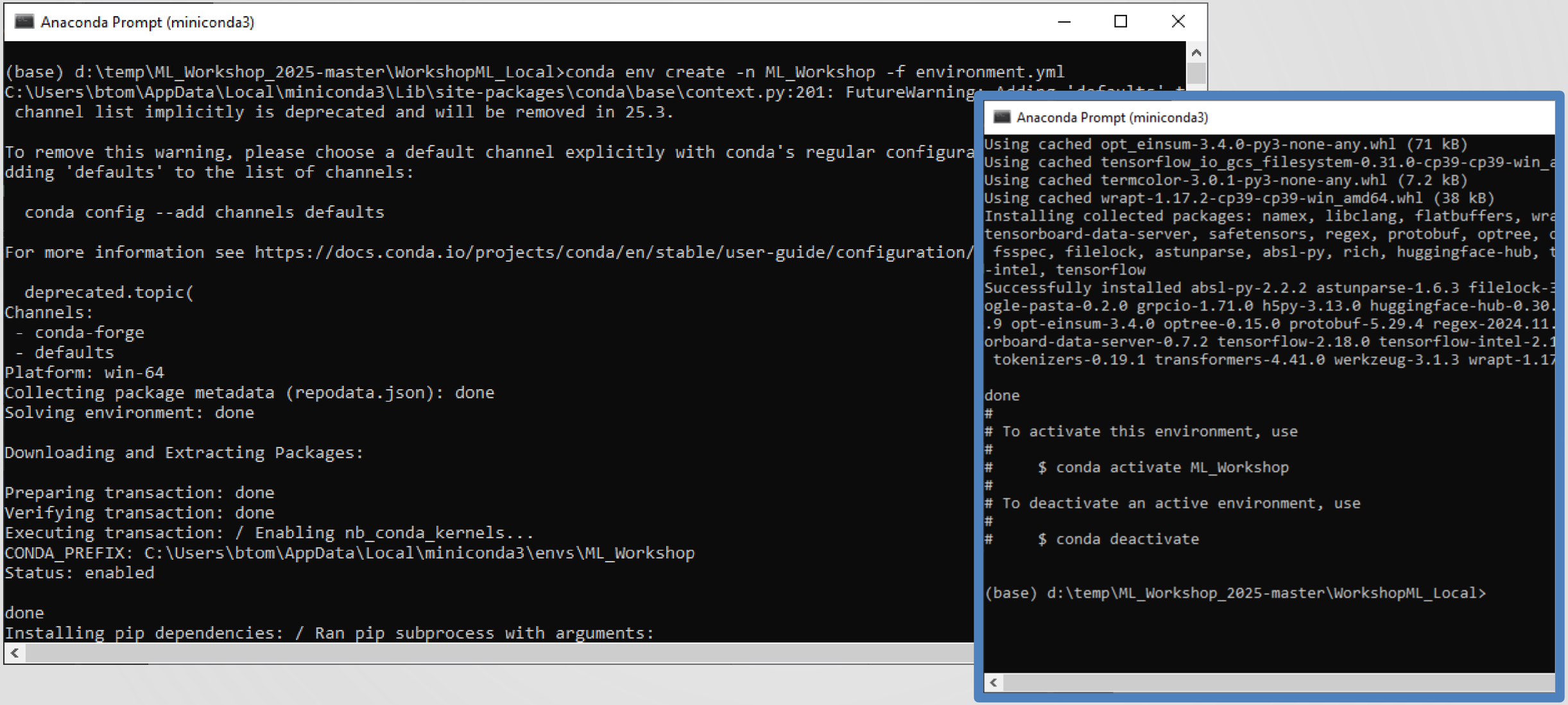
- Use the “cd” command to navigate to the folder containing your code/data





Create a conda environment

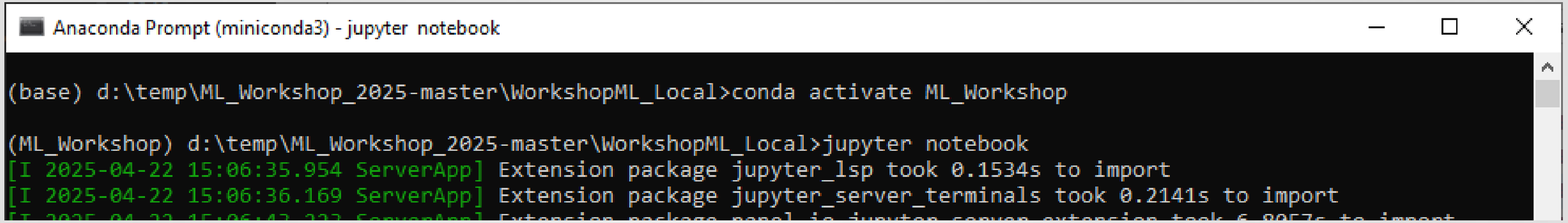
- conda env create -n ML_Workshop -f environment.yml



Open Jupyter Notebook

conda activate ML_Workshop

jupyter notebook



If the above command fails, try these two commands:

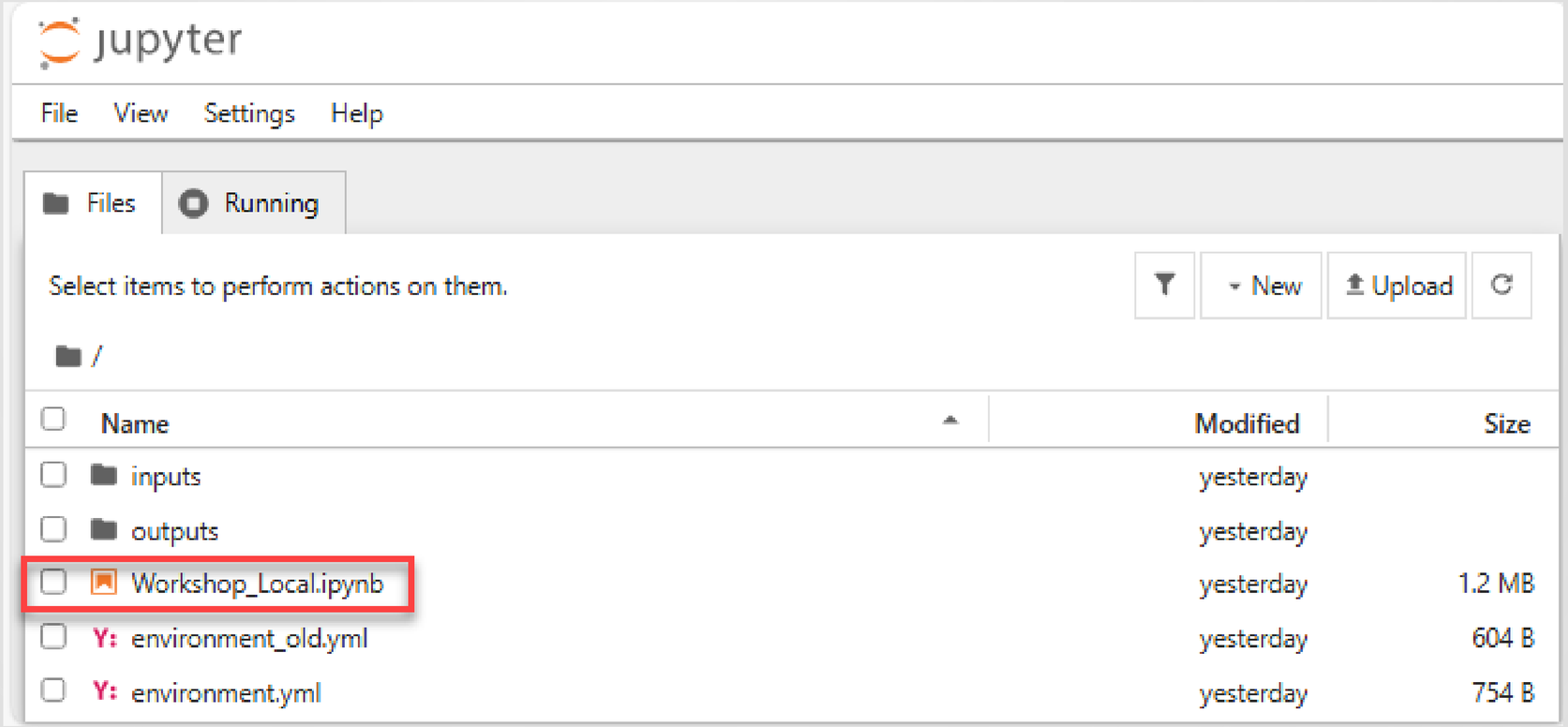
call conda activate jupyter

start jupyter notebook

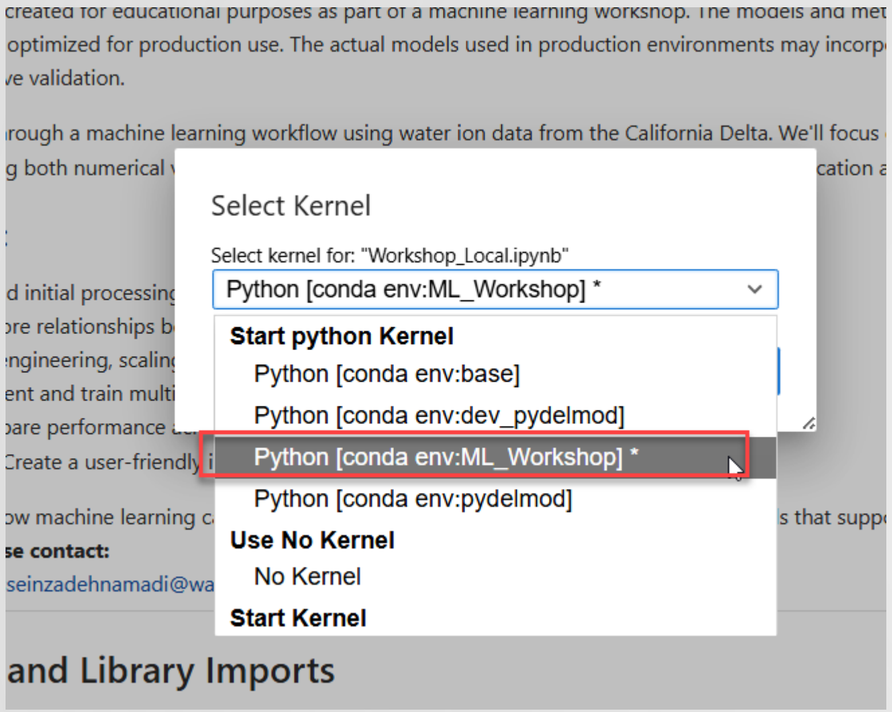
Open the ANN Jupyter Notebook

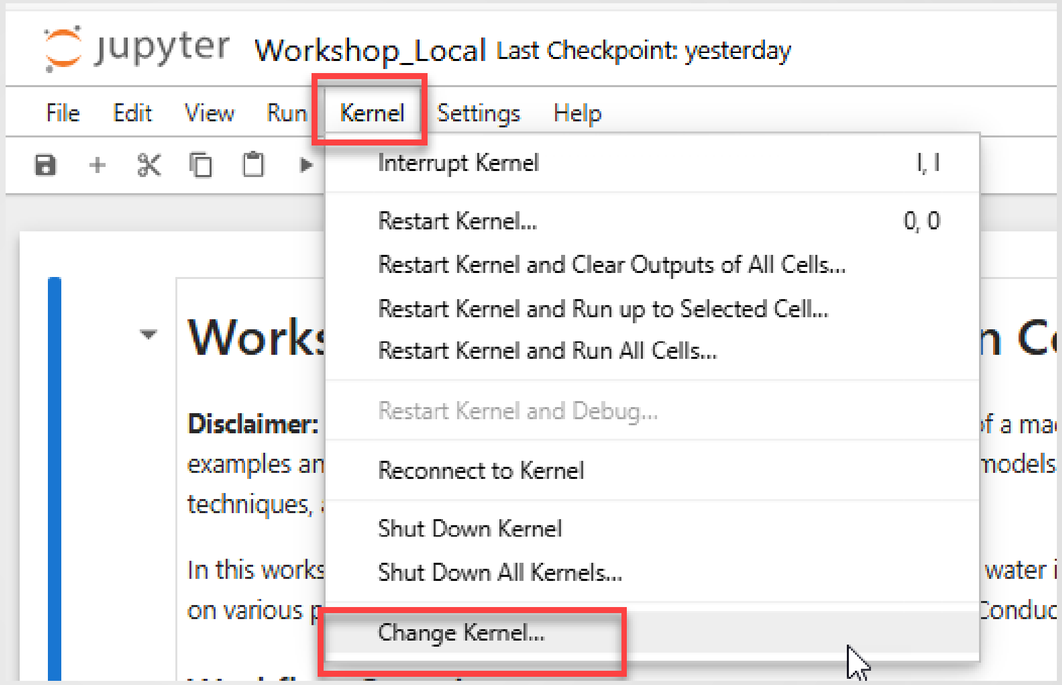
click on

Workshop_Local.ipynb

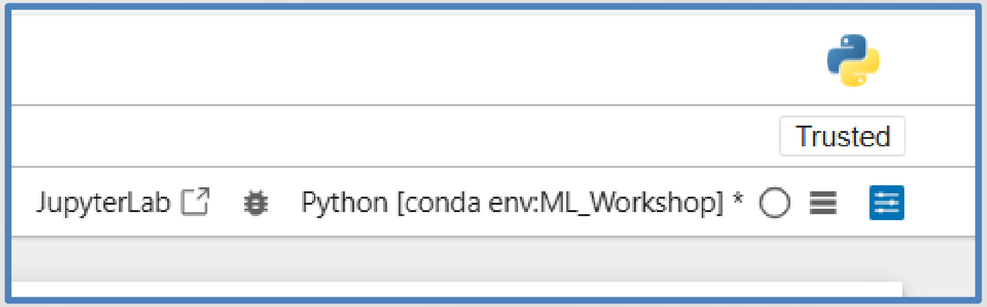


Select notebook kernel

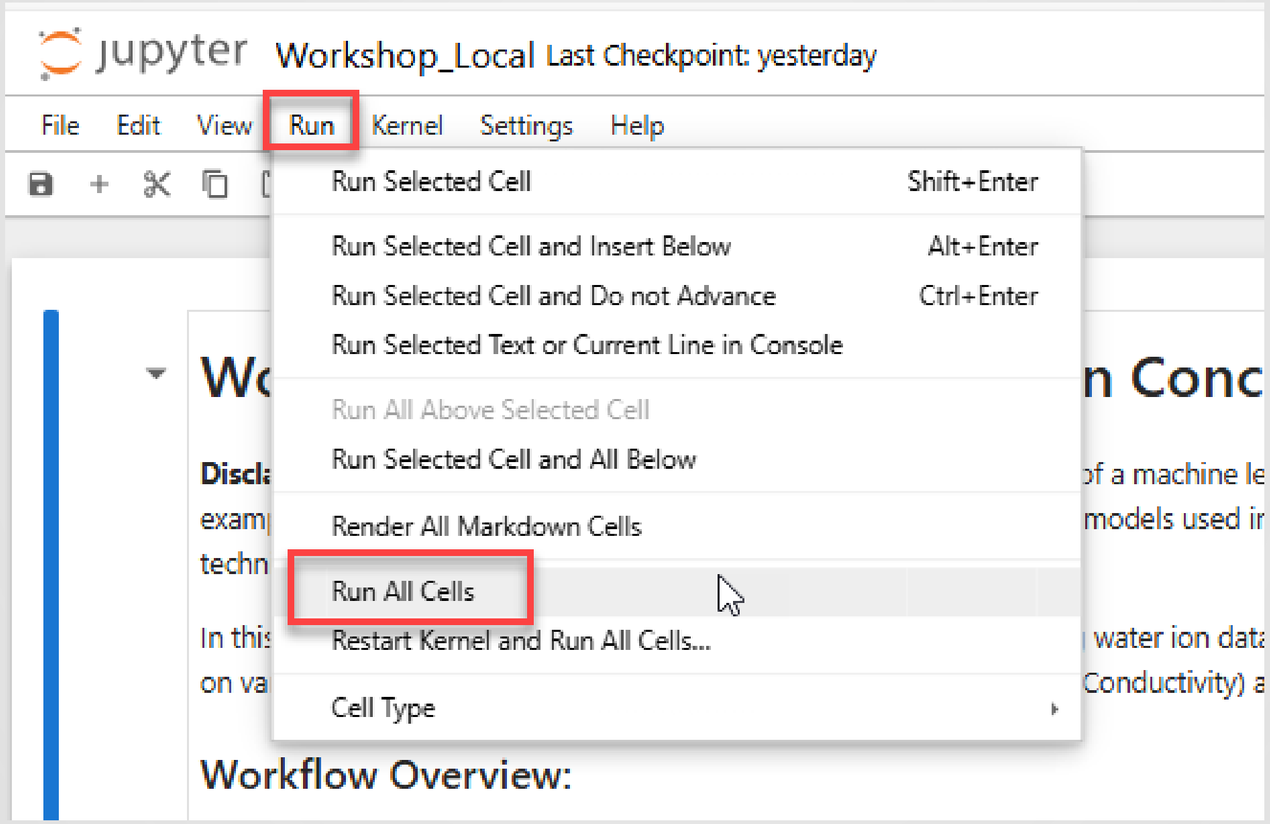




A notebook kernel is **a** “computational engine” that executes the code contained in a Notebook document.



Run the notebook code



Run the notebook code

Br Concentration Prediction Dashboard

This dashboard uses three machine learning models to predict Br concentration based on environmental parameters. Adjust the sliders and dropdown menus to see how different conditions affect predicted Br levels.

Input Parameters

Electrical Conductivity (EC) [µS/cm]: **1000**

Sacramento X2: **80**

Location

OMR

Water Year Type (WYT)

AN

Month

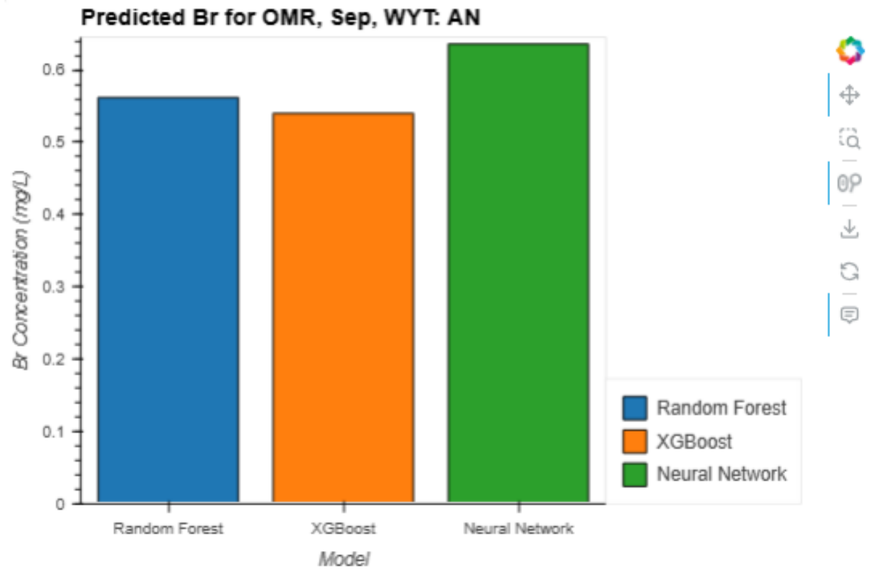
Sep

Predictions updated

Prediction Results

Predicted Br Concentration

Model	Br (mg/L)
Random Forest	0.56
XGBoost	0.54
Neural Network	0.64



Completed cell vs running cell

Section 0: Setup and Library Imports

[1]:
```
# ================= SECTION 0: IMPORT LIBRARIES =================
# Core data handling and analysis
import pandas as pd
import numpy as np
import os
import joblib
from datetime import datetime
```

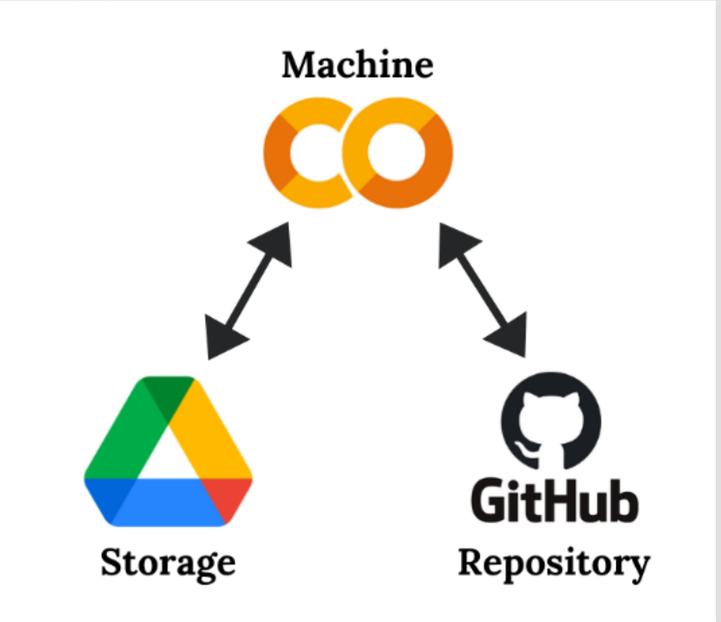
1. Data Loading

[*]:
```
import pandas as pd
import os

# Define the path to the inputs folder
input_folder = 'inputs'

# List of ion files to load
```

Part B:

Advantages of Google Colab:

- Free GPU/TPU access (to a certain limit)
- No setup required on your computer
- Real-time collaboration



Limitations of Google Colab:

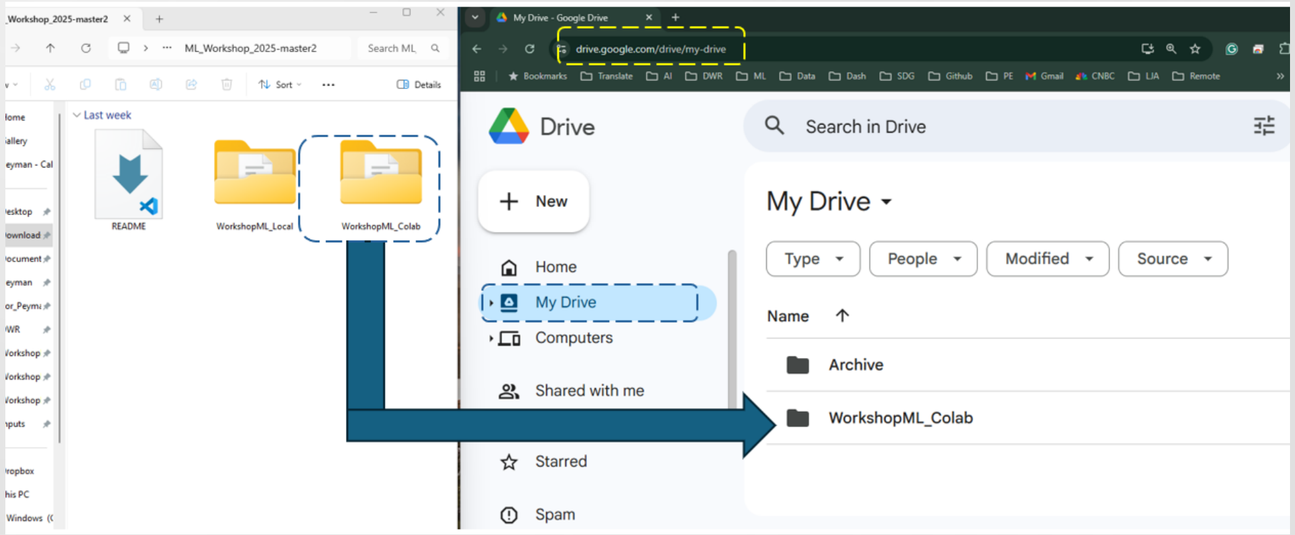
- **Session time limits**: Free → 12-hour maximum runtime; paid tiers → 24 hours
- **Limited persistent storage**: Files not saved to Google Drive are deleted when sessions end
- **Inconsistent performance**: GPU/TPU availability and performance vary based on usage demand
- **Resource constraints**: Limited RAM (typically 12-25GB) & storage capacity, especially on free tier

Step 1: Create a **Google account** or use an existing Google account and log in to your **Google Drive** at https://drive.google.com/drive/my-drive

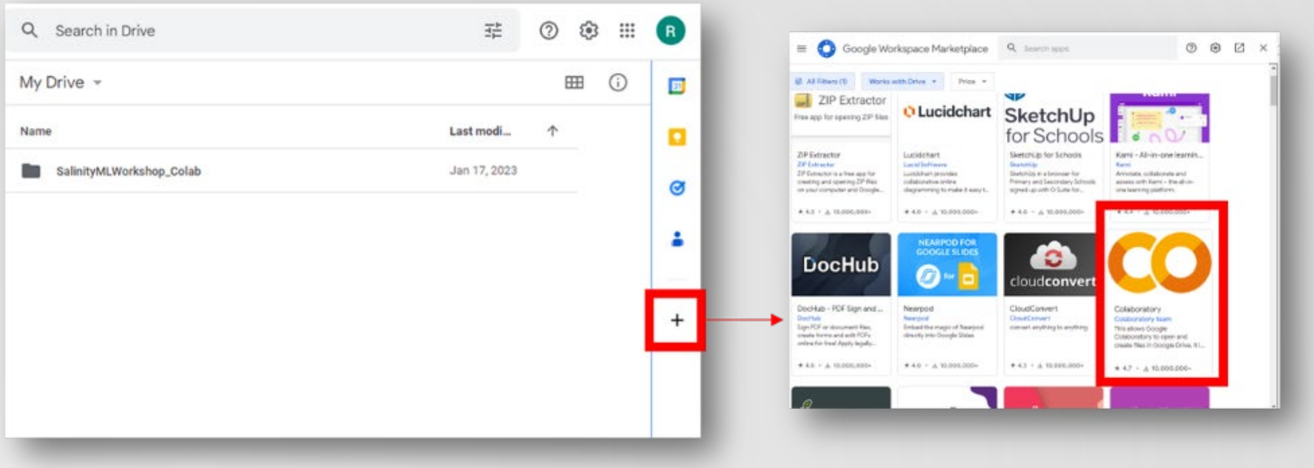
Step 2: Use this link to download the folder containing data and code: https://github.com/CADWRDeltaModeling/ML_Workshop_2025/tree/master

Step 3: Navigate to your download location and **unzip** the folder

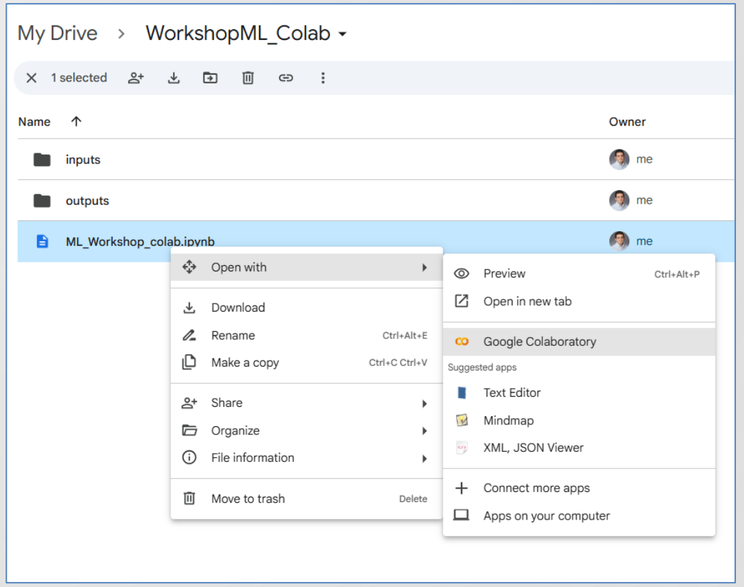
Step 4: After unzipping the file, you'll see two folders: "WorkshopML_Local" and "WorkshopML_Colab". Take the "**WorkshopML_Colab**" folder and upload it directly to the main area of your Google Drive (the "My Drive" section). Make sure you're uploading the actual folder with the workshop materials, not a parent folder that contains another folder with the same name.



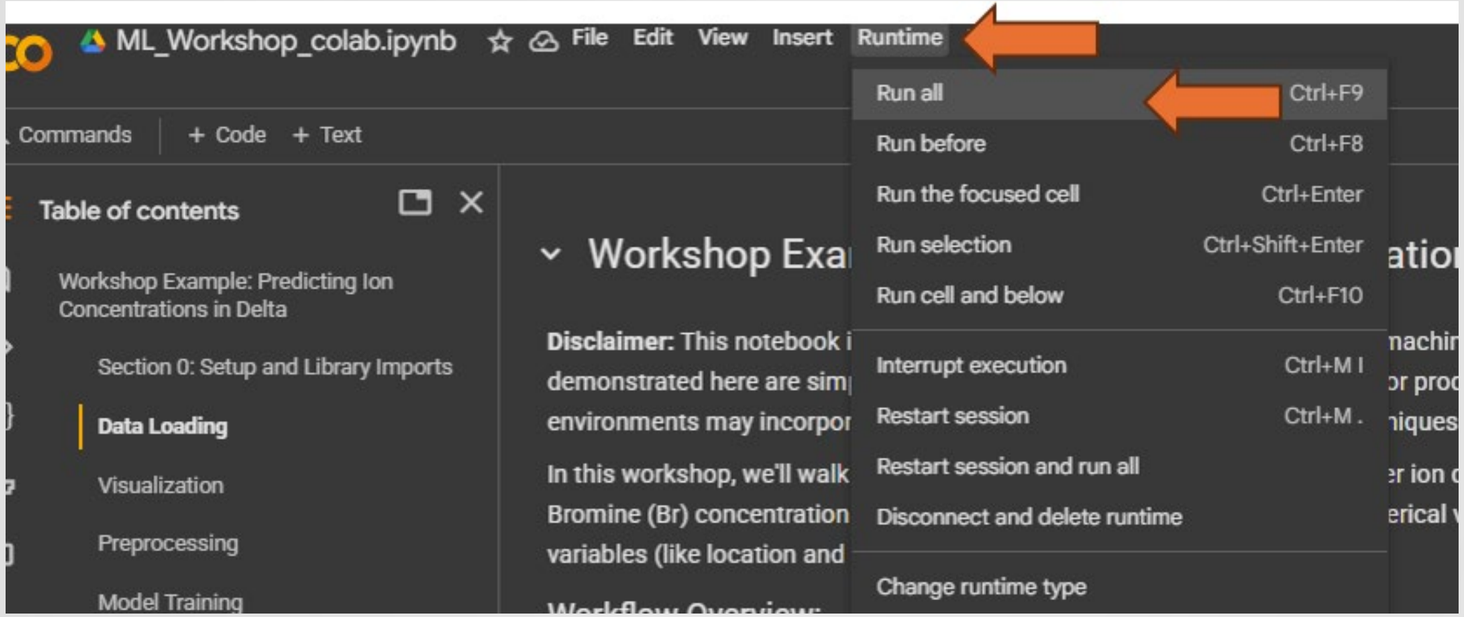
Step 5: Link the **Google Colab** Add-On to your Google Drive. See screenshot below. When complete, refresh your browser for the changes to take effect.



Step 6: Go into the "WorkshopML_Colab" folder you uploaded and open *ML_Workshop_colab.ipynb*. Right-click and Open With the Google Colaboratory.



Step 7: With the script now open in Colab, run the script by clicking *Runtime* and *Run All*. The script takes about 2 minutes to complete. Note: upon launching the code, Google Colab will request permission to access the folder – follow the prompts to allow access.



Questions?

CALIFORNIA DEPARTMENT OF
WATER RESOURCES

Contact: Kevin He Kevin.He@water.ca.gov