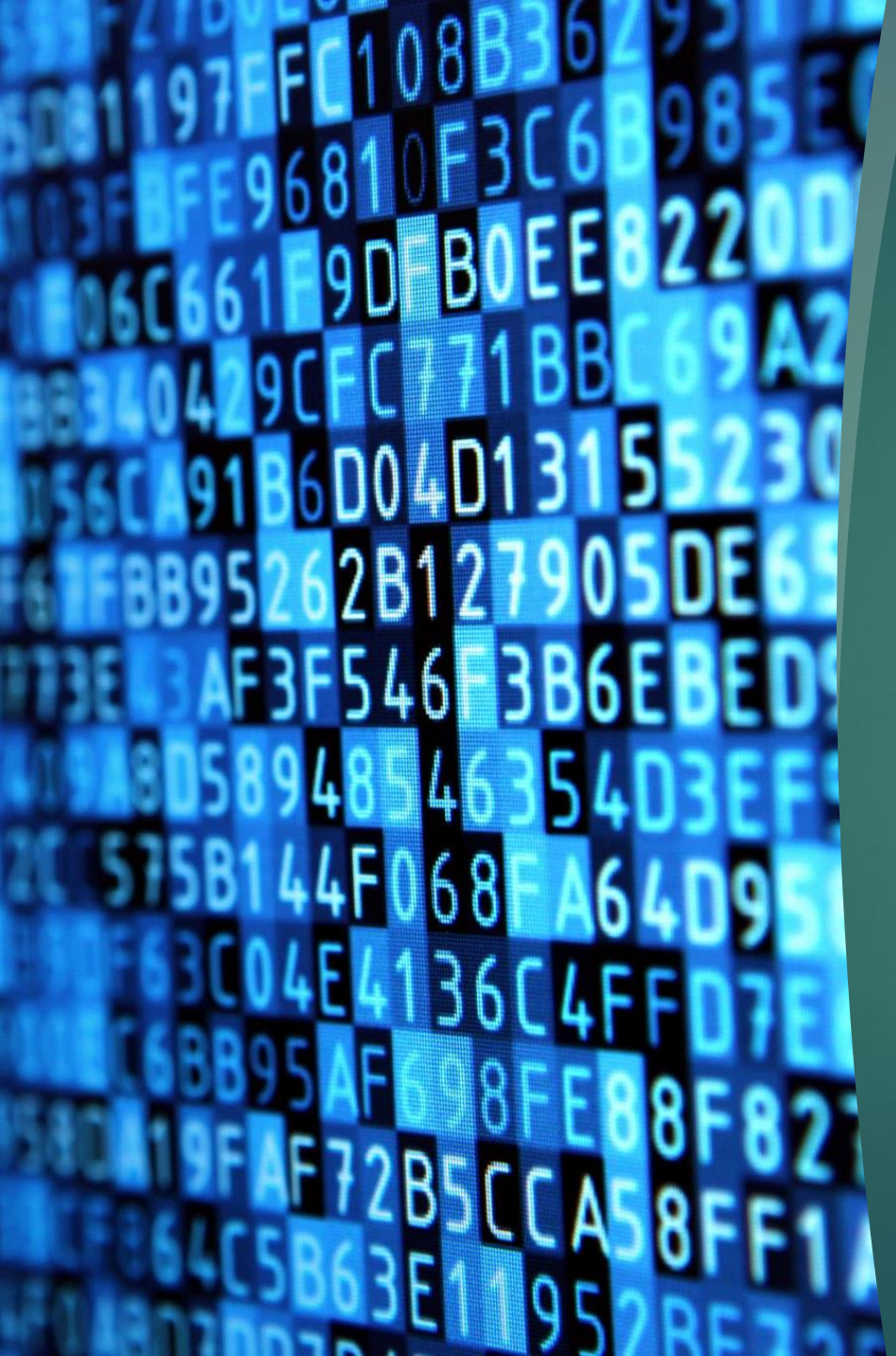


Introduction to Data Science and Machine Learning

COMP-1702

Module 4: Tools

Module 4: Tools

Learning Outcomes

- By the end of this module, you should be able to:
  - Use industry standard tools to build notebooks.
  - Use a plot to visualize and gain insights into data.
- What are you going to learn in this module?
  - You'll learn how to install and use the Anaconda distribution & Jupyter Notebook, which is the starting point for many DSML tools.
- Why are you going to learn this?
  - To become familiar with setting up the tools you'll use in subsequent courses.

Module 4: Tools

Options

- There are different ways you can install Python, Jupyter Notebook, NumPy, Pandas & Matplotlib on your computer.
- One way is to install Python on its own, then use Python's built in PIP tool to install the rest.
- This gives very fine grained control over the versions you install and how they get installed.
- This is also the method most prone to problems and dependency issues.
- Since the DSML program isn't about becoming a programmer, but becoming a data scientist, we'll use the industry standard Anaconda distribution instead, which packages everything you'll need to proceed.

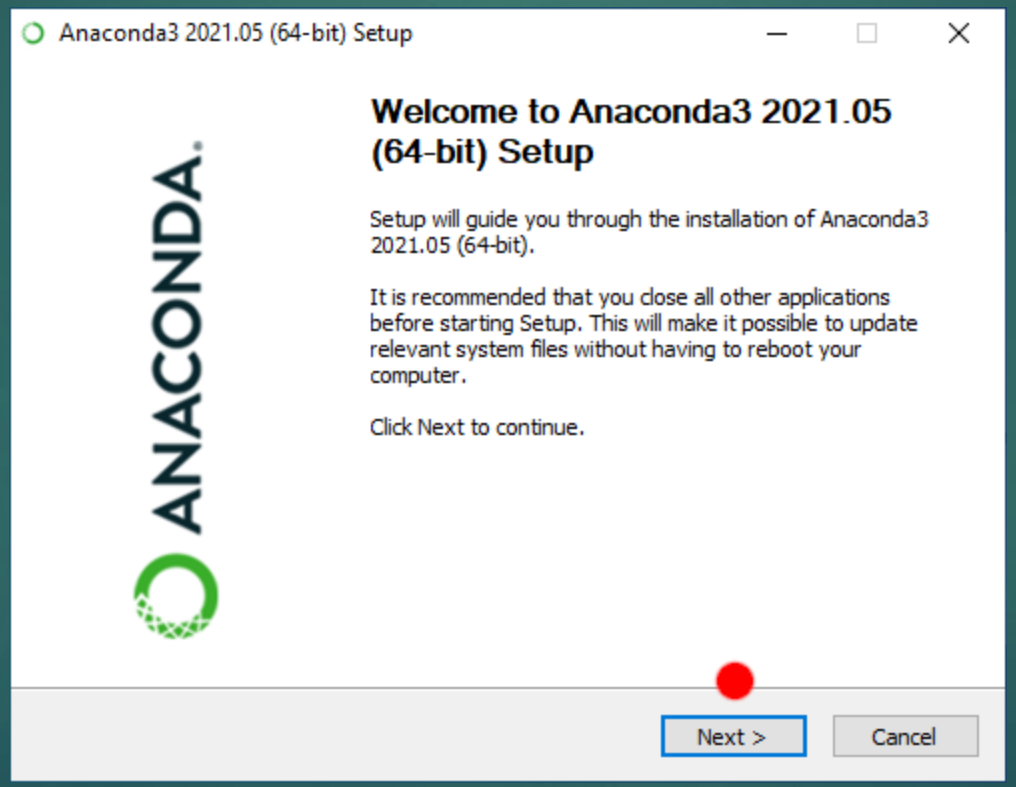
Module 4: Tools

Anaconda Distribution

- Please navigate to the following URL:
  - https://www.anaconda.com/products/individual
- Click the "Download" button on the page, or scroll all the way down to the bottom.
- Choose the 64 bit graphical installer for your platform.
- 32 bit has been obsolete for quite some time, and is only available for backward compatibility with older systems.
- Whenever you have a choice, you should be installing the 64 bit version of applications.

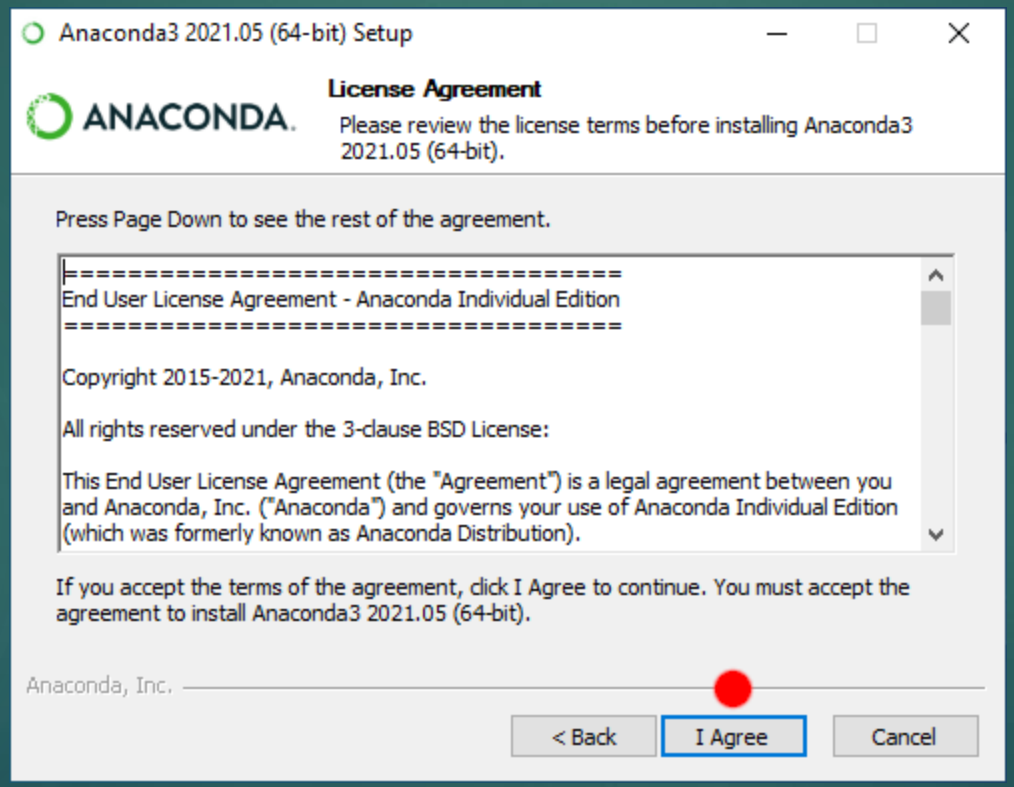
Module 4: Tools

Anaconda Distribution



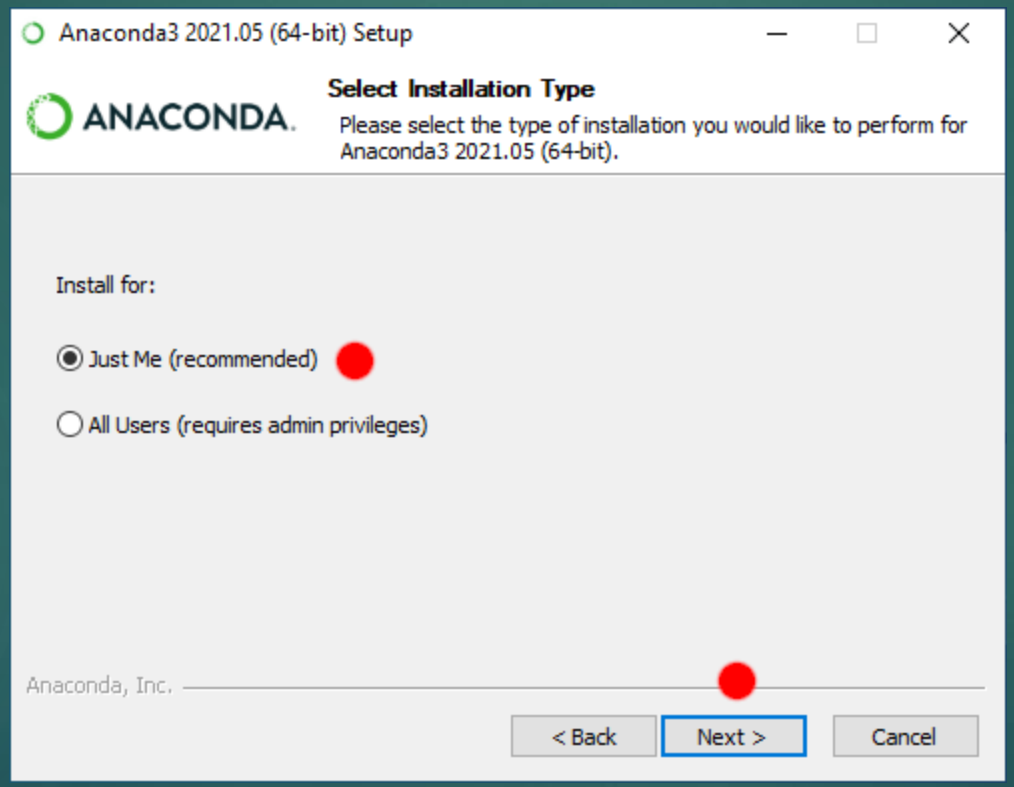
Module 4: Tools

Anaconda Distribution



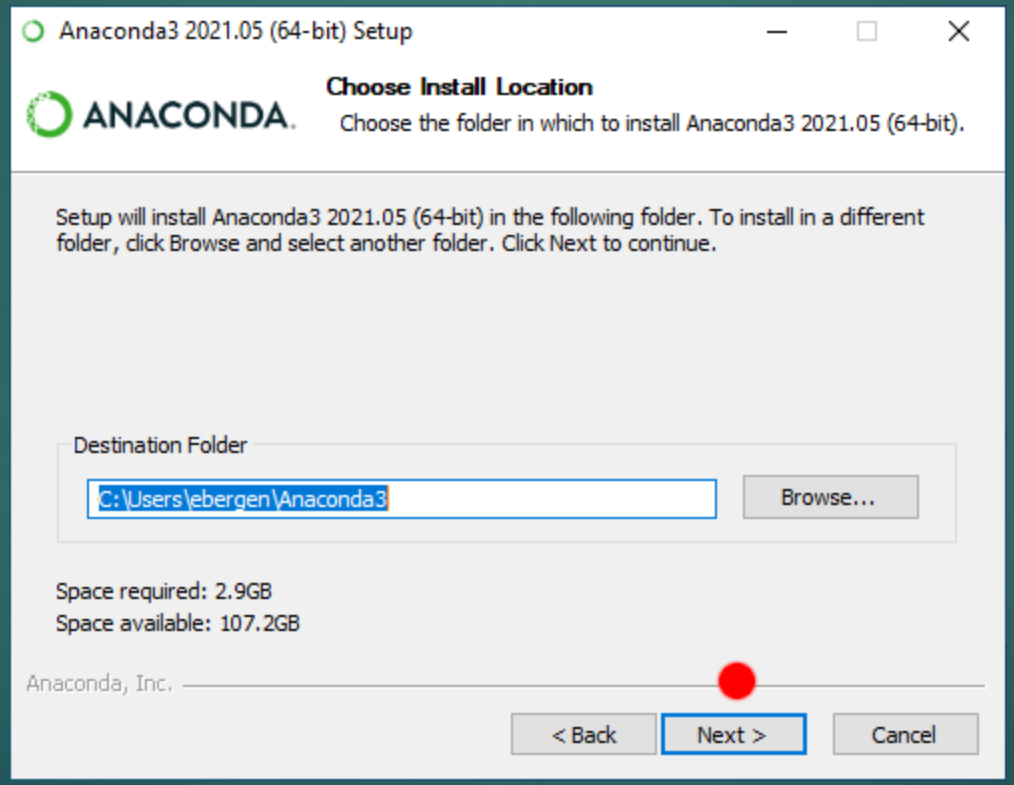
Module 4: Tools

Anaconda Distribution



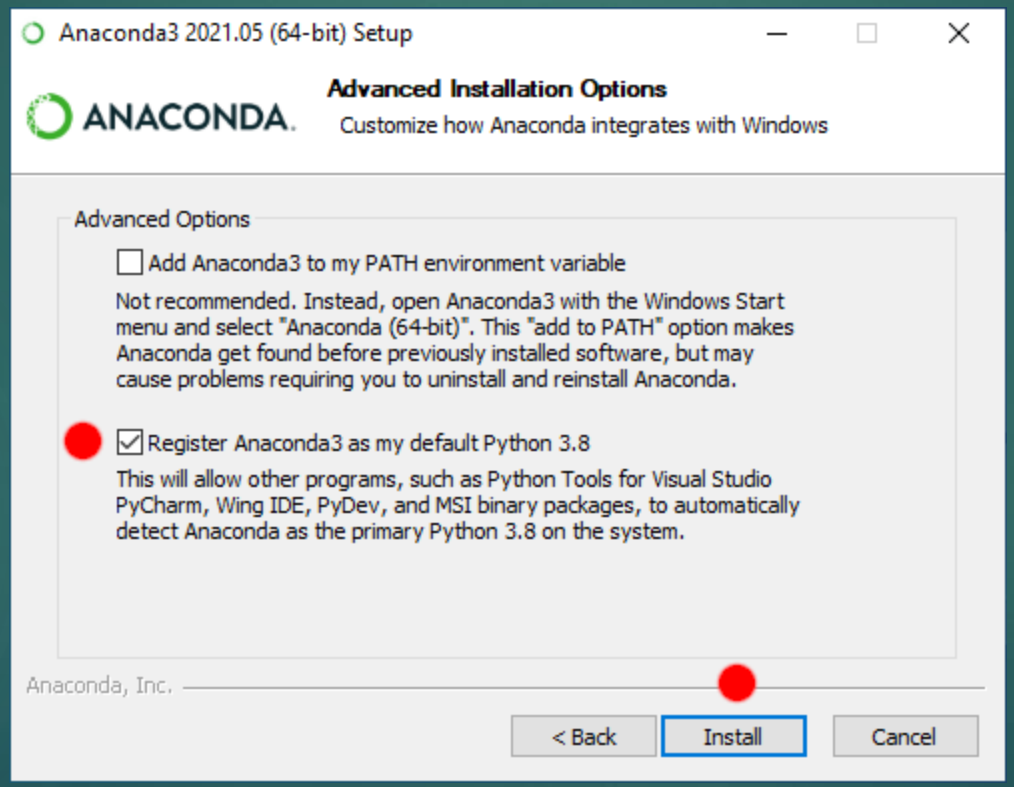
Module 4: Tools

Anaconda Distribution



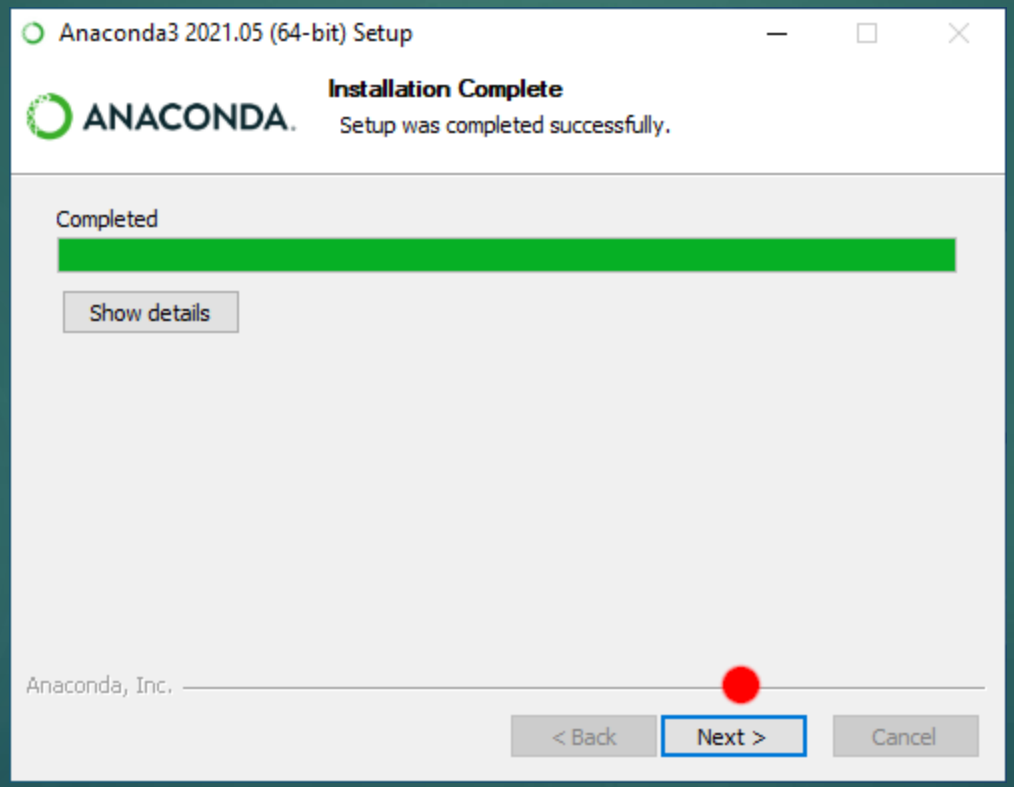
Module 4: Tools

Anaconda Distribution



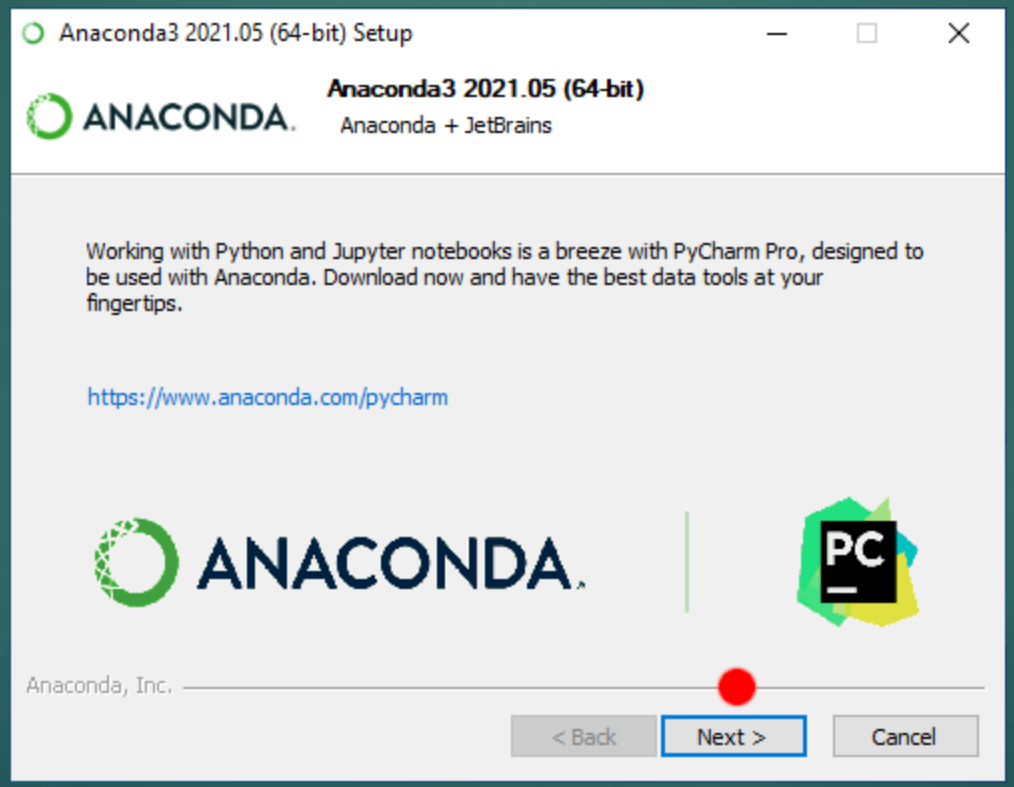
Module 4: Tools

Anaconda Distribution



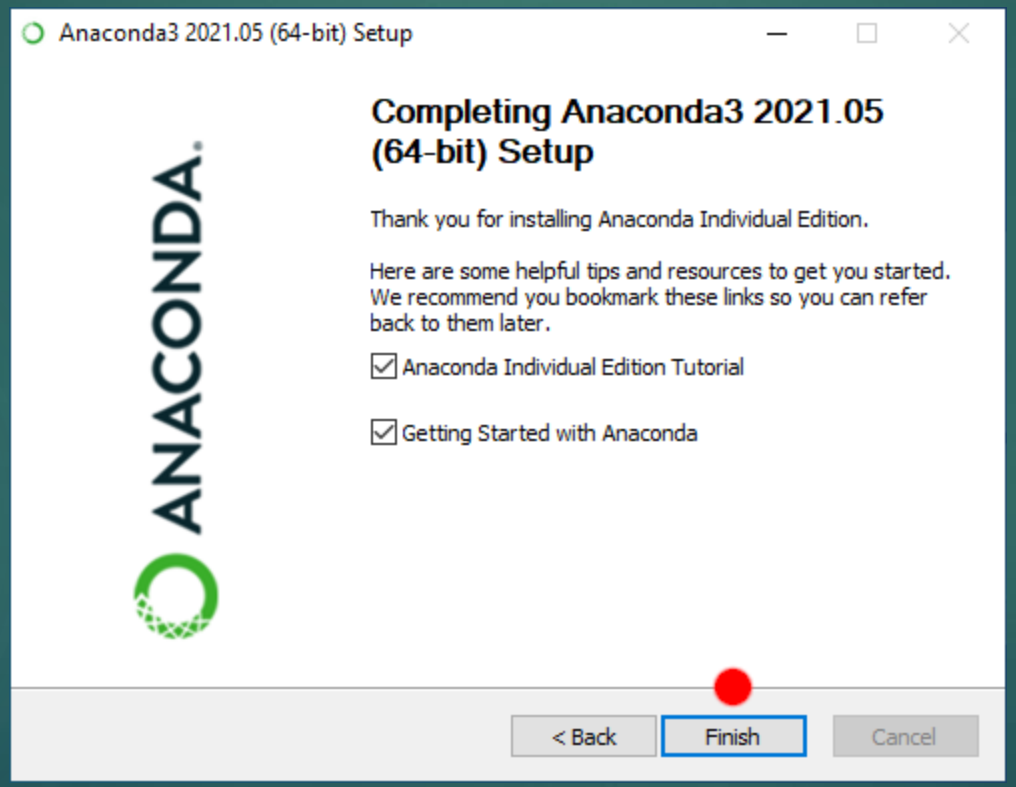
Module 4: Tools

Anaconda Distribution



Module 7: Tools

Anaconda Distribution



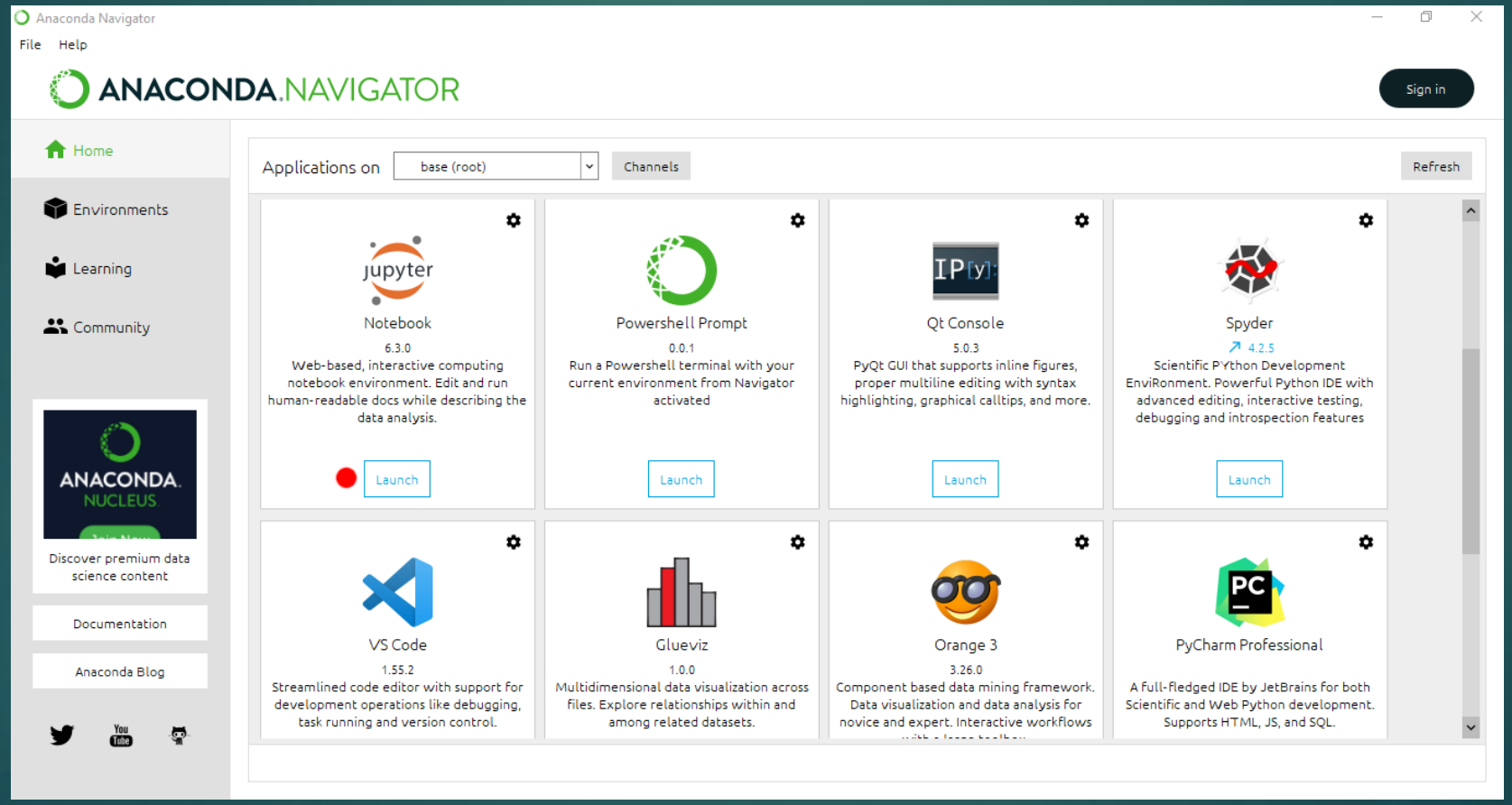
Module 4: Tools

Anaconda Navigator

- Now that you have successfully installed the Anaconda distribution of Python and related tools, lets do a little test run!
- Look for the Anaconda navigator entry in your Windows start menu, and select it.
- Choose Jupyter Notebook from the app screen, see screenshot on next slide…

Module 4: Tools

Anaconda Navigator



Module 4: Tools

Jupyter Notebook

- When Jupyter Notebook first launches, it may prompt you to choose which web browser to use (if you have more than one web browser on your computer).
- Jupyter Notebook is web based, so it will always open in the browser you choose.
- Open your first notebook by clicking on "New" in the top right, and choose "Python3".
- A new empty notebook should open for you.

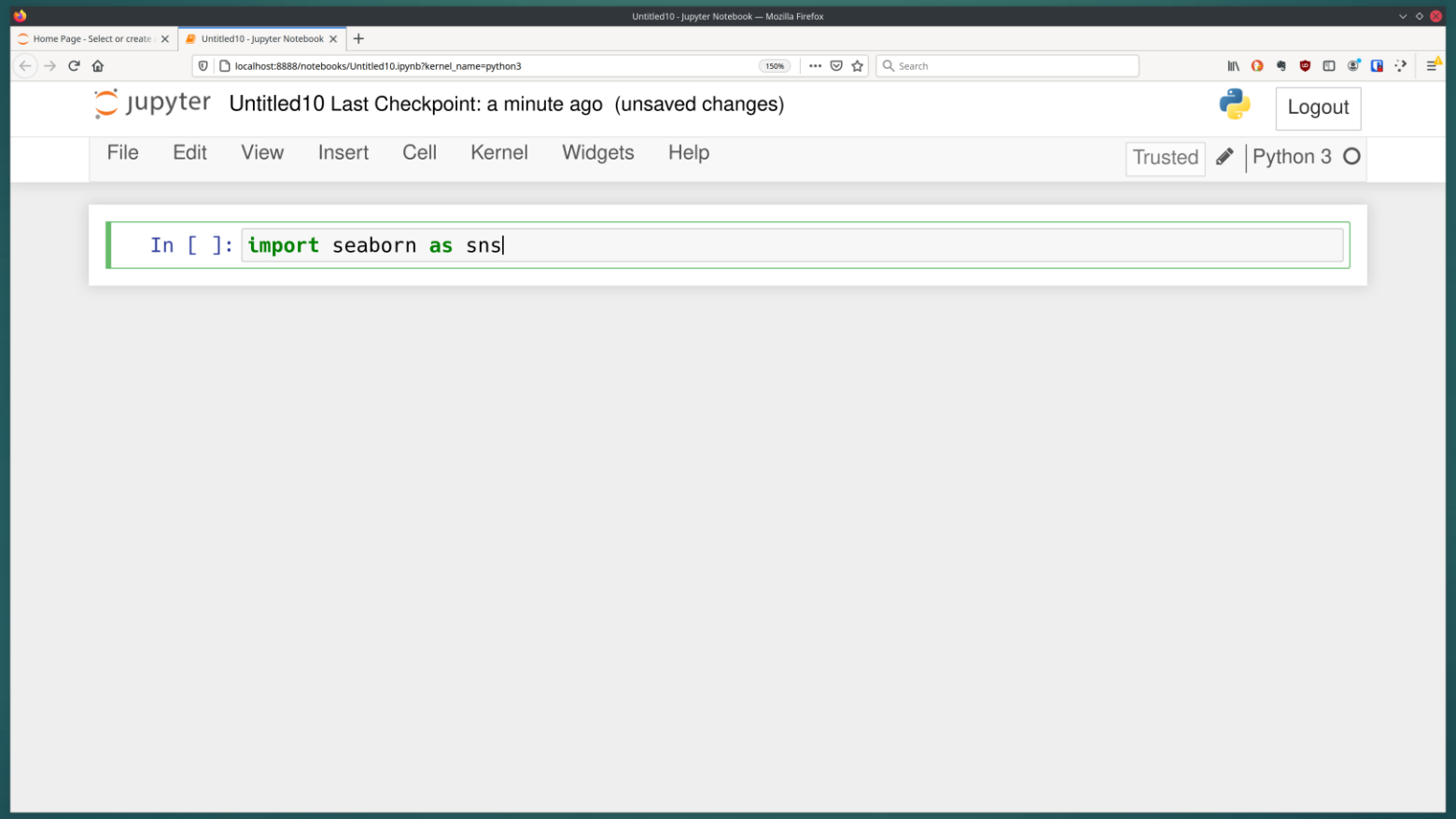
Module 4: Tools

Jupyter Notebook

- Execute selected cells:
  - ctrl-enter
- Execute current cell:
  - shift-enter
- Execute a cell and insert new one below:
  - alt-enter
- Insert a markdown cell to document your code:
  - Cell -> Cell Type -> Markdown
- Execute all cells in the notebook:
  - Cell -> Run All

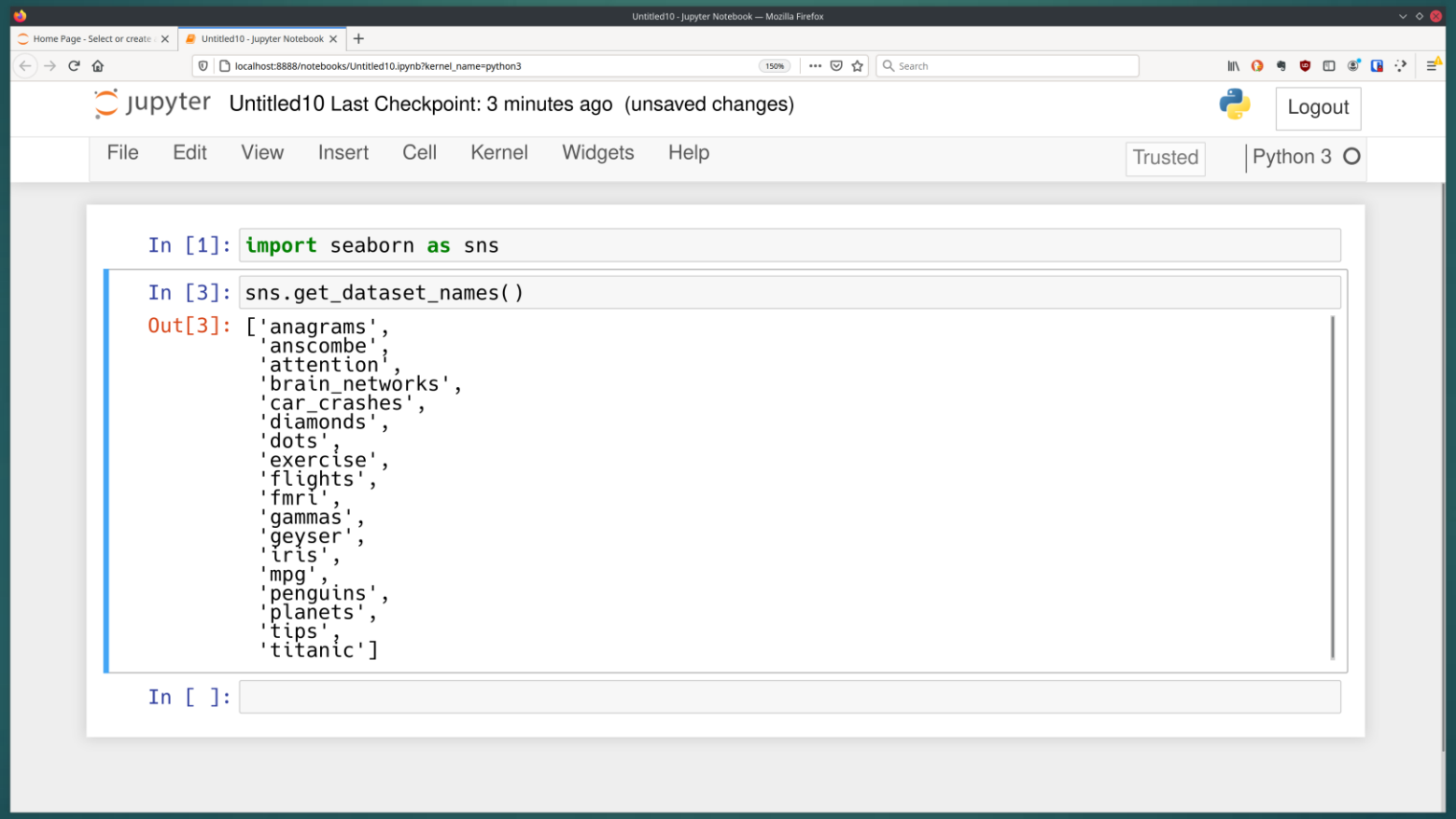
Module 4: Tools

Jupyter Notebook



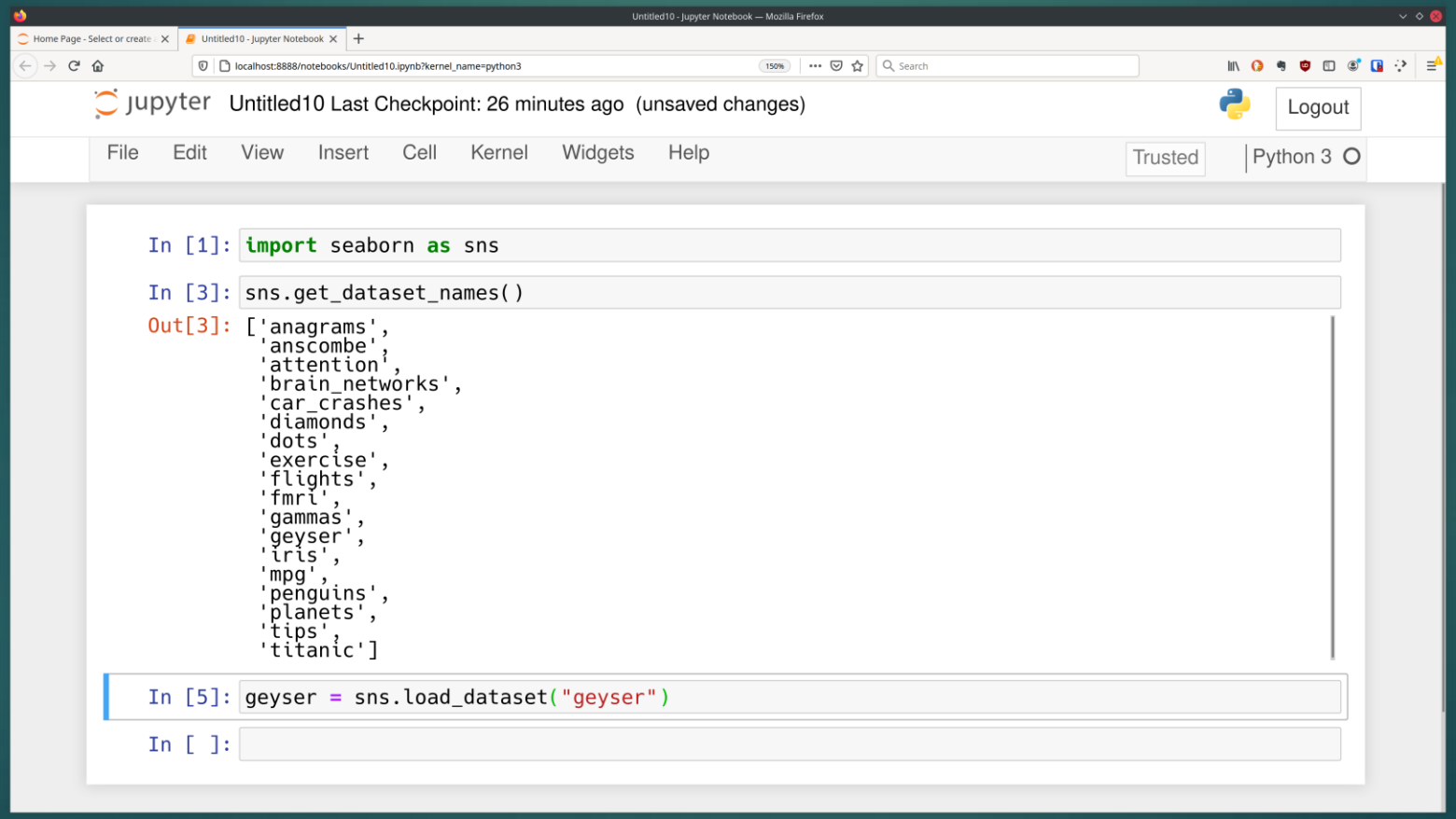
Module 4: Tools

Jupyter Notebook



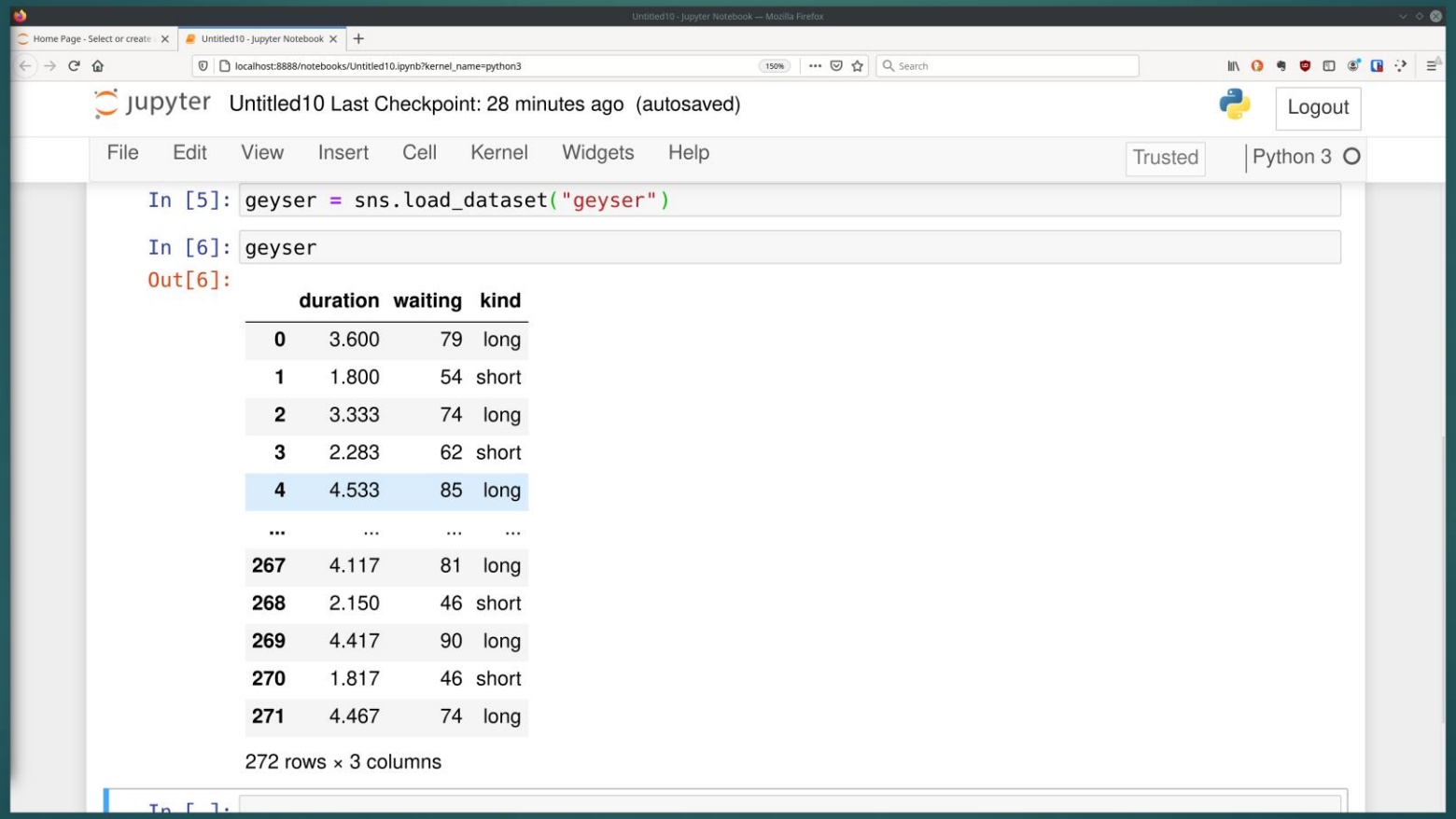
Module 4: Tools

Jupyter Notebook



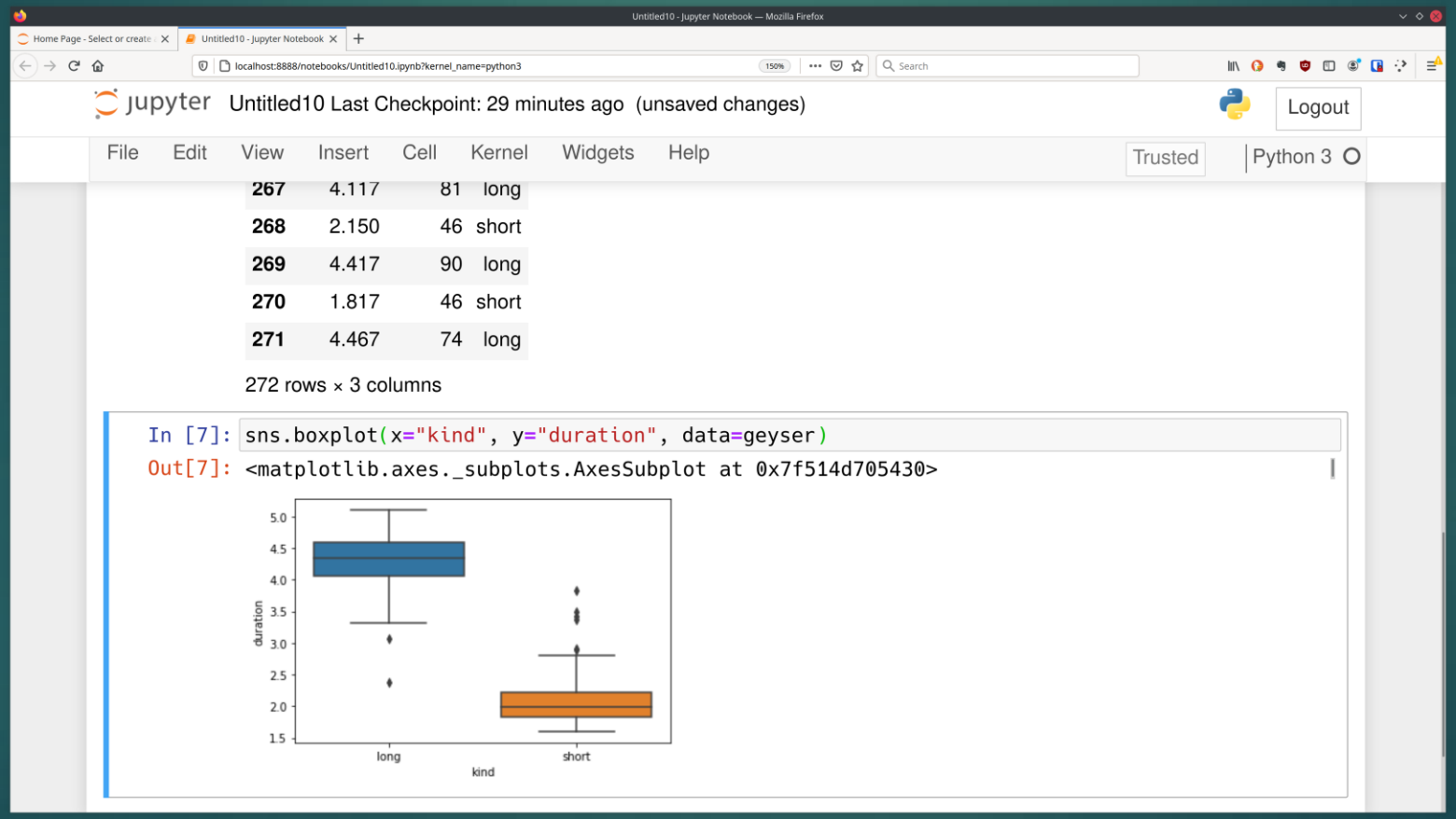
Module 4: Tools

Jupyter Notebook



Module 4: Tools

Jupyter Notebook



Module 4: Tools

Assignment

- If you were able to get the boxplot to appear, then you've successfully installed the Anaconda distribution and accessed data science tools for the first time! This is the first assignment for this week.
- For the second assignment, please navigate to the following link and work through the tutorial. Skip the installation steps, since you've already done that.
  - https://www.pybloggers.com/2018/04/jupyter-notebook-for-beginners-a-tutorial/
  - Start at "Example Analysis" and stop before "Sharing Notebooks".
  - The fortune500.csv file is attached to the dropbox for you.
- When you're finished, please submit your notebook to the dropbox.
- Be prepared to demonstrate your notebook during face to face evaluation.

Module 4: Tool

Assignment

- We'll be covering Jupyter Notebook, NumPy, Pandas, Matplotlib, Seaborn and other DSML tools in a lot more detail in the next course, so don't worry if the tutorial seemed a little complicated for now.
- Review the assignment instructions and associated rubrics in Module 4 in Learn.