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. 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

16 hours

Teaching mode

This course will be only taught <u>in person</u>. Online mode will not be provided.

Calendar

Lecture	Date	Time	Room
1	Thu 02/11/2023	18.15 – 19.45	N31 (Velodromo)
2	Tue 07/11/2023	18.15 – 19.45	N31 (Velodromo)
3	Thu 09/11/2023	18.15 – 19.45	N31 (Velodromo)
4	Tue 14/11/2023	18.15 – 19.45	N31 (Velodromo)
5	Thu 16/11/2023	18.15 – 19.45	N31 (Velodromo)
6	Tue 21/11/2023	18.15 – 19.45	N31 (Velodromo)
7	Thu 23/11/2023	18.15 – 19.45	N31 (Velodromo)
8	Tue 28/11/2023	18.15 – 19.45	InfoAS04

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 Visual Studio Code** - Preliminaries - Install Visual Studio Code - Walk through the development environment - Built-in data structures and sequences. *Exercises*	**Ch. 1, 2, and 3**
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**

Lecture	Topics	Book reference
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
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

Python, version 3.9+. Current version is 3.11.5.
Python interpreter can be downloaded for free from here:
https://www.python.org/downloads/.

Microsoft Visual Studio Code (VS Code). Current version is 1.82.
Visual Studio Code is a free coding editor that helps to start coding quickly. It supports multiple programming languages, and the use of a Python web-based interactive computing platform (Jupyter Notebook).
Supported in: Windows 10 and 11 (32-bit and 64-bit), macOS versions with Apple security update support, Linux Ubuntu Desktop 10.64, Debian 9, Red Hat Enterprise Linux 7, CentOS 7, Fedora 34.

VS Code can be downloaded from here: https://code.visualstudio.com/

Suggested bibliography

McKinney W., *Python for Data Analysis, second edition. Data Wrangling with Pandas, NumPy and IPython*, O'Reilly Media, 2017.

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.