

Ljubljana Doctoral Winter School

12 – 16 February 2024

14:00 – 18:00 (CET, Ljubljana)

Introduction to Data Science with Python (ECTS: 4)

Matej GUID & Martin MOZINA, University of Ljubljana, Faculty of Computer and Information Science, Slovenia

Aims of the course

The course Introduction to Data Science with Python gives an overview of some of the basic topics in data science, such as data analysis, data visualisation, machine learning, and time series forecasting. The course is designed for students who want to learn about Python's powerful data science ecosystem to apply data analysis techniques, information visualisation, and inferential statistical analysis to gain new insights into data. The course is taught in a tutorial format. The emphasis is not on computer programming, but mainly on the use of various practical tools and libraries in the Python programming language. They will be introduced through various case studies and practical examples from different fields of economics.

Syllabus

Students will learn how to use the popular pandas data science library and jupyter notebooks as a working environment for data analysis, as well as how to effectively use data handling functions. At the end of the course, students will be able to load data from a file or retrieve it from an online service, cleanse and manipulate it, perform basic inferential statistical analysis, and create comprehensive data visualisations. They will also know how to use powerful libraries such as scikit-learn and statsmodels to apply machine learning techniques such as clustering, classification and regression, and to perform time series forecasting. They will learn how to use the Python data science ecosystem in several practical case studies, such as market basket analysis, portfolio optimization, and online advertising in social networks.

The course consists of five sessions. Each session lasts about four hours.

Session 1: Introduction to Python data science ecosystem

Session 2: Data wrangling and data exploring

Session 3: Data visualizations and time series analysis

Session 4: Machine learning and time series forecasting

Session 5: Practical case studies from finance and marketing

Course delivery

Online lectures will be held on ZOOM. Before the start of the course, the course leaders will provide instructions for installing Python using the Anaconda distribution. This installation comes with many useful packages and is easy to install.

Course materials / List of readings

Students will receive jupyter notebooks that illustrate various concepts and data science techniques. The recommended literature:

- Wes Mckinney. Python for Data Analysis, 2nd Edition. O'Reilly Media, 2017.
- Grus, Joel. Data Science from Scratch: First Principles with Python. O'Reilly Media, 2015.

Course credit

Grading is based on four graded assignments, which are assigned at the end of each session. Detailed instructions are given during the course.

Course leaders

GUID Matej is an assistant professor at the Faculty of Computer and Information Science at the University of Ljubljana. He has extensive experience as a leading scientist, researcher, mentor, and lecturer in the fields of Artificial Intelligence and Data Science.

MOZINA Martin is a senior data scientist in a multinational company and a senior researcher at the Faculty of Computer and Information Science at the University of Ljubljana. He has more than ten years of research experience in Artificial Intelligence and Data Science.

Contact: Matej.Guid@fri.uni-lj.si