

Laptop Prep for “Hands-on: Introduction to Text Analytics 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:

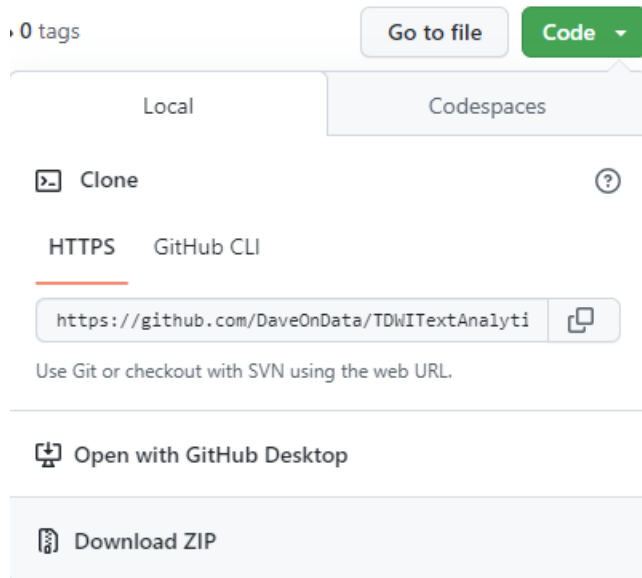
- <https://github.com/DaveOnData/TDWClusterAnalysisWithPython>

Hardware Requirements

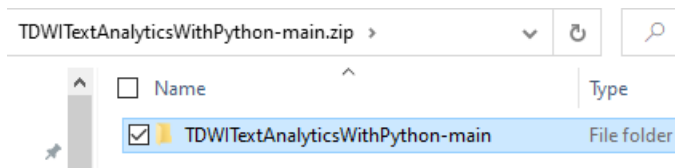
1. Windows or Mac OS X preferred (instructors have no experience with Linux)
2. 64-bit operating system
3. 8GB of RAM, 16GB preferred
4. 4GB of free drive space

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:

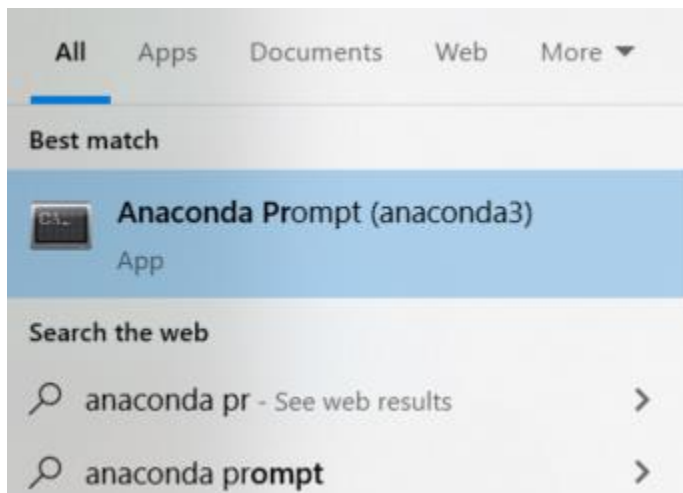
- 📄 Hands-On Lab 1 - Tokenization.ipynb
- 📄 Hands-On Lab 2 - Token Normalization.ipynb
- 📄 Hands-On Lab 3 - Vector Space Model.ipynb
- 📄 Hands-On Lab 4 - Grouping Documents.ipynb
- 📄 Hands-On Lab 5 - Document Classification.ipynb
- 📄 HotelReviews.zip
- 📄 HotelReviewsTest.zip
- 📄 LaptopPrepTextAnalyticsWithPython.pdf
- 📄 NLTK Installation.ipynb
- 📄 README.md
- 📄 SarcasticHeadlines.zip
- 📄 SarcasticHeadlinesTest.zip

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. With Anaconda Python installed, start the Anaconda Prompt:



2. At the command prompt type the following without quotes and hit <enter>:
 - a. “conda install -c conda-forge prince-factor-analysis”

Anaconda Prompt (anaconda3)

```
(base) C:\Users\david>conda install -c conda-forge prince-factor-analysis
```

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

```
Anaconda Prompt (anaconda3) - conda install -c conda-forge prince-factor-analysis

(base) C:\Users\david>conda install -c conda-forge prince-factor-analysis
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: C:\Users\david\anaconda3

  added / updated specs:
    - prince-factor-analysis

The following packages will be downloaded:

  package | build
  -----|-----
  prince-factor-analysis-0.7.1 | pyhd8ed1ab_1      21 KB  conda-forge
  -----|-----
                             Total:      21 KB

The following NEW packages will be INSTALLED:

  prince-factor-ana~ conda-forge/noarch::prince-factor-analysis-0.7.1-pyhd8ed1ab_1

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

4. You should see something like the following:

```
Anaconda Prompt (anaconda3)

environment location: C:\Users\david\anaconda3

added / updated specs:
- prince-factor-analysis

The following packages will be downloaded:

package | build
-----|-----
prince-factor-analysis-0.7.1 | pyhd8ed1ab_1 21 KB conda-forge
-----|-----
Total: 21 KB

The following NEW packages will be INSTALLED:

prince-factor-ana~ conda-forge/noarch::prince-factor-analysis-0.7.1-pyhd8ed1ab_1

Proceed ([y]/n)? y

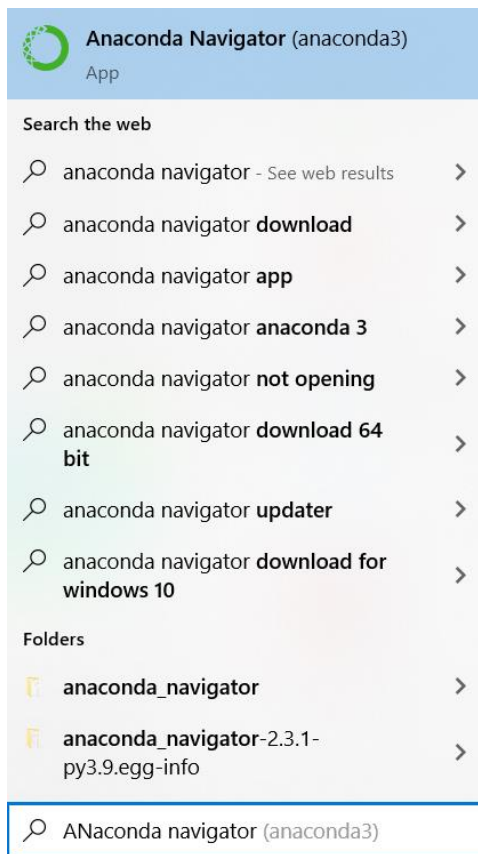
Downloading and Extracting Packages
Preparing transaction: done
Verifying transaction: done
Executing transaction: done

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

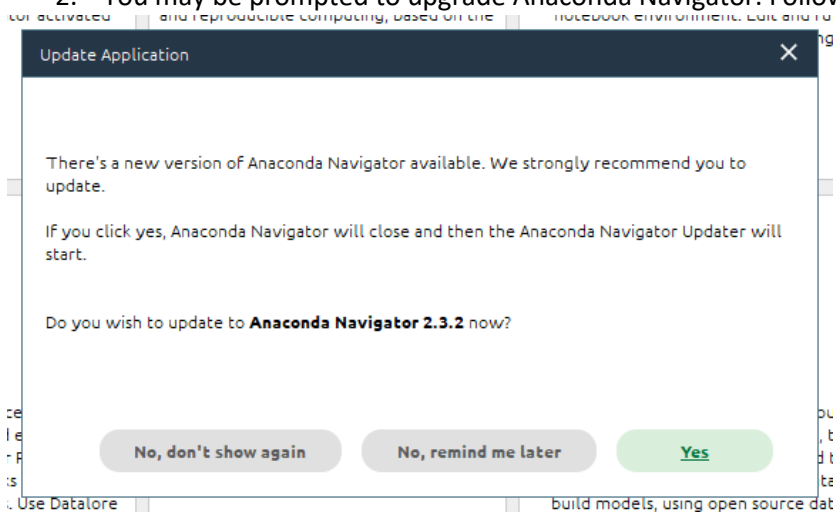
5. Repeat the above process by typing the following on the command line (without quotes) and hitting <enter>:
 - a. "conda install -c conda-forge plotnine"

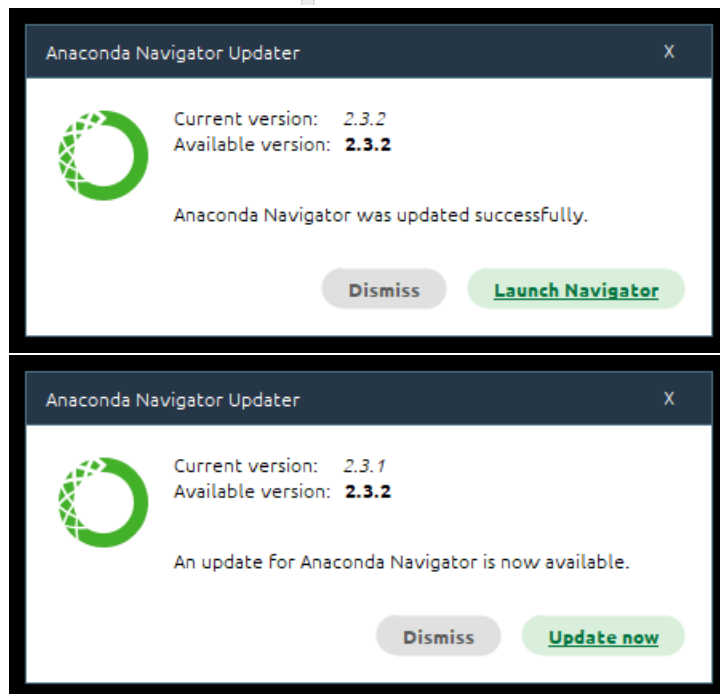
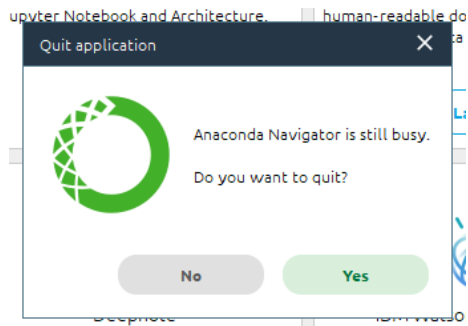
Step 4 – Verify Installation

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

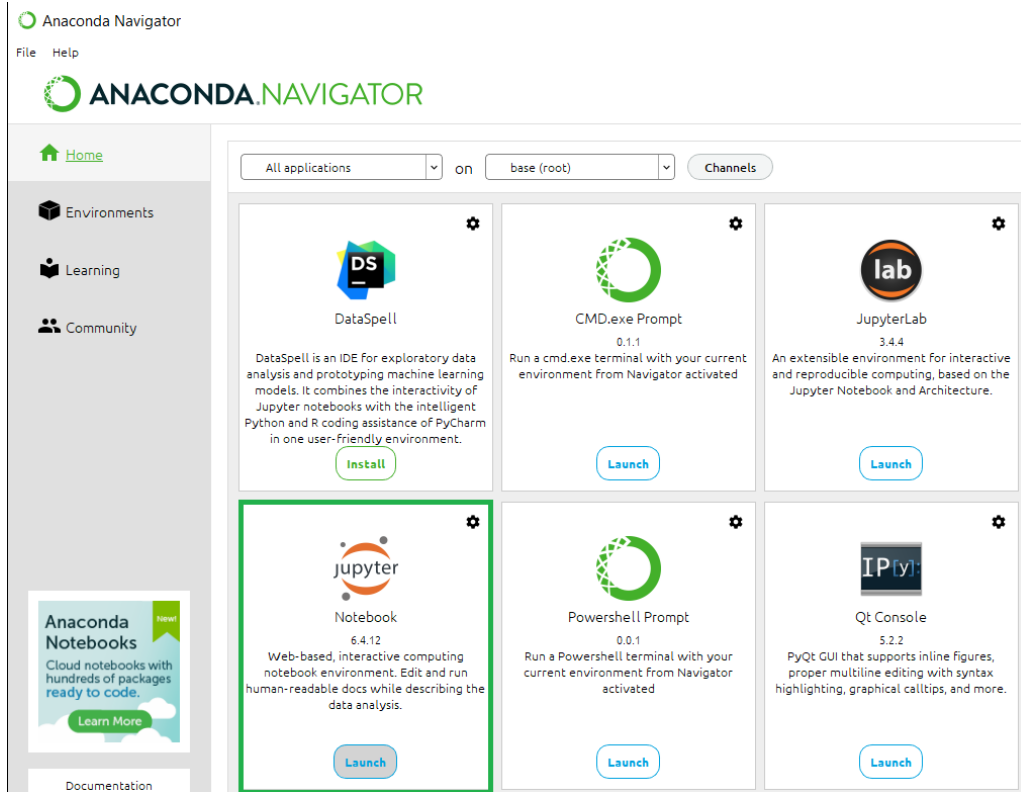


2. You may be prompted to upgrade Anaconda Navigator. Follow the dialogs to do so:

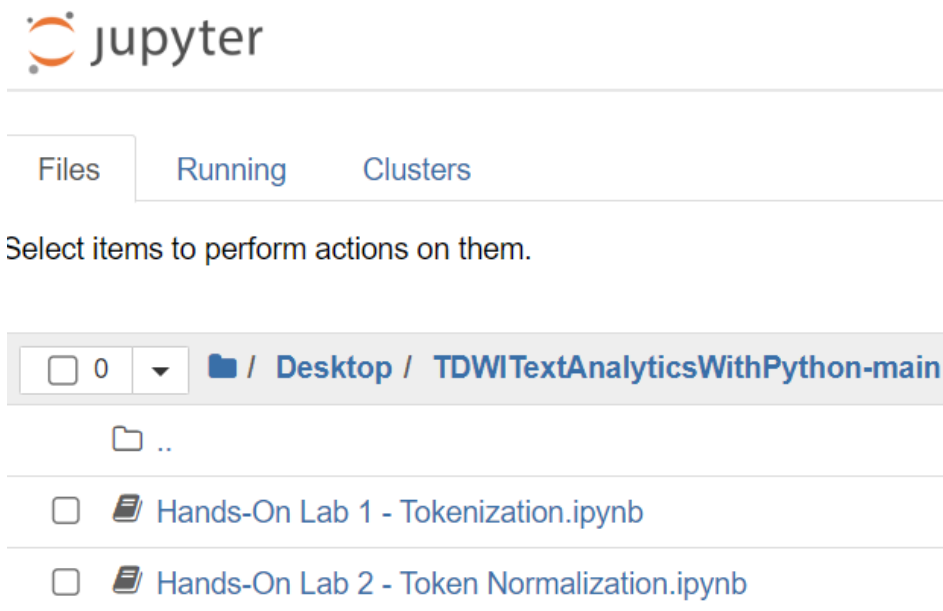




3. If needed, relaunch Anaconda Navigator
4. NOTE – Your Anaconda Navigator window might not look exactly like the following. Within Anaconda Navigator, launch Jupyter Notebook:



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



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

☐  Hands-On Lab 5 - Document Classification.ipynb

☐  NLTK Installation.ipynb

☐  HotelReviews.zip

7. Run all the cells in the notebook:

NLTK Installation (autosaved)



8. All of the cells should have output similar to 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!

1.