Syllabus

Analyzing Data with Python

- **Instructor:** Eszter Somos (somoseszter@gmail.com office hours: Tuesday 18-19.30 by appointment)
- **Credits:** 2 (4 ECTS)
- **Term:** Spring 2017-2018
- **Course level:** [MA/MSc]
- **Prerequisites:** Data Analysis 1a: Foundation of Data management in R or equivalent experience

Course description

This course will provide a comprehensive introduction to programming with Python, starting from the basics. Beyond confidently using Python, the class will focus on solving problems around Data Processing and Analysis. Additionally, we will discuss for what types of problems Python is the right choice and how to further extend your knowledge after the class. The overarching goal is to equip students with enough programming experience to start working in any area of computation and data-intensive research.

The course will run with a mix of theoretical classes and hands-on sessions organized into 6 classes of 200min. Use of a computer will be required during most of the lectures. Students can use their own laptops or the facilities provided by CEU. If you haven't completed Foundation of Data management in R course or you don't have programming experience this class will be too advanced for you and we recommend you to take the beginner version in the Winter Semester. The classes use knowledge and skills from the previous ones thus it is important to attend all classes. We will provide extra exercises between classes in the form of homework. The course can accommodate a maximum of 30 students.

Course coordinator: Eszter Somos, somoseszter@gmail.com Teaching Assistant: Zsombor Koman, Koman_Zsombor@student.ceu.edu - please send him all inquiries about schedule, classroom and technical requests No. of Credits: 2

Learning outcomes

By the end of the course, students will have experience with techniques which are vital to effective data management:

- The basic syntax and use of Python as a data analysis tool, including writing and executing scripts to automate common tasks, using the IPython interpreter for interactive exploration of data and code, and using the Jupyter notebook to share and collaborate.
- Loading data from a variety of common formats
- Manipulating data efficiently with Pandas
- Basic web scraping
- Use of web APIs
- Use of special python packages such as data visualization libraries

- Performing basic data mining

Grading policy

- Students shall not miss more than 2 classes. Failing to do so will yield an administrative fail grade. (If you have a major impediment please contact the Instructor.)
- To pass, students will need to get at least 50% of the overall grade. The grade will be based 80% on homework and 20% on class participation.

Tentative Calendar

Week 1

- Why Python? Notebook, variables, if/else, logical operators, loops
- Data structures, Numpy

Week 2

- File operations, regular expressions, parsing
- Functions, libraries

Week 3

- Web scraping, using APIs
- Storing data: csv, json, html. Parsing

Week 4

- Data cleaning using Pandas, data transformations
- Basic visualization using Pandas

Week 5

- Data analysis using Pandas
- Advanced data visualization using MatPlotLib, Seaborn

Week 6

- Intro to machine learning, business use cases
- Additional useful and fun python tools

Suggested reading

- Online resources and documentation provided during classes - Bill Mark Lutz, Learning Python, O'Reilly (2013)

- Also available for free online

- Bill Lubanovic, Introducing Python, O'Reilly (2014)

- Wes McKinney, Python for Data Analysis, O'Reilly (2013)

Further information, such as the course website, assessment deadlines, office hours, contact details etc. will be given during the course. The instructor reserves the right to modify this syllabus as deemed necessary any time during the term. Any modifications to the syllabus will be discussed with students during a class period. Students are responsible for information given in class.