

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

## Overview

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

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

NOTE – Administrator permission may be required to complete laptop prep. Also, often it is necessary to disable anti-virus software to allow for the installation. As such, disabling any anti-virus is recommended before laptop prep. 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:

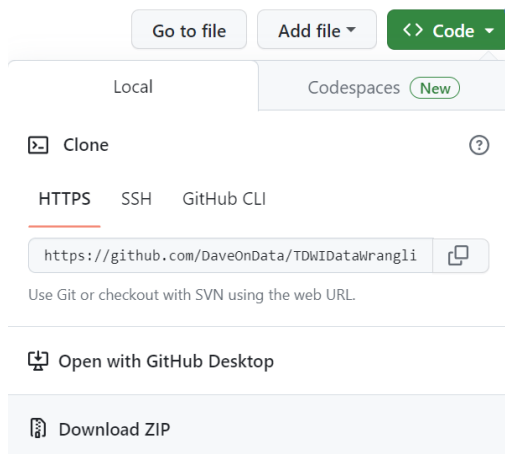
- <https://github.com/DaveOnData/TDWIDataWranglingForMLWithPython>

## Hardware Requirements

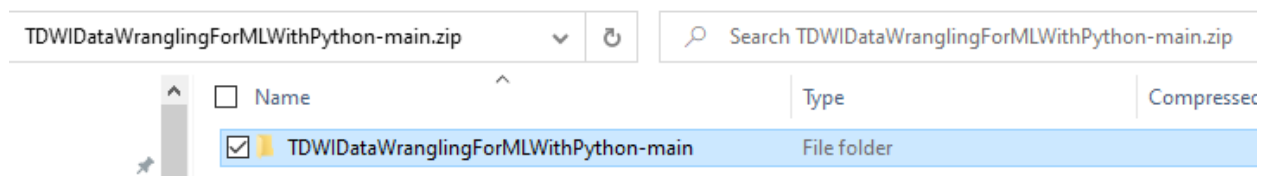
1. Windows or Mac OS X preferred (instructors have no experience with Linux)
2. 64-bit operating system
3. 8GB of RAM, 16GB preferred
4. 4GB of free drive space

## 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:

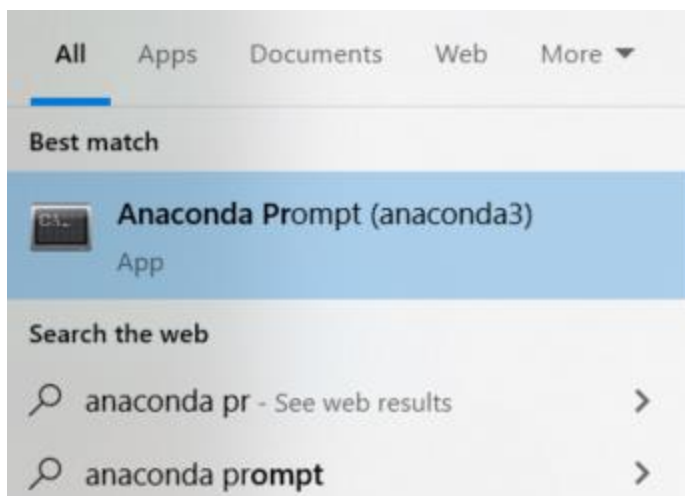
- flights.zip
- Hands-On Lab 1 - EDA.ipynb
- Hands-On Lab 2 - pandas.ipynb
- Hands-On Lab 3 - Wrangling Strings.ipynb
- Hands-On Lab 4 - Joining Data.ipynb
- LaptopPrepDataWranglingForMachineLearningPython.pdf
- README.md
- titanic\_test.csv
- titanic\_titles.csv
- titanic\_train.csv
- Verify Installation.ipynb

## Step 2 – Anaconda Python Installation

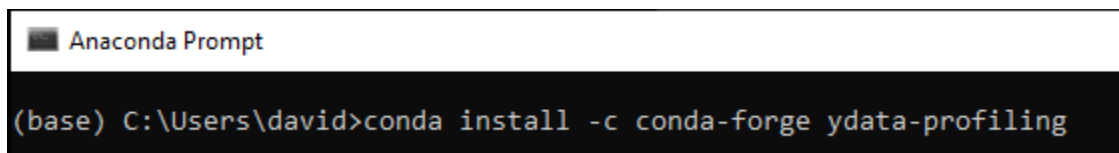
1. Open your browser and navigate to: <https://www.anaconda.com/products/distribution>
2. Click the download button.
3. When the installer has downloaded, start the installer and follow the instructions (accepting defaults) to complete the installation.

## 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:



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



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

```

Anaconda Prompt - conda install -c conda-forge ydata-profiling

## Package Plan ##

environment location: C:\Users\david\anaconda3

added / updated specs:
- ydata-profiling

The following packages will be downloaded:

package | build | size | channel
-----|-----|-----|-----
dacite-1.7.0 | pyhd8ed1ab_0 | 17 KB | conda-forge
dataclasses-0.8 | pyhc8e2a94_3 | 10 KB | conda-forge
htmlmin-0.1.12 | py_1 | 21 KB | conda-forge
imagehash-4.3.1 | pyhd8ed1ab_0 | 294 KB | conda-forge
multimethod-1.4 | py_0 | 10 KB | conda-forge
phik-0.12.3 | py310hcf51aa5_0 | 636 KB | conda-forge
pybind11-abi-4 | hd8ed1ab_3 | 10 KB | conda-forge
pydantic-1.10.8 | py310h8d17308_0 | 1.4 MB | conda-forge
tangled-up-in-unicode-0.2.0 | pyhd8ed1ab_0 | 2.9 MB | conda-forge
typeguard-2.13.3 | pyhd8ed1ab_0 | 19 KB | conda-forge
visions-0.7.5 | pyhd8ed1ab_0 | 59 KB | conda-forge
wordcloud-1.9.2 | py310h8d17308_0 | 173 KB | conda-forge
ydata-profiling-4.2.0 | pyhd8ed1ab_1 | 195 KB | conda-forge
-----|-----|-----|-----
Total: | | 5.7 MB |

The following NEW packages will be INSTALLED:

dacite conda-forge/noarch::dacite-1.7.0-pyhd8ed1ab_0
dataclasses conda-forge/noarch::dataclasses-0.8-pyhc8e2a94_3
htmlmin conda-forge/noarch::htmlmin-0.1.12-py_1
imagehash conda-forge/noarch::imagehash-4.3.1-pyhd8ed1ab_0
multimethod conda-forge/noarch::multimethod-1.4-py_0
phik conda-forge/win-64::phik-0.12.3-py310hcf51aa5_0
pybind11-abi conda-forge/noarch::pybind11-abi-4-hd8ed1ab_3
pydantic conda-forge/win-64::pydantic-1.10.8-py310h8d17308_0
tangled-up-in-uni~ conda-forge/noarch::tangled-up-in-unicode-0.2.0-pyhd8ed1ab_0
typeguard conda-forge/noarch::typeguard-2.13.3-pyhd8ed1ab_0
visions conda-forge/noarch::visions-0.7.5-pyhd8ed1ab_0
wordcloud conda-forge/win-64::wordcloud-1.9.2-py310h8d17308_0
ydata-profiling conda-forge/noarch::ydata-profiling-4.2.0-pyhd8ed1ab_1

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

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

```

Proceed ([y]/n)? y

Downloading and Extracting Packages

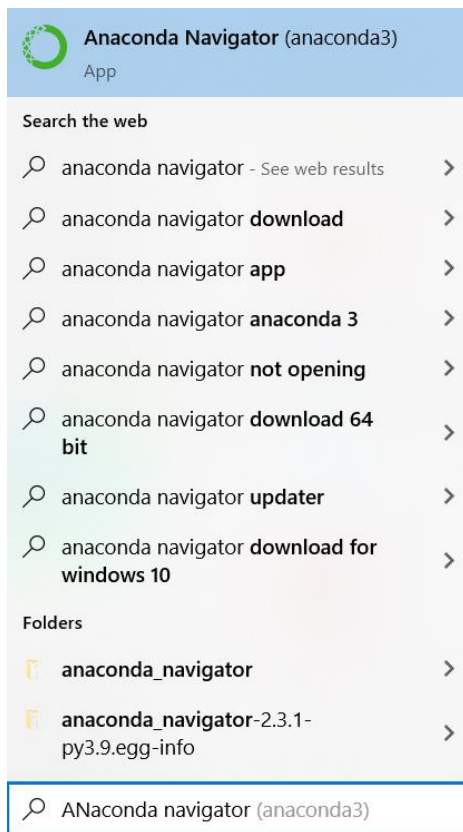
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

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

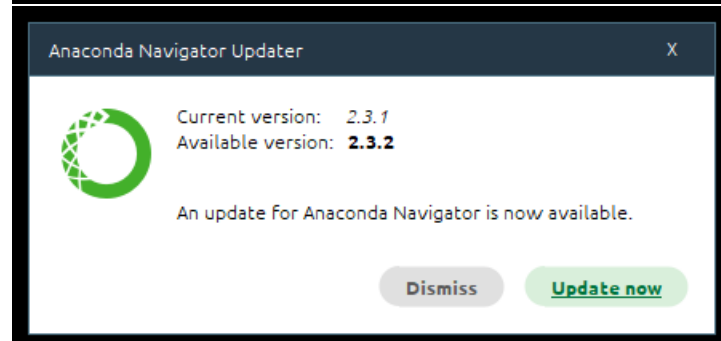
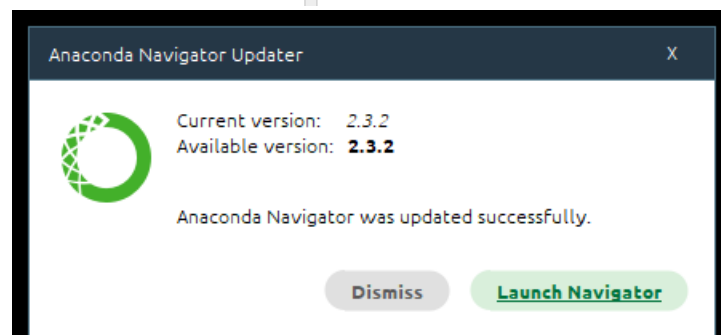
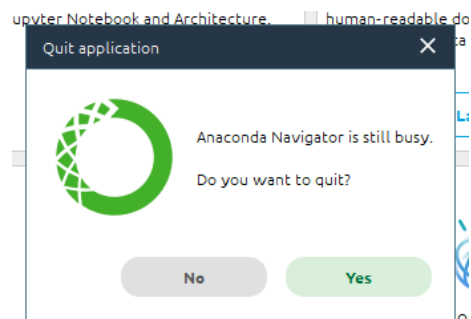
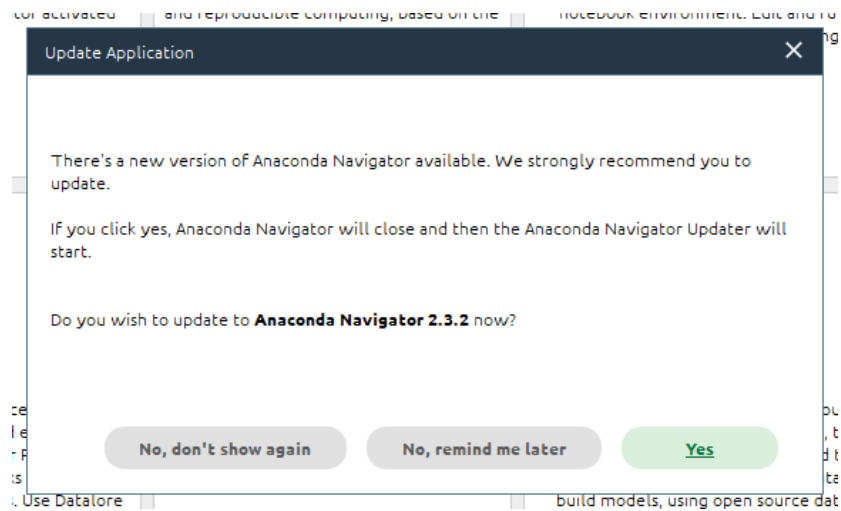
6. Repeat the above process at the command prompt, but now type the following without quotes and hit <enter>:
  - a. "conda install -c conda-forge plotnine"

## Step 4 – Verify Installation

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

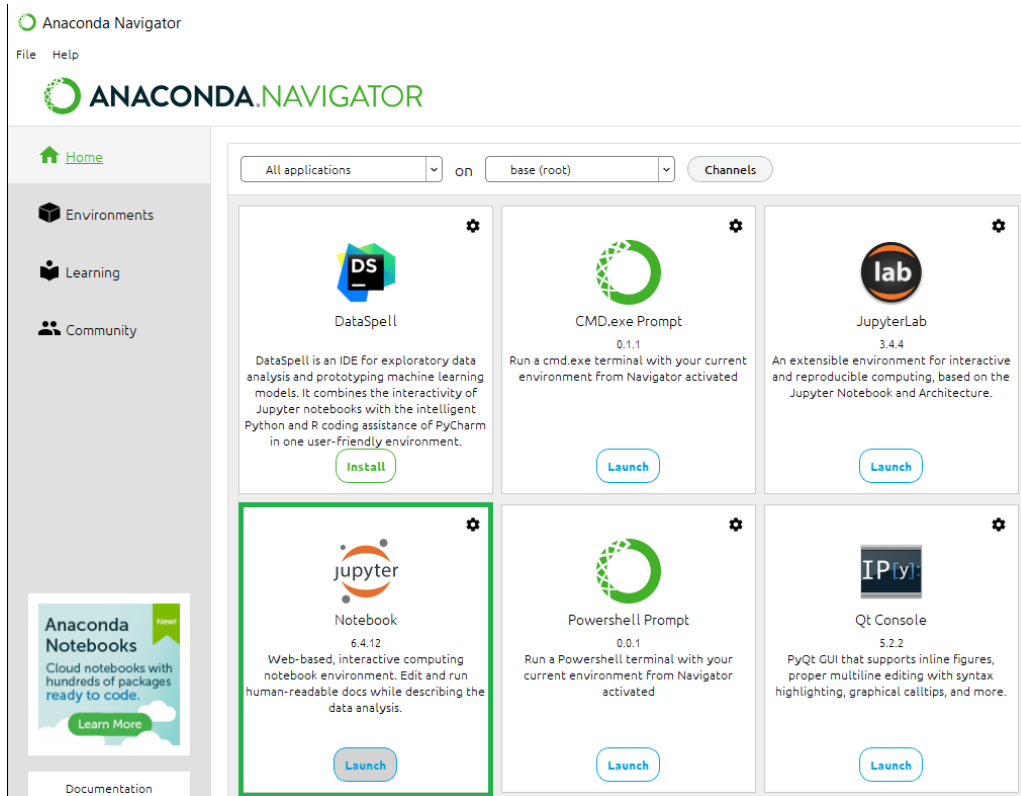


2. You may be prompted to upgrade Anaconda Navigator. Follow the dialogs to do so:

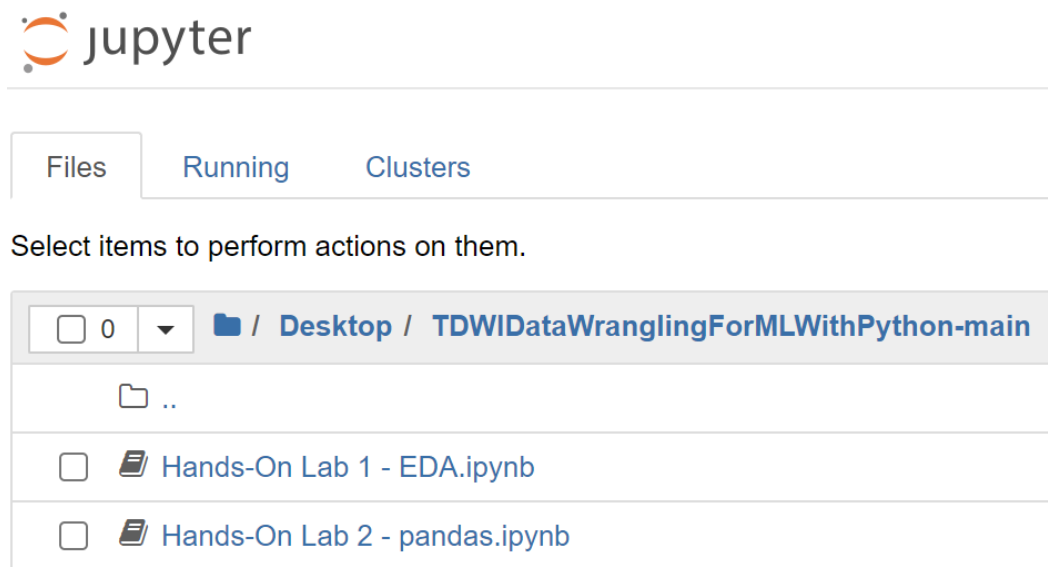


3. If needed, relaunch Anaconda Navigator

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



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

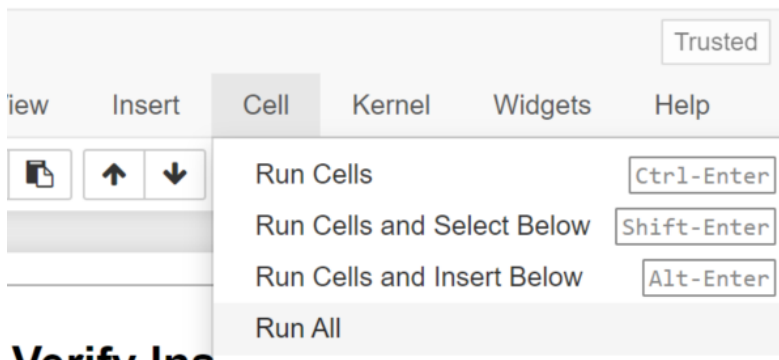


6. Click on the “Verify Installation.ipynb” entry:



7. Run all the cells in the notebook:

## Verify Installation



8. 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.

```
from ydata_profiling import ProfileReport
from plotnine import ggplot
```

C:\Users\david\anaconda3\Lib\site-packages\numba\core\decorators.py:262: NumbaDeprecationWarning: numba.generated\_jit is deprecated. Please see the documentation at: <https://numba.readthedocs.io/en/stable/reference/deprecation.html#deprecation-of-generated-jit> for more information and advice on a suitable replacement.  
warnings.warn(msg, NumbaDeprecationWarning)

C:\Users\david\anaconda3\Lib\site-packages\visions\backends\shared\nan\_handling.py:50: NumbaDeprecationWarning: The 'nopython' keyword argument was not supplied to the 'numba.jit' decorator. The implicit default value for this argument is currently False, but it will be changed to True in Numba 0.59.0. See <https://numba.readthedocs.io/en/stable/reference/deprecation.html#deprecation-of-object-mode-fall-back-behaviour-when-using-jit> for details.  
@nb.jit



9. Close Anaconda Navigator and quit Jupyter Notebook when prompted.

**Congratulations! You are now ready for the class!**