INTRODUCTION TO DATA SCIENCE WITH PYTHON

University *of Ljubljana*

FACULTY OF ECONOMICS

Doctoral seminar (4 ECTS)

Faculty of Economics, University of Ljubljana

September 2018

COURSE SUMMARY

The Introduction to Data Science with Python course surveys some of the foundational topics in data science, such as data analysis, data visualization, machine Learning, and time series forecasting. The course is intended for students who wish to learn about the powerful Python data science ecosystem in order to apply data analysis techniques, information visualization, and inferential statistical analyses to gain new insights into the data. The course is taught in a tutorial format. The focus is *not* on computer programming, but most of all on the use of several practical tools and libraries in Python programming language. They will be introduced through various case studies and practical examples from different fields of economics.

COURSE OBJECTIVES

Students will learn how to use the popular *pandas* data science library and *jupyter notebooks* as their working environment for data analysis, along with the effective use of functions for handling data. At the end of the course, students will be able to load the data from a file or fetch it from an online service, clean it, manipulate it, run basic inferential statistical analyses and produce comprehensive data visualizations. They will also know how to use the powerful libraries such as *scikit-learn* and *statsmodels* in order to apply machine learning techniques such as clustering, classification and regression, and to perform time series forecasting. They will learn about the use of Python data science ecosystem on several practical case studies, such as market basket analysis, portfolio optimization and online advertising on social networks.

SCHEDULE

The seminar consists of five sessions. Each session lasts about four hours and there will be at least one week between two sessions. The seminar will be held between March and May in 2019. The exact dates can be determined according to the preferences of the participants.

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

GRADING

The grading will be based on four graded assignments, each given at the end of the session. Precise instructions will be given during the course.

REQUIREMENTS AND LITERATURE REVIEW

Only a minimal background in programming is expected, and only a minimal knowledge of statistics.

The recommended literature:

- Wes Mckinney. Python for Data Analysis. O'Reilly Media, 2013.
- Grus, Joel. Data Science from Scratch: First Principles with Python. O'Reilly Media, 2015.

INSTRUCTORS

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

Martin Možina is chief data scientist in a multinational corporation and senior researcher at the Faculty of Computer and Information Science, University of Ljubljana. He has more than ten years of research experience from the fields of Artificial Intelligence and Data Science.