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 latest version of Anaconda Python is recommended – even if you have Python already installed.

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

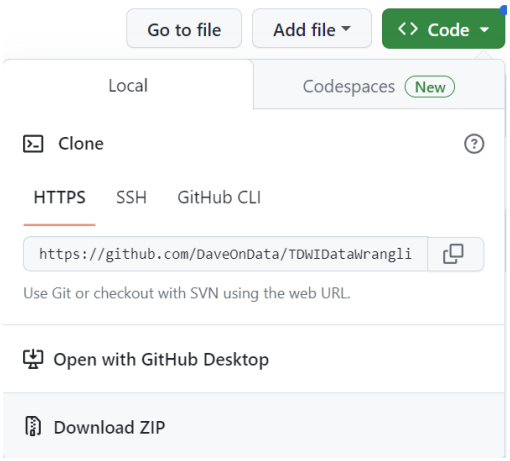
- https://github.com/DaveOnData/TDWIDataWranglingForMLWithPython

Hardware Requirements

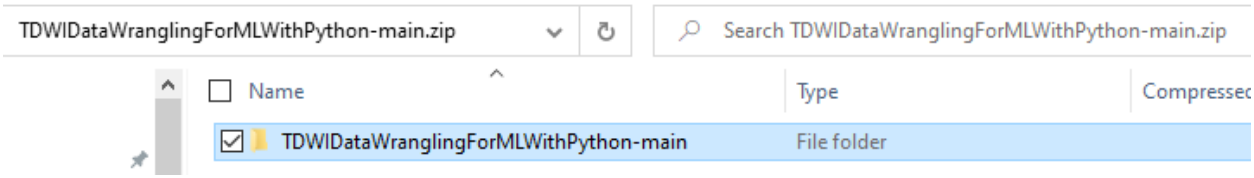
1. Windows or Mac OS X preferred (instructors have no experience with Linux)
2. 64-bit operating system
3. 8GB of RAM, 16GB preferred
4. 5GB of free drive space

Step 1 - Download the files from GitHub

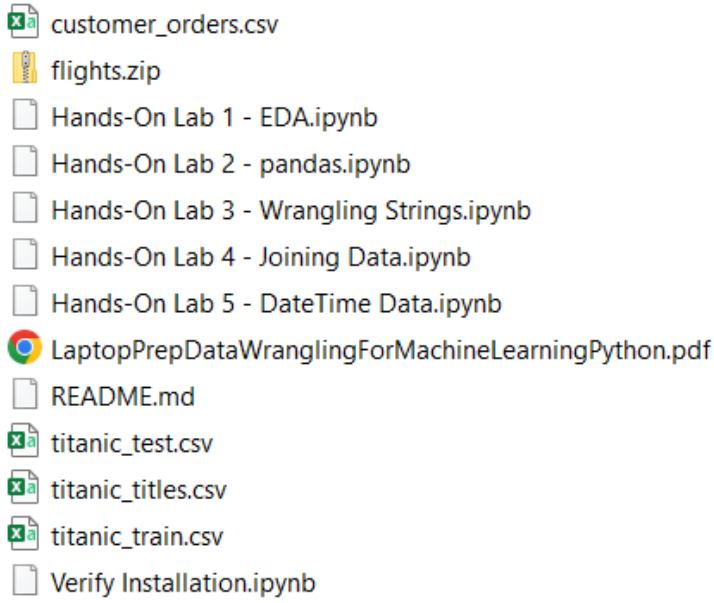
1. Within the GitHub repository page, click on the “Code” button and select “Download ZIP”:



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



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

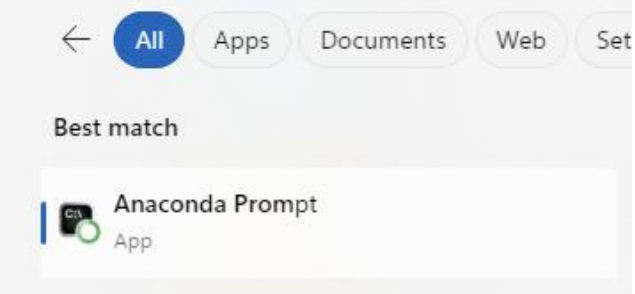


Step 2 – Anaconda Python Installation

1. Open your browser and navigate to: https://www.anaconda.com/download
2. Click “Skip registration.”
3. On the next page, select the download for your operating system (e.g., Windows).
4. When the installer has downloaded, start the installer and follow the instructions (accepting defaults) to complete the installation.

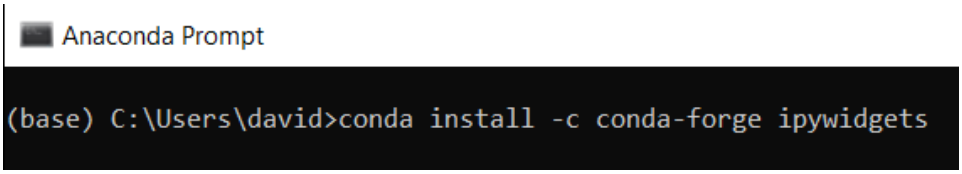
Step 3 – Package Downloads

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.
2. With Anaconda Python installed, start the Anaconda Prompt:

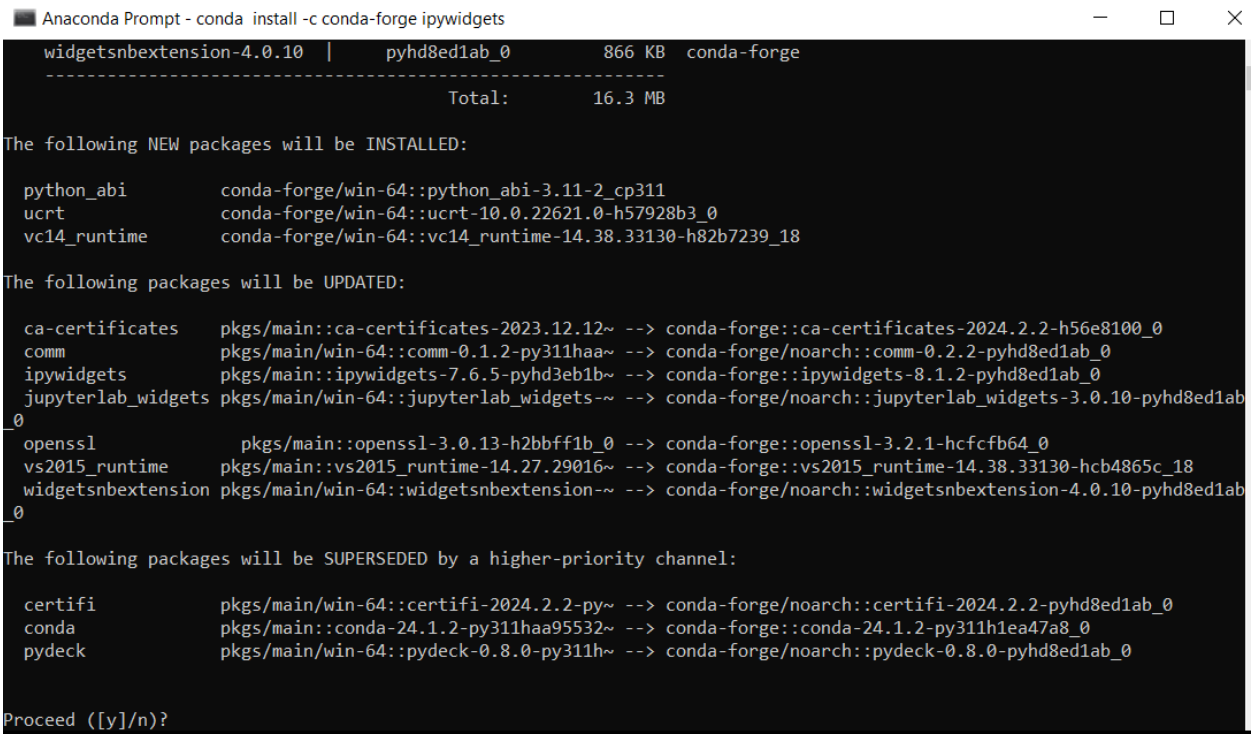


NOTE – On Mac, use the Terminal app. The commands are the same.

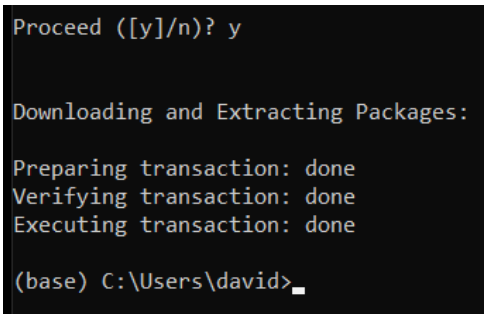
3. At the command prompt type the following without quotes and hit <enter>:
    a. “conda install -c conda-forge ipywidgets”



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



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



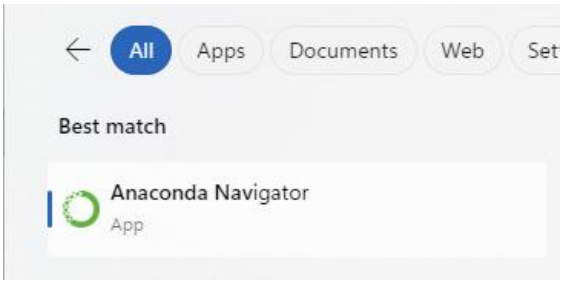
6. Repeat the above process at the command prompt, but now type the following without quotes and hit <enter>:
    a. “conda install -c conda-forge ydata-profiling”

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

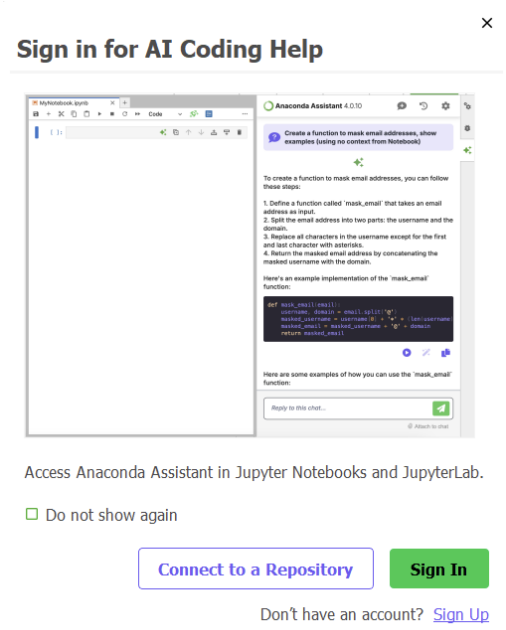
7. Repeat the above process at the command prompt, but now type the following without quotes and hit <enter>:
    a. “conda install -c conda-forge plotnine”

Step 4 – Verify Installation

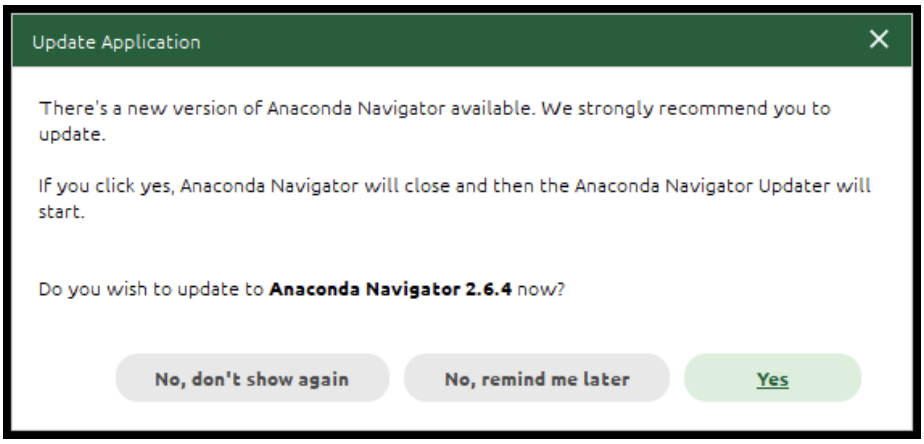
1. With Anaconda Python installed, start the Anaconda Navigator application:

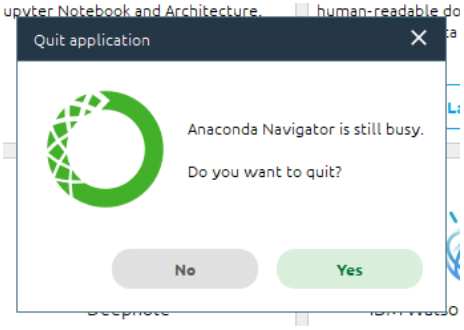


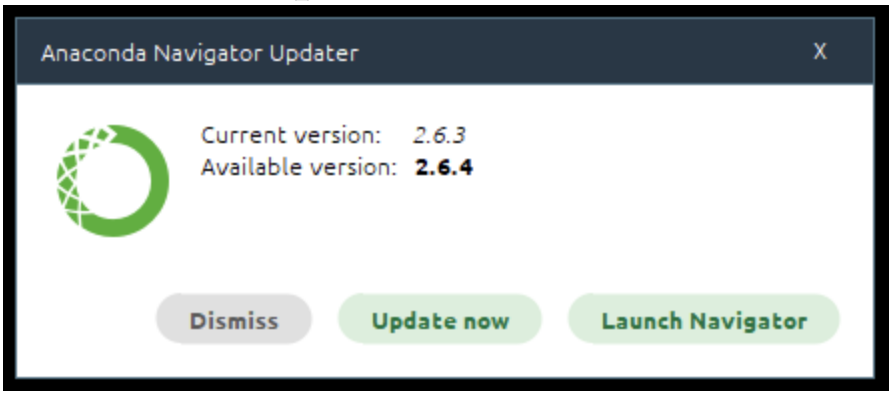
2. If you see the following dialog, close it (no need to sign in):



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.

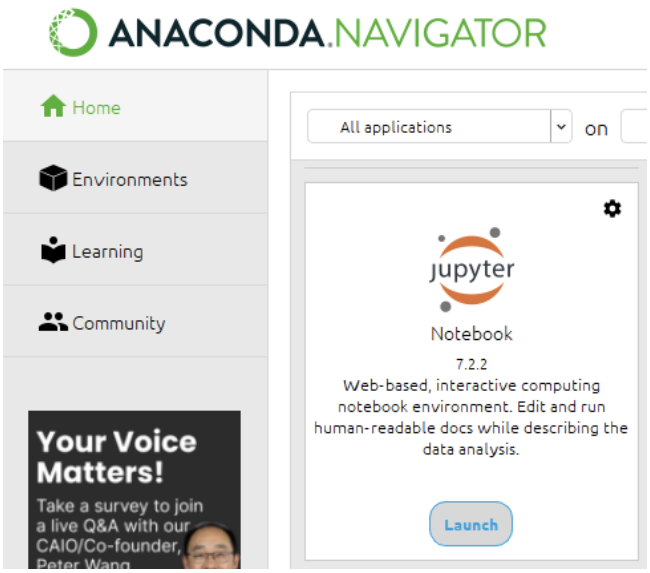






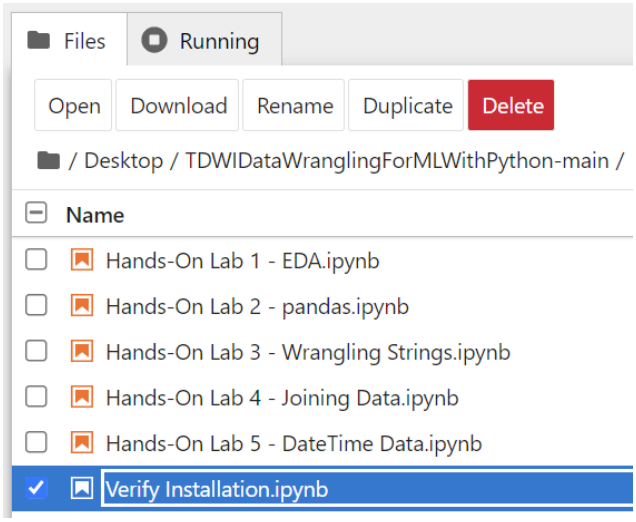
4. If needed, relaunch Anaconda Navigator.

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

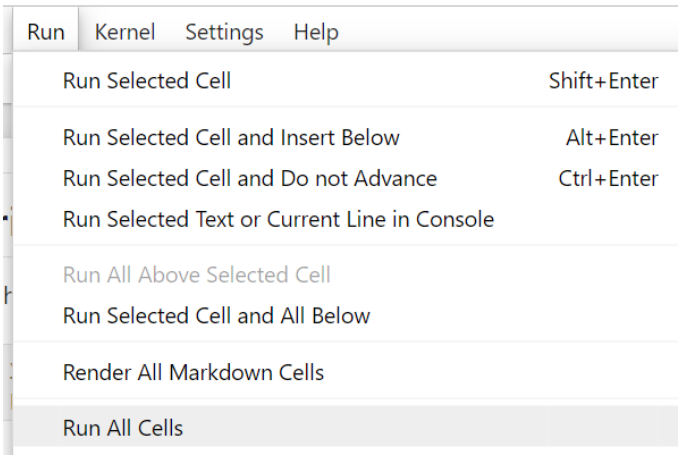


6. Within the Jupyter browser, navigate to where you copied the course file folder.

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!