



Workshop 1 – Intro: Anaconda, Jupyter, Pandas, Scikit-Learn

Advanced Analytics and Applications [AAA]

General Information

GitHub: Workshop material will be uploaded to GitHub



https://github.com/IS3UniCologne/Advanced_Analytics_And_Applications_2023

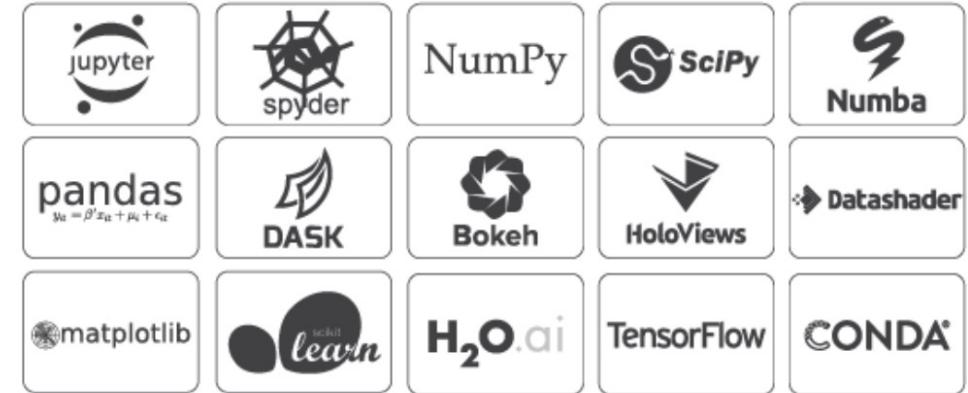
Tools: Python, Anaconda, Jupyter Notebooks

Libaries: Numpy, Pandas and Scikit-Learn

Anaconda: Why Anaconda?

The open-source **Anaconda Individual Edition** (formally Anaconda Distribution) is the easiest way to perform Python/R data science and machine learning on Linux, Windows, and Mac OS X. With over 19 million users worldwide, it is the industry standard for developing, testing, and training on a single machine, enabling *individual data scientists* to:

- Quickly download 7,500+ Python/R data science packages
- Manage libraries, dependencies, and environments with **Conda**
- Develop and train machine learning and deep learning models with **scikit-learn**, **TensorFlow**, and **Theano**
- Analyze data with scalability and performance with **Dask**, **NumPy**, **pandas**, and **Numba**
- Visualize results with **Matplotlib**, **Bokeh**, **Datashader**, and **Holoviews**



Anaconda: Installation

1

Download Link – <https://www.anaconda.com/products/distribution>

2

Python Version – 3.9



3

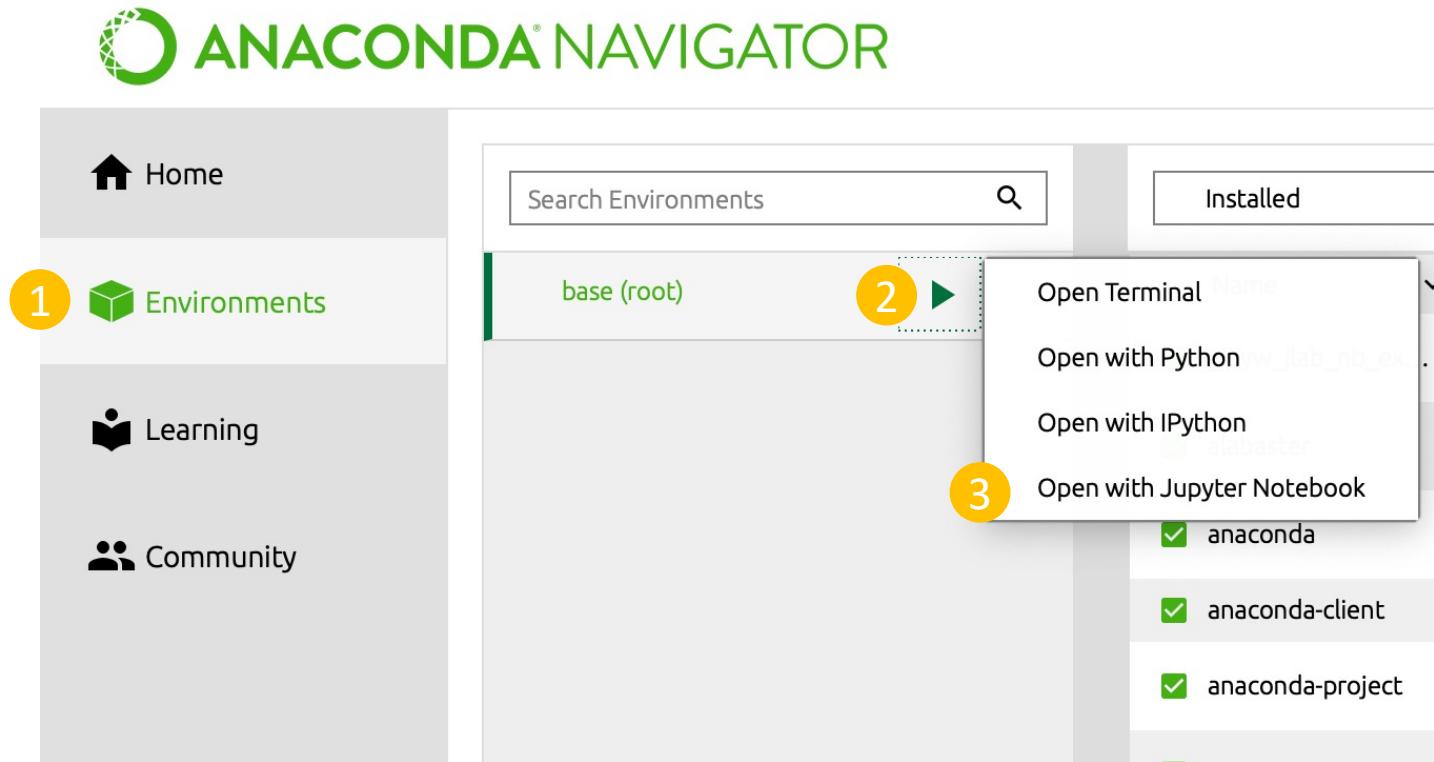
Install on your device – If you have enough space, select all packages and libraries.

4

Start Anaconda-Navigator – If the installation process was successful, you should be able to start “Anaconda Navigator” now.



Jupyter Notebook: Starting a Notebook



Tools: Python, Anaconda, Jupyter Notebooks

Libaries: Numpy, Pandas and Scikit-Learn

Pandas: A Data Scientist's favourite



This is not a tutorial on Python! We expect from you that you are familiar with the fundamentals of Python or at least are willing to learn it by yourselves.

Pandas: A Data Scientist's favourite

- **Main intended use:** Read, clean, transform, write and analyze Data
- Pandas is built on top of **NumPy**, and it can seamlessly work with different well-known Data-Science libraries such as **SciPy**, **Matplotlib** and **Scikit-Learn**.
- Core concept of Pandas: Series and DataFrame:

grade	mat_id
1	45
2	12
3	14
1	15

Data Frame



grade	mat_id
1	45
2	12
3	14
1	15

Series



Contact



For general questions and enquiries on **research**, **teaching**, **job openings** and new **projects** refer to our website at www.is3.uni-koeln.de



For specific enquiries regarding this course contact us by sending an email to the **IS3 teaching** address at is3-teaching@wiso.uni-koeln.de