

# PYTHON BLOCK COURSE

23. - 25. + 30. - 31. May 2022

or

30. May - 03. June 2022

---

**Instructor:** Niklas Benner

**Email:** [benner@statistik.tu-dortmund.de](mailto:benner@statistik.tu-dortmund.de)

**Time:** 10:00 – 16:00

**Place:** To be determined

---

## 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 (via Anaconda) <https://www.anaconda.com/products/individual/>
- 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.