

Laptop Prep for “Hands-on: Data Wrangling for Machine Learning” with Python

Overview

Laptop preparation for the class consists of four high-level steps, with detailed instructions below:

1. Download course files from GitHub
2. Installation of Anaconda Python
3. Package downloads
4. Verify installation

NOTE – When using a work laptop, please keep the following in mind:

- Administrator permission may be required to complete laptop prep.
- It is often necessary to disable anti-virus software to allow for the installation. As such, disabling any anti-virus is recommended before laptop prep.
- Corporate proxy servers and firewalls can block the installation. Be sure to consult your IT department as needed.
- Lastly, installing the version of Anaconda Python listed below is highly recommended, even if you have Python already installed. The course has been tested only with the version listed below.

Because of the above, many attendees find using their personal laptop for the course is an easier option.

Hardware Requirements

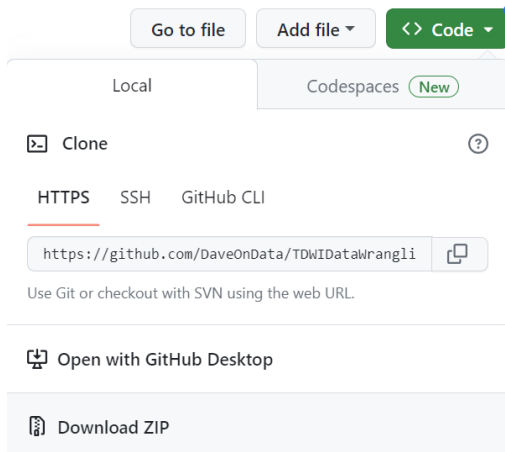
1. Windows or Mac OS X preferred (instructors have no experience with Linux)
2. 64-bit operating system
3. 8GB of RAM, 16GB preferred
4. 5GB of free drive space

Step 1 - Download the files from GitHub

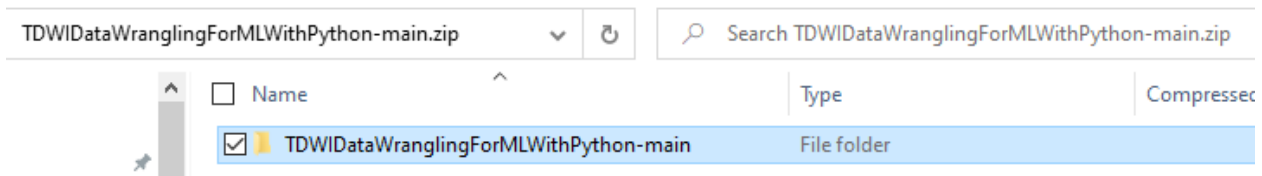
The GitHub repository with all required course files is located here:

1. The GitHub repository with all required course files is located here:
 - a. <https://github.com/DaveOnData/TDWIDataWranglingForMLWithPython>

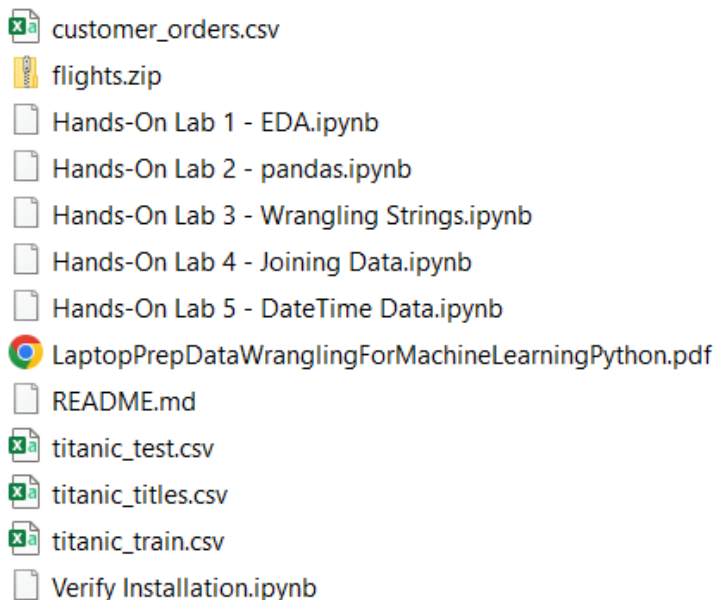
2. Within the GitHub repository page, click on the “Code” button and select “Download ZIP”:



3. Copy the file folder within the downloaded ZIP to a well-known location on your laptop (e.g., the Desktop):



4. Open the file folder. You should see the following files:



Step 2 – Anaconda Python Installation

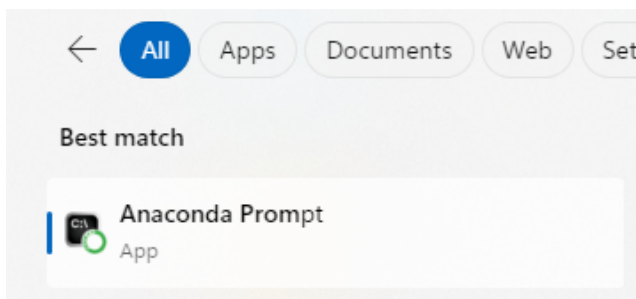
1. Click the link below for your operating system to download the Anaconda installer:
 - Windows: https://repo.anaconda.com/archive/Anaconda3-2024.10-1-Windows-x86_64.exe
 - Mac (Apple CPU): <https://repo.anaconda.com/archive/Anaconda3-2024.10-1-MacOSX-arm64.pkg>
 - Mac (Intel CPU): https://repo.anaconda.com/archive/Anaconda3-2024.10-1-MacOSX-x86_64.pkg

NOTE – The above versions are not the latest Anaconda distribution because of versioning conflicts with libraries used in the ML Bootcamp. The course has been tested with this version only. Also, if you installed the 2024.10 version of Anaconda as part of the Python Quick Start you do not need to install it again.

2. When the installer has downloaded, start the installer and follow the instructions (accepting defaults) to complete the installation.

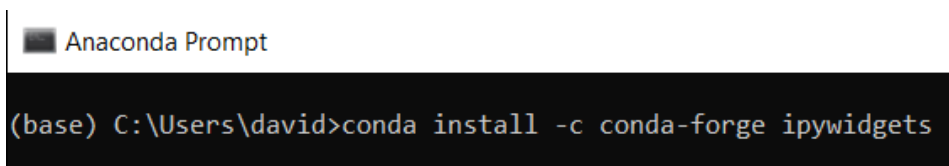
Step 3 – Package Downloads

1. NOTE – Some packages are used across multiple TDWI classes. If you are taking multiple classes as part of the same training (e.g., conference or bootcamp), you only need to install the packages once. For example, if you've installed plotnine as part of the Python Quick Start, you don't need to install it again.
2. With Anaconda Python installed, start the Anaconda Prompt:



NOTE – On Mac, use the Terminal app. The commands are the same.

1. At the command prompt type the following without quotes and hit <enter>:
 - a. "conda install -c conda-forge ipywidgets"



2. If prompted, hit the “y” key and <enter> to proceed:

```
Anaconda Prompt - conda install -c conda-forge ipywidgets

widgetsnbextension-4.0.10 | pyhd8ed1ab_0      866 KB  conda-forge
-----
Total: 16.3 MB

The following NEW packages will be INSTALLED:

python_abi      conda-forge/win-64::python_abi-3.11-2_cp311
ucrt             conda-forge/win-64::ucrt-10.0.22621.0-h57928b3_0
vc14_runtime    conda-forge/win-64::vc14_runtime-14.38.33130-h82b7239_18

The following packages will be UPDATED:

ca-certificates  pkgs/main::ca-certificates-2023.12.12~ --> conda-forge::ca-certificates-2024.2.2-h56e8100_0
comm             pkgs/main/win-64::comm-0.1.2-py311haa~ --> conda-forge/noarch::comm-0.2.2-pyhd8ed1ab_0
ipywidgets      pkgs/main::ipywidgets-7.6.5-pyhd3eb1b~ --> conda-forge::ipywidgets-8.1.2-pyhd8ed1ab_0
jupyterlab_widgets pkgs/main/win-64::jupyterlab_widgets~ --> conda-forge/noarch::jupyterlab_widgets-3.0.10-pyhd8ed1ab_0
openssl         pkgs/main::openssl-3.0.13-h2bfff1b_0 --> conda-forge::openssl-3.2.1-hcfcfb64_0
vs2015_runtime  pkgs/main::vs2015_runtime-14.27.29016~ --> conda-forge::vs2015_runtime-14.38.33130-hcb4865c_18
widgetsnbextension pkgs/main/win-64::widgetsnbextension~ --> conda-forge/noarch::widgetsnbextension-4.0.10-pyhd8ed1ab_0

The following packages will be SUPERSEDED by a higher-priority channel:

certifi         pkgs/main/win-64::certifi-2024.2.2-py~ --> conda-forge/noarch::certifi-2024.2.2-pyhd8ed1ab_0
conda           pkgs/main::conda-24.1.2-py311haa95532~ --> conda-forge::conda-24.1.2-py311h1ea47a8_0
pydeck         pkgs/main/win-64::pydeck-0.8.0-py311h~ --> conda-forge/noarch::pydeck-0.8.0-pyhd8ed1ab_0

Proceed ([y]/n)?
```

3. When the installation is completed, you should see something like the following:

```
Proceed ([y]/n)? y

Downloading and Extracting Packages:

Preparing transaction: done
Verifying transaction: done
Executing transaction: done

(base) C:\Users\david>
```

4. Repeat the above process at the command prompt, but now type the following without quotes and hit <enter>:
 - a. “conda install -c conda-forge ydata-profiling”

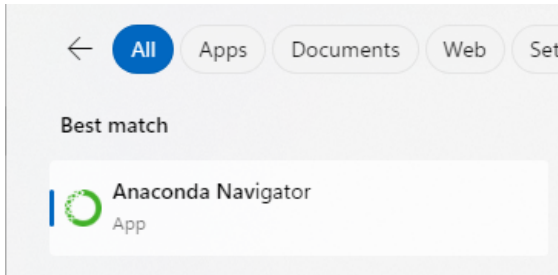
NOTE – This installation can take quite a while to run (e.g., 60 minutes). Please be patient.

5. Repeat the above process at the command prompt, but now type the following without quotes and hit <enter>:
 - a. “conda install -c conda-forge plotnine”

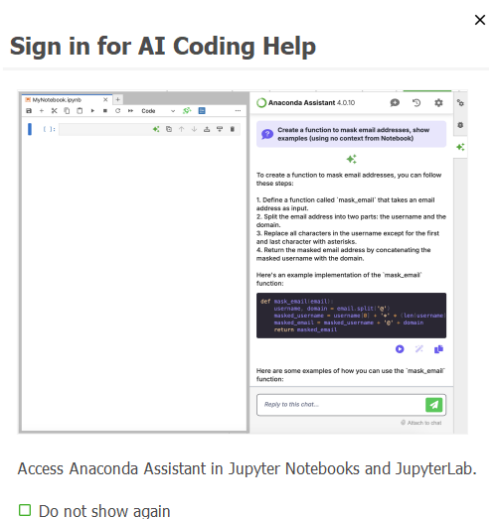
NOTE – This installation can take quite a while to run. Please be patient.

Step 4 – Verify Installation

1. With Anaconda Python installed, start the Anaconda Navigator application:



2. If you see the following dialog, close it (no need to sign in):



Access Anaconda Assistant in Jupyter Notebooks and JupyterLab.

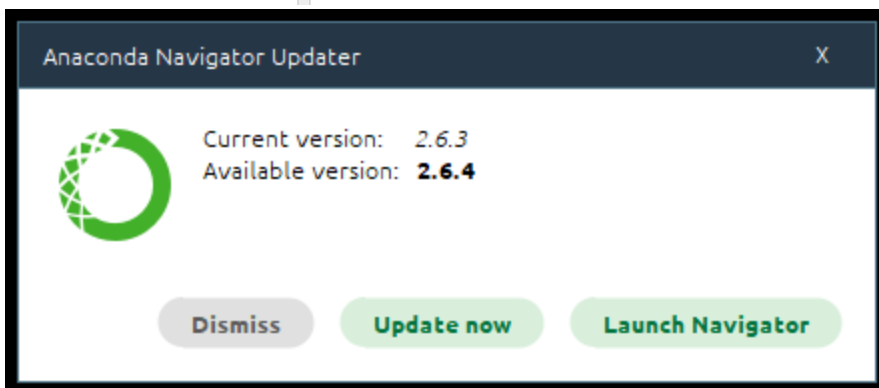
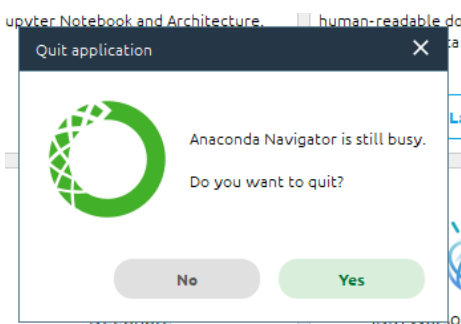
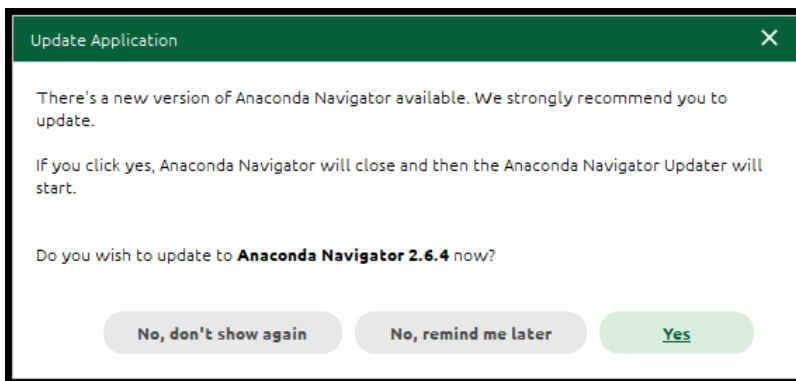
☐ Do not show again

[Connect to a Repository](#)

[Sign In](#)

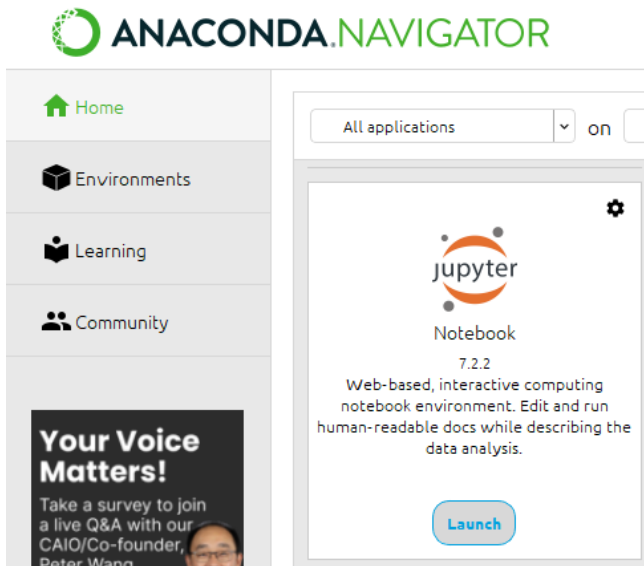
Don't have an account? [Sign Up](#)

3. You may be prompted to upgrade Anaconda Navigator. Follow the dialogs (i.e., click “Yes” and “Update now”) to do so. It may take a bit of time for this process to complete.



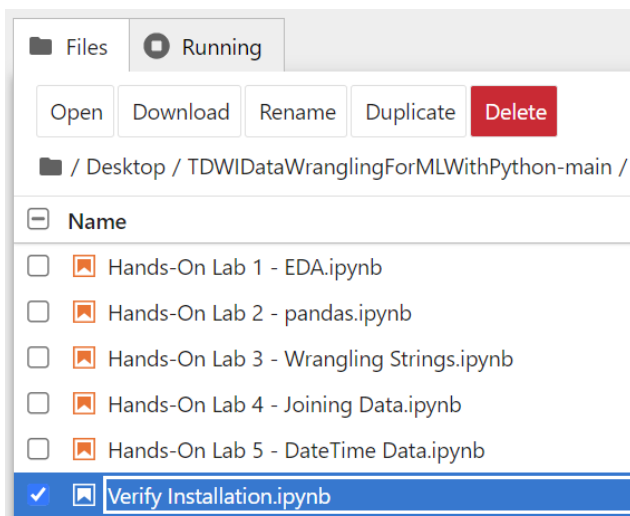
4. If needed, relaunch Anaconda Navigator.

5. NOTE – Your Anaconda Navigator window might not look exactly like the following. Within Anaconda Navigator, launch Jupyter Notebook:

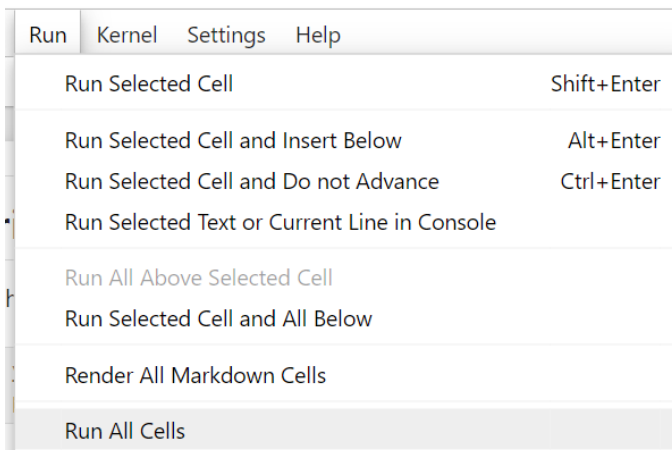


6. Within the Jupyter browser, navigate to where you copied the course file folder.

7. Double-click on the “Verify Installation.ipynb” entry:



8. Run all the cells in the notebook:



9. Your output should look like the following, with no errors. The following warnings are expected:

Verify Installation

Run the following code cell you should see no errors as a result of the running the code.

```
[1]: from ydata_profiling import ProfileReport
      from plotnine import ggplot
```

Congratulations! You are now ready for the class!