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 students of the Master of Science Programs at Bocconi University and is part of the Enhancing Experience activities that are worth 2 credits (subject to 75% attendance and to passing the final test). More specifically, it is targeted 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

16 hours

Teaching mode

It will be possible to join the course in presence and/or in distance, by connecting remotely and following the streaming of the lesson held in the classroom.

The **final test** on the last day of class, however, can **only** be taken **in physical presence**. Online mode will not be provided.

Calendar

Lecture	Date	Time	Room
1	Wed 22/03/2023	18.15 - 19.45	Aula 1.1 (Sarfatti)
2	Tue 28/03/2023	18.15 - 19.45	Aula 1.1 (Sarfatti)
3	Thu 30/03/2023	18.15 - 19.45	Aula 3.1 (Sarfatti)
4	Tue 04/04/2023	18.15 - 19.45	Aula 1.1 (Sarfatti)
5	Thu 06/04/2023	18.15 - 19.45	Aula 3.1 (Sarfatti)
6	Tue 18/04/2023	18.15 - 19.45	Aula 1.1 (Sarfatti)
7	Wed 19/04/2023	18.15 - 19.45	Aula 1.1 (Sarfatti)
8	Thu 27/04/2023	18.15 - 19.45	InfoAS04/AS05

Note: lessons will be held in the traditional room and **all the students have to bring their own device**.

Syllabus of the course

Lecture	Topics	Book reference
1	**Introduction to Anaconda** - Preliminaries - Install Anaconda - Walk through the development environment - Built-in data structures and sequences. *Exercises*	**Ch. 1, 2, and 3**

Lecture	Topics	Book reference
2	**Arrays and vectorized computation** - NumPy basics - Working with multidimensional array objects - Indexing, slicing, and transposing arrays - Array-Oriented Programming - Mathematical and statistical methods.	Ch. 4
3	**Plotting and visualization** - Data visualization using matplotlib - Figures and Axes - Saving figures to file - Sub-plots - Multiple line plots - Colors, line styles, axes limits, labels plot title, legend and other chart elements - Histograms.	Ch. 9
4	**Data manipulation with pandas** - Pandas basics - Introduction to Series, DataFrame, Index objects - Essential functionalities of pandas library - Summary statistics methods - Data visualization using pandas. *Exercises*	Ch. 5
5	**Problem requiring data analysis** - Data loading, storage and file formats - Dataset analysis - Reading and writing data in text format - Interacting with Web APIs - Interacting with Databases via pyodbc. *Exercises*	Ch. 6
6	**Data Cleaning and Preparation** - Handling missing data - Data formatting and string manipulation - Data transformation (normalization and binning) - Categorical values *Exercises*	Ch. 7

Lecture	Topics	Book reference
7	**Exploratory Data Analysis** - Descriptive statistics - GroupBy mechanics - The analysis of variance - Correlation between different variables - Pearson correlation and correlation heatmaps. *Exercises*	**Ch. 8, 10, 12**
8	**Final Exam**	

Software used

Anaconda3 2022.10
Anaconda Individual Edition is free version for solo practitioners, students and researchers, and supports Python version 3.9.
Available for Windows, Linux and OS X, for 32-bit or 64-bit systems, can be downloaded from here:
https://www.anaconda.com/products/distribution#Downloads.

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 and reserved to **students of the Master of Science Programs**. Registrations cannot be carried out once this number has been reached or after closing of the registration period.