JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

GGS
Gießener Graduiertenzentrum
Sozial-, Wirtschafts- und Rechtswissenschaften

Workshop

Data Preparation with Python

Organisational details

Instructor:	Dr Tobias Keller
Dates:	September 8, 12, and 15, 2025, 9.00 am – 5.00 pm respectively
Venue:	Room 001, Licher Str. 68, 35394 Giessen
ECTS:	3
Max. participants:	10

Objectives

The amount of time required bringing the data into shape for machine learning and artificial intelligence algorithms or statistical analysis is often underestimated. Furthermore, introductions to data science typically focus on the methods and algorithms and do not cover the required data preparation appropriately.

This workshop aims at enabling participants to go beyond the unrealistically clean datasets provided in data science and machine learning tutorials. Instead, participants learn how to handle data as they would face it in real-life situations in research and business, where errors, inconsistencies, incompleteness, duplicates and many more problems are commonplace. They learn how to combine data from different sources and how to perform computations, aggregations, and other typical data preparation steps efficiently. Finally, participants are introduced to special data pre-processing steps required for machine learning.

Having completed this course will give participants an edge in the labour market, where most newcomers have little experience with real-life datasets – especially those aiming for a career in consulting or other areas related to data science and artificial intelligence.

This course is also an ideal complement for participants taking the course "Machine Learning with Python".

Content and methods

The course covers the typical data preparation techniques required for analytics:

- Loading and joining data from different types of data sources
- Data types and conversions
- Filtering
- Computations and data modifications

JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

GGS
Gießener Graduiertenzentrum
Sozial-, Wirtschafts- und Rechtswissenschaften

- Aggregations
- Pivoting / reshaping
- Handling inconsistencies and errors in the data
- Time series operations
- Special pre-processing operations for machine learning

We may emphasize or skip topics based on questions or suggestions during the workshop and based on the pace of the group.

Hands-on tutorials and exercises will make up about 90 per cent of the course time. Participants will complete those using their own computers. Please see the preparation requirements below for a list of software that needs to be installed to that end. The exercises will be based on exemplary datasets that will be provided to the participants before the course.

Requirements to register for the course

- Knowledge of the topics covered in the previous course, "Introduction to Python (for Data Preparation)", is required for participating in this follow-up course
- In particular, the participants need to be familiar with the following concepts and methods:
  - Numeric data types
  - String manipulation and string formatting
  - Utilizing methods and method chaining
  - Working with boolean data types and boolean operators
  - Lists and their functionality
  - Dictionaries and their usage
  - Conditional statements with if statements
  - Loop structures
  - List comprehensions for concise data manipulation
  - The creation and use of functions
  - Understanding and implementing lambda functions
  - Using modules and packages

To gain the ECTS credit points participants have to:

- Refresh their knowledge from the previous course, "Introduction to Python for Data Preparation",
- Follow the installation instructions provided in time before the course,
- Download and extract the course material and exemplary datasets that will be provided in time before the course,
- Actively participate during the workshop.

Target group

Doctoral candidates or postdoctoral researchers doing empirical research or intending to work as data scientists.

Course language

English (German, if only German participants)

Please note: As this is not an English language course proficiency in English at the C1 level of competency is required.

About the Instructor

Dr Tobias Keller has been working as data scientist at ZERO.ONE.DATA, the big data startup of Deutsche Bahn AG since 2016. He consults on and applies machine learning and statistics for artificial intelligence systems in a big data environment. Furthermore, he teaches data science at Deutsche Bahn, in the doctoral education programmes at Justus Liebig University Giessen, and in the doctoral programme and master programme at WHU – Otto Beisheim School of Management. His research interests include machine learning and artificial intelligence, finance and accounting, strategic management.

Registration

By August 25, 2025 via e-mail at info@ggs.uni-giessen.de.