Laptop Prep for "Hands-on: Text Analytics with Python"

Introduction

Thank you for registering for Day 1 of my 3-day Machine Learning Bootcamp!

There are a few things that I want to mention to help you prepare for the course:

- If you're new to Python, TDWI will send you a link to my free Python Quick Start online tutorial. This will cover everything you need to know before class.
- If you're attending other courses in the ML Bootcamp, you will receive a laptop prep document like this one for each course.
- If you're attending all three days of the ML Bootcamp, you will receive my Introduction to Machine Learning course as a gift. TDWI will email you a link after the Bootcamp.
- If you would like to enroll in all three days of the Bootcamp, but are not, drop an email to info@tdwi.org.

I look forward to seeing you in class.
-Dave

Overview

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

1. Download course files from GitHub
2. Installation of Anaconda Python
3. NLTK downloads

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.

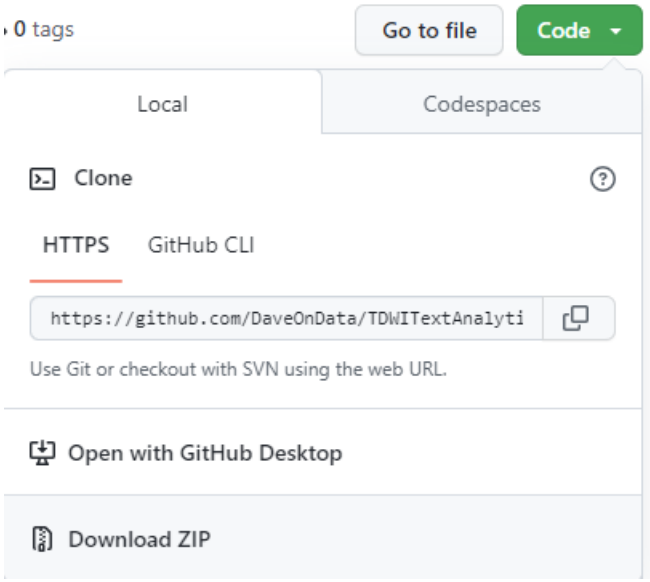
Hardware Requirements

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

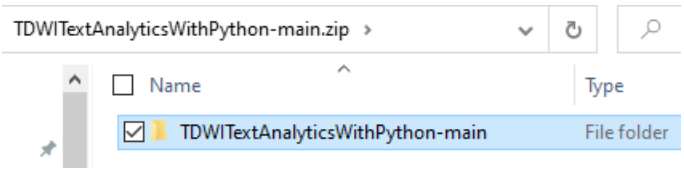
Step 1 - Download the files from GitHub

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

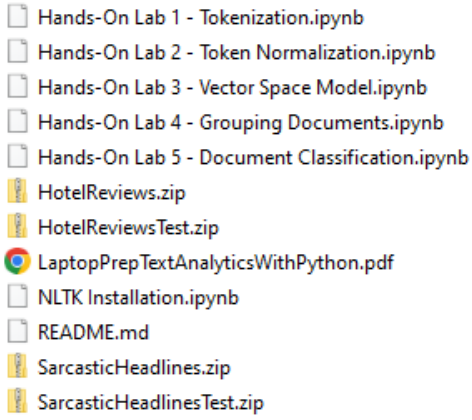
2. Within the GitHub repository page, click on the "Code" button and select "Download ZIP":



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

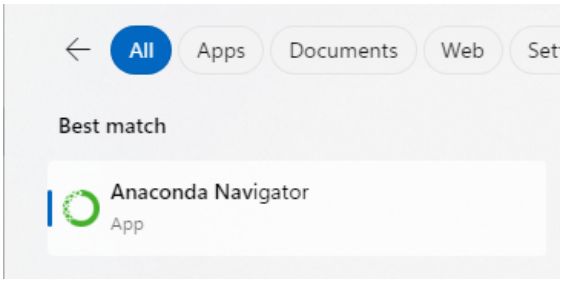
1. Click the link below for your operating system to download the Anaconda installer:
    - Windows: https://repo.anaconda.com/archive/Anaconda3-2025.06-1-Windows-x86_64.exe
    - Mac (Apple CPU): https://repo.anaconda.com/archive/Anaconda3-2025.06-1-MacOSX-arm64.pkg
    - Mac (Intel CPU): https://repo.anaconda.com/archive/Anaconda3-2025.06-1-MacOSX-x86_64.pkg

    NOTE – The above versions might not be the latest Anaconda distribution at the time of the ML Bootcamp. The course has been tested with this version only. Also, if you installed the 2025.06 version of Anaconda as part of the Python Quick Start, you do not need to install it again.

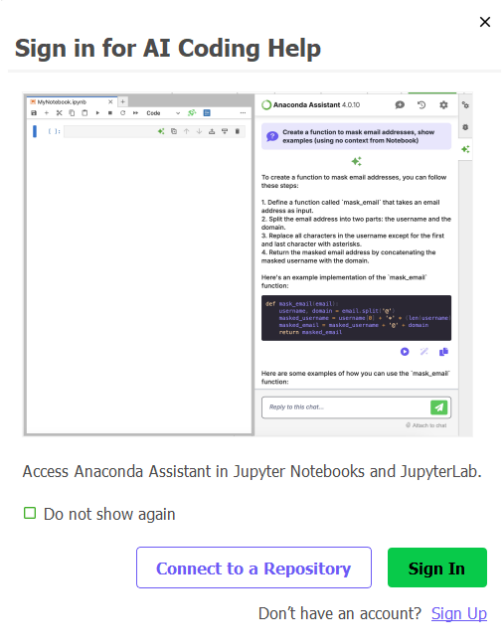
2. When the installer has downloaded, start the installer and follow the instructions (accepting defaults) to complete the installation.

Step 3 – NLTK Downloads

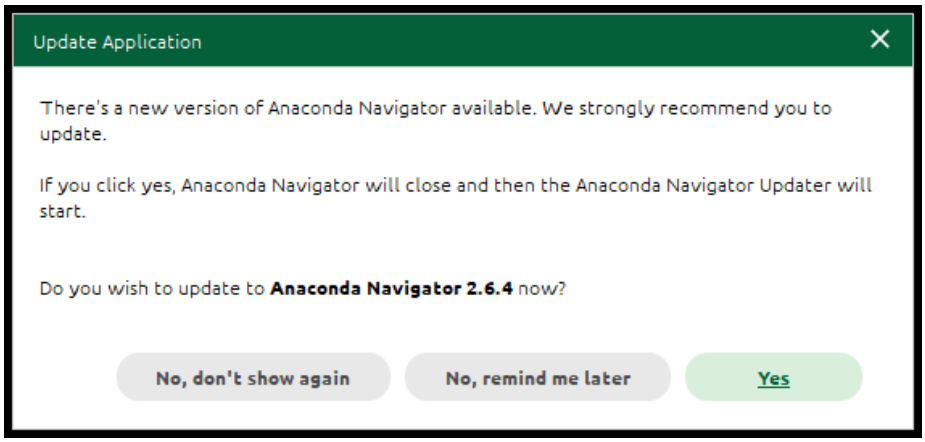
1. With Anaconda Python installed, start the Anaconda Navigator application:

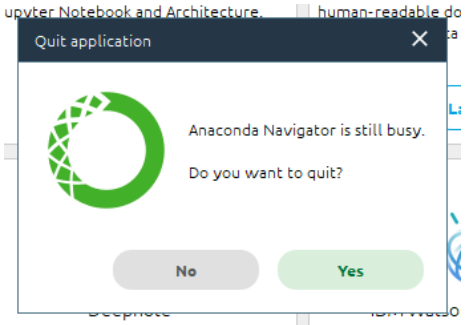


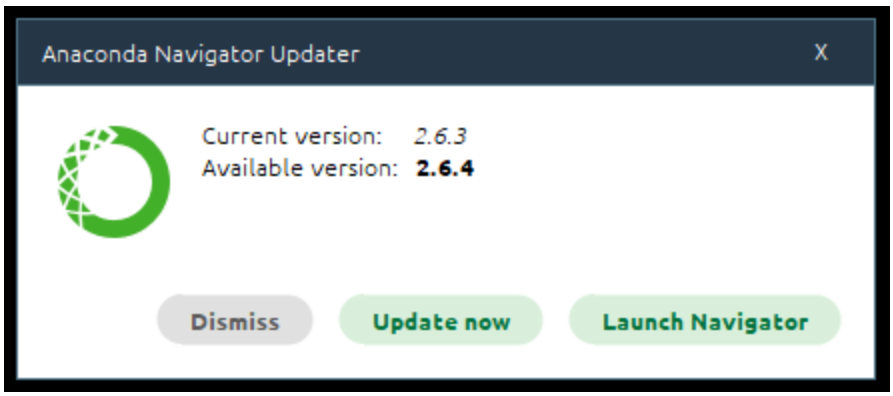
2. If you see the following dialog, close it (no need to sign in):



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.

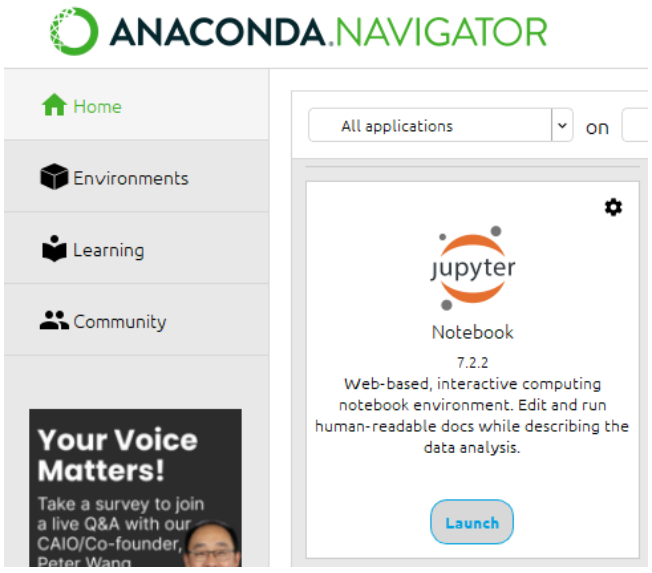




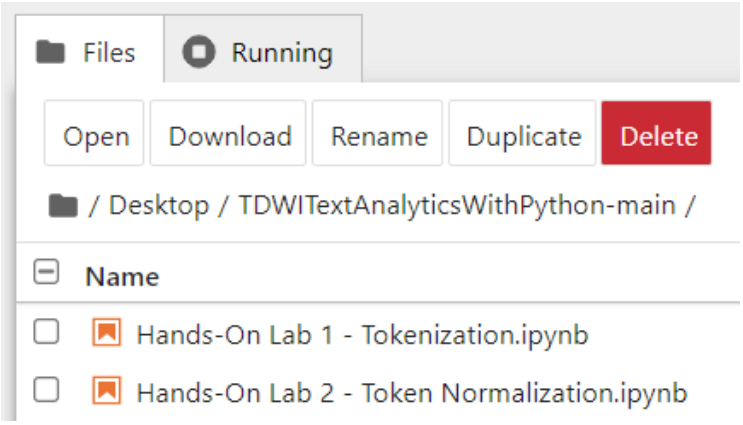


4. If needed, relaunch Anaconda Navigator

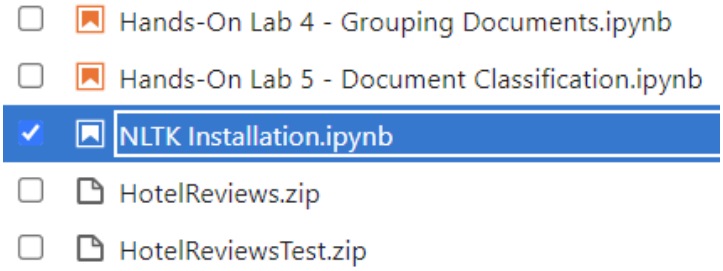
5. NOTE – Your Anaconda Navigator window might not look exactly like the following. Within Anaconda Navigator, launch Jupyter Notebook:



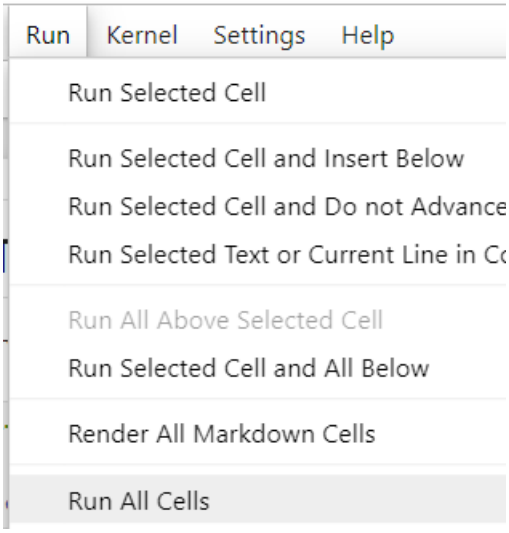
6. Within the Jupyter browser, navigate to where you copied the course file folder:



7. Double-click on the "NLTK Installation.ipynb" entry:



8. Run all the cells in the notebook:



9. All the cells should have output like the following:

```
In [1]: import nltk

        nltk.download('punkt')

        [nltk_data] Downloading package punkt to
        [nltk_data]     C:\Users\david\AppData\Roaming\nltk_data...
        [nltk_data]   Unzipping tokenizers\punkt.zip.

Out[1]: True
```

Congratulations! You are now ready for the class!