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

<u>NOTE</u> – 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:
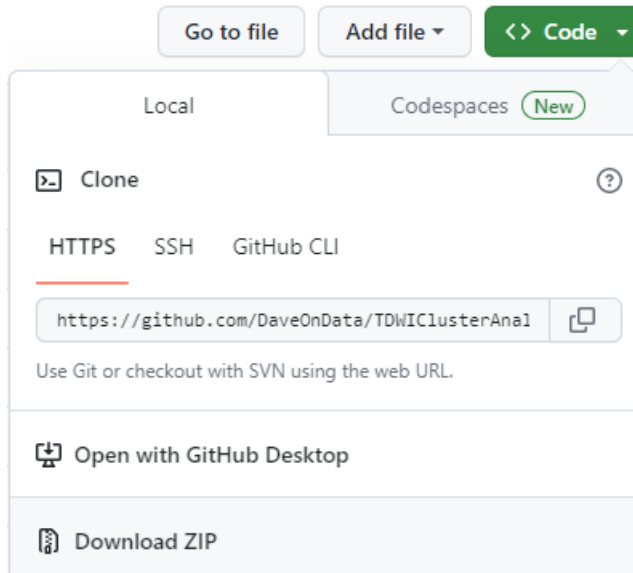
- https://github.com/DaveOnData/TDWIClusterAnalysisWithPython

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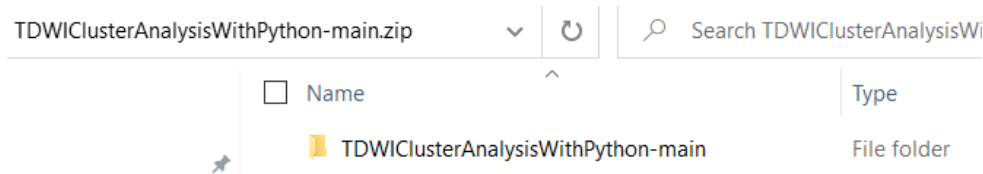
3. Open the file folder. You should see the following files:

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:
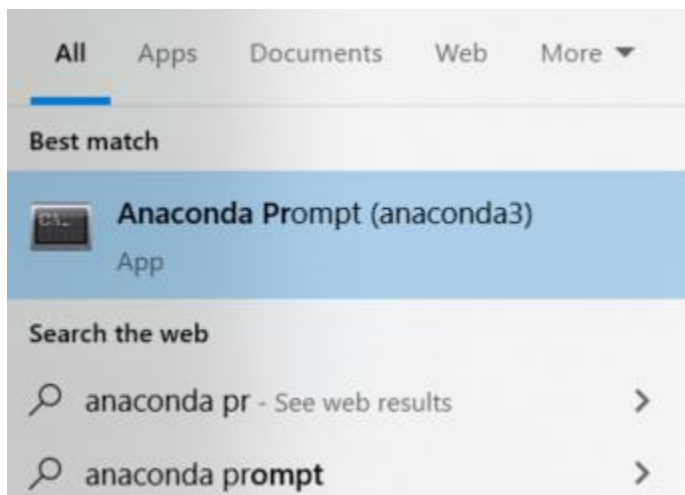
2. At the command prompt type the following without quotes and hit <enter>:
   a. "conda install -c conda-forge prince-factor-analysis"
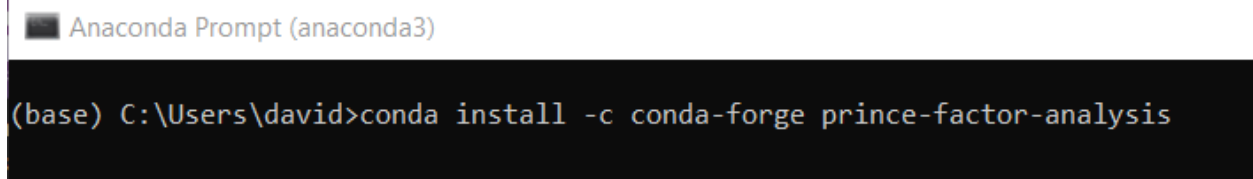
3. If prompted, hit the "y" key and <enter> to proceed:

```
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>
```
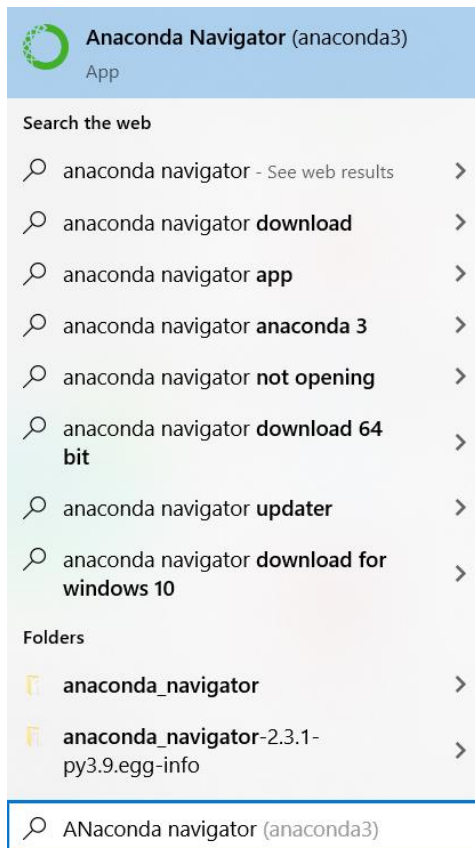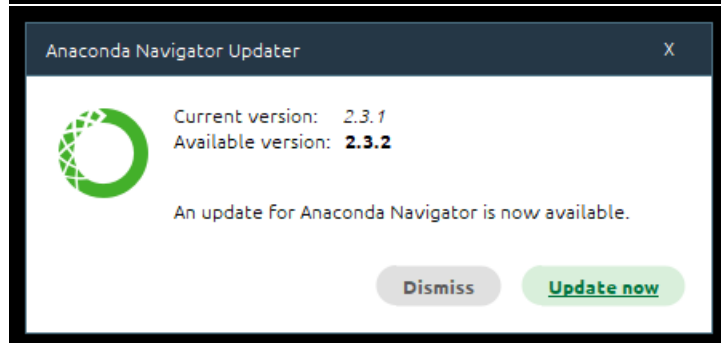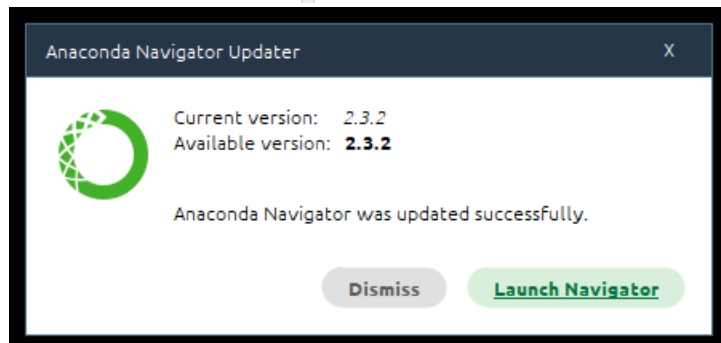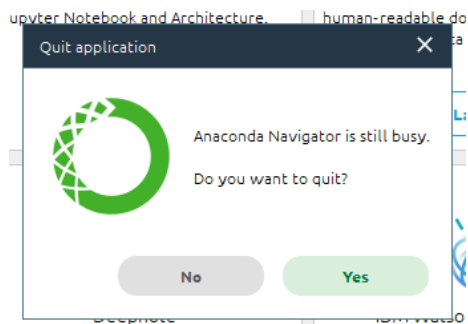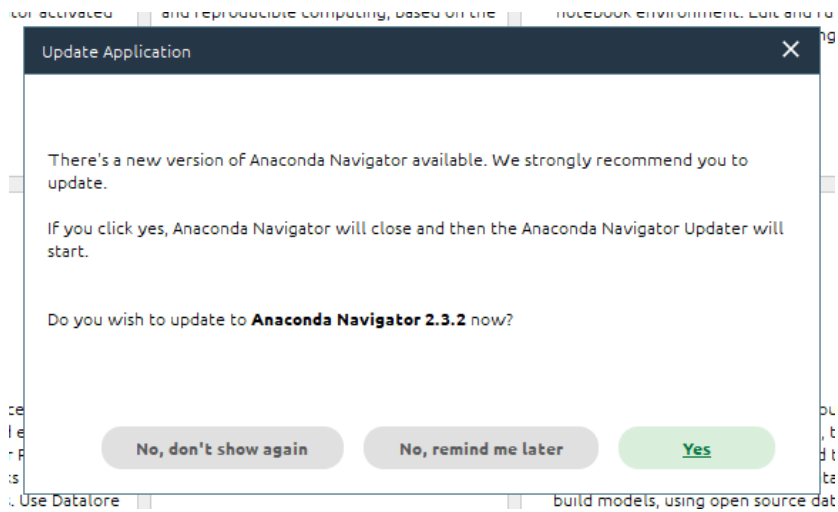
5. Repeat the above process by typing the following on the command line (without quotes) and hitting <enter>:
   a. "conda install -c conda-forge plotnine"
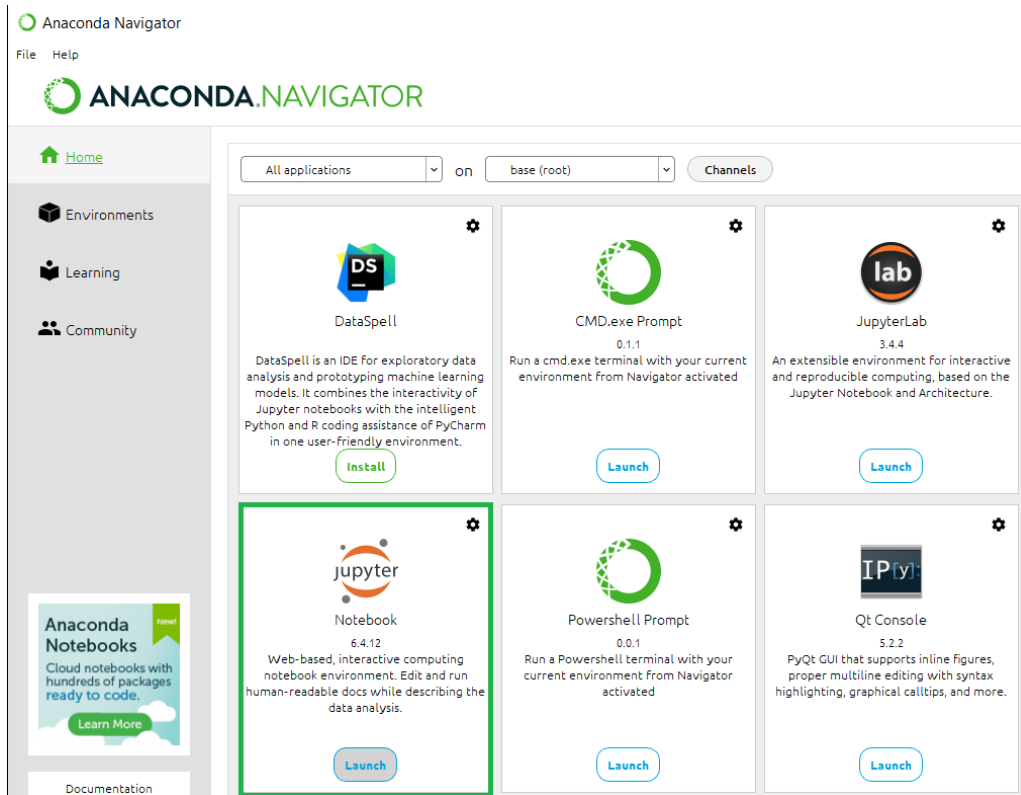
Step 4 – Verify Installation

1. With Anaconda Python installed, start the Anaconda Navigator application:
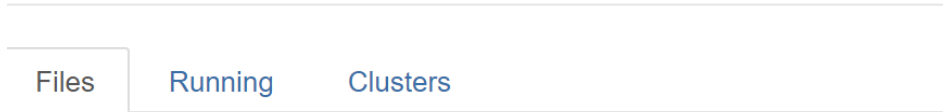
2. You may be prompted to upgrade Anaconda Navigator. Follow the dialogs to do so:

Update Application   ✕

There's a new version of Anaconda Navigator available. We strongly recommend you to update.

If you click yes, Anaconda Navigator will close and then the Anaconda Navigator Updater will start.

Do you wish to update to **Anaconda Navigator 2.3.2** now?

No, don't show again	No, remind me later	Yes

Quit application   ✕

Anaconda Navigator is still busy.

Do you want to quit?

No	Yes

Anaconda Navigator Updater   X

Current version: 2.3.2
Available version: **2.3.2**

Anaconda Navigator was updated successfully.

Dismiss	Launch Navigator

Anaconda Navigator Updater   X

Current version: 2.3.1
Available version: **2.3.2**

An update for Anaconda Navigator is now available.
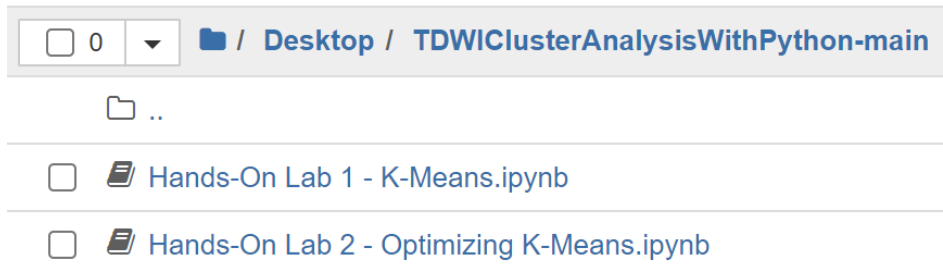
Dismiss	Update now

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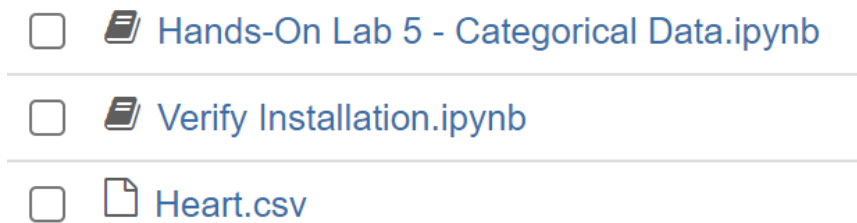
5. Within the Jupyter browser, navigate to where you copied the course file folder:
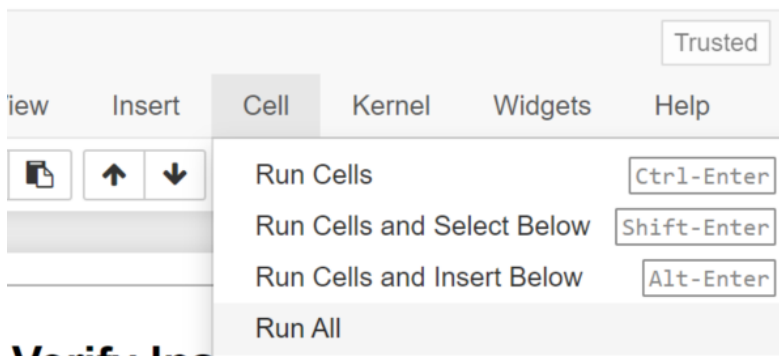
Jupyter

| Files | Running | Clusters |

Select items to perform actions on them.

| ☐ 0 | ▼ | 📁 / **Desktop** / **TDWIClusterAnalysisWithPython-main** |

📁 ..

☐ 📓 Hands-On Lab 1 - K-Means.ipynb

☐ 📓 Hands-On Lab 2 - Optimizing K-Means.ipynb

6. Click on the "Verify Installation.ipynb" entry:

☐ 📓 Hands-On Lab 5 - Categorical Data.ipynb

☐ 📓 Verify Installation.ipynb

☐ 📄 Heart.csv

7. Run all the cells in the notebook:

Verify Installation

Trusted

| iew | Insert | Cell | Kernel | Widgets | Help |

| 📋 | ↑ | ↓ |

Run Cells   Ctrl-Enter

Run Cells and Select Below   Shift-Enter

Run Cells and Insert Below   Alt-Enter

Run All

Verify Ins

8. Your output should look like the following, with no errors:

Verify Installation

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

```
In [1]:    ▶|  from prince import FAMD
               from plotnine import ggplot
```

Congratulations! You are now ready for the class!