

RGS PYTHON BLOCK COURSE

11.10.2021 – 15.10.2021

Instructor: Niklas Benner	Time: 10:00 – 16:00
Email: nbenner@rwi-essen.de	Place: Room E27 TU Dortmund

Core python (*standard library*)

- data types and their import functions
- operators and statements
- functions and classes

Python modules

- IDEs and terminal
- pip and virtual environments

Webscaping (*selenium, beautifulsoup, requests*)

- dynamic vs. static websites
- error handling, proxy servers

Text Mining (*spacy, sklearn, gensim*)

- preprocessing, tokenization
- sentiment analysis and topic modeling

Data processing (*pandas, numpy*)

- efficient vector processing
- working with data frames
- saving and writing data
- performance tricks

Data visualization (*matplotlib*)

- plot types, sub plots
- legends, customizing

Regressions and ML (*statsmodels, sklearn*)

- Statistics in Python
- Machine Learning Algorithms

Objectives:

This course is aimed at PhD students with no or intermediate Python knowledge. Experience with other languages is helpful but not necessary.

Required software:

- Python <https://www.python.org/downloads/>
- Jupyter Notebook <https://jupyter.org/>
- PyCharm <https://www.jetbrains.com/de-de/pycharm/>
(You can get the professional version for free with a student mail account.)
- Notepad++ <https://notepad-plus-plus.org/downloads/>

Useful literature:

- VanderPlas, Jake. Python data science handbook: Essential tools for working with data. ” O’Reilly Media, Inc.”, 2016.
- Mitchell, Ryan. Web scraping with Python: Collecting more data from the modern web. ” O’Reilly Media, Inc.”, 2018.
- Hammond, Michael. Python for Linguists. Cambridge University Press, 2020.