

Python for Data Analysis

Lecturer: Ivan Renesto

Course language

English

Course description and objectives

Python is a widely used high-level, general-purpose, interpreted, dynamic programming language.

Through this course you will learn how to manipulate, process, and clean data with Python, using its data-oriented library ecosystem and tools that will lay the foundations to let you become an effective data analyst.

At the end of the course, participants will be able to:

- work with arrays and vectorized computation
- work with tabular or heterogeneous data
- plot and visualize data

Audience

The course is open to all students of Bocconi University. In particular, it is aimed at:

- those who want to approach the world of data analysis;
- students who want to acquire the basic knowledge to develop future expertise in the area of Data Science;
- those who are interested in facing advanced topics in Python or are planning to be part of projects where to extract information from a data set.

Prerequisites

Knowledge of Python basics, having attended the curricular course 30424 Computer Science, or the extracurricular course: Python start, or having equivalent knowledge and skills.

Duration

12 hours



Calendar

Lecture	Date	Time	Room
1	Sat 01/06/2019	09.30 - 11.00	Info AS05
2	Sat 01/06/2019	11.15 - 12.45	Info AS05
4	Sat 08/06/2019	09.30 - 11.00	Info AS05
3	Sat 08/06/2019	11.15 - 12.45	Info AS05
5	Sat 15/06/2019	09.30 - 11.00	Info AS05
6	Sat 15/06/2019	11.15 - 12.45	Info AS05

Syllabus of the course

Lecture	Topics	Book reference
1	Introduction to Anaconda <ul style="list-style-type: none"> - Preliminaries - Install Anaconda - Walk through the Development Environment - Built-in Data Structures and Sequences. <i>Exercises</i>	Ch. 1, 2, and 3
2	Arrays and Vectorized Computation <ul style="list-style-type: none"> - NumPy basics - Working with multidimensional array objects - Indexing, slicing, and transposing arrays - Array-Oriented Programming - Mathematical and statistical methods. 	Ch. 4
3	Tabular and heterogeneous data <ul style="list-style-type: none"> - Pandas basics - Introduction to Series, DataFrame, Index objects - Essential functionalities of pandas library - Summary statistics methods. <i>Exercises</i>	Ch. 5
4	Data Loading, Storage and File Formats <ul style="list-style-type: none"> - Reading and writing data in text format - Binary Data Format - Interacting with Web APIs and Databases. <i>Exercises</i>	Ch. 6

Lecture	Topics	Book reference
5	Data Cleaning and Preparation <ul style="list-style-type: none"> - Handling missing data - Data transformation - String manipulation. <i>Exercises</i>	Ch. 7
6	Plotting and Visualization <ul style="list-style-type: none"> - Data visualization using matplotlib - Data visualization using pandas. <i>Final exercise</i>	Ch. 9

Software used

Anaconda version 5.2 | Release Date: May 30, 2018.

At June 2018, it supports Python version 3.6 or 2.7.

Available for Windows, Linux and OS X, for 32 bit or 64 bit systems, can be downloaded here: <https://www.anaconda.com/download/>.

Detailed information on how to install Anaconda [here](#):

- for [Windows](#)
- for [Mac OS](#)
- for [Linux](#)

Suggested bibliography

Wes McKinney, *Python for Data Analysis, second edition. Data Wrangling with Pandas, NumPy and IPython*, O'Reilly.

Available seats

This activity is limited to **110** participants. Registrations cannot be carried out once this number has been reached or after closing of the registration period.