

# 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

The course will be held in distance learning mode. It will be possible to follow the live streaming (Live Session) of each lesson accessing through Blackboard to the corresponding virtual room.

## Calendar

Lecture	Date	Time
1	Thu 10/06/2021	18.40 – 20.10
2	Tue 15/06/2021	18.40 – 20.10
4	Thu 17/06/2021	18.40 – 20.10
3	Tue 22/06/2021	18.40 – 20.10
5	Thu 24/06/2021	18.40 – 20.10
6	Tue 29/06/2021	18.40 – 20.10
7	Thu 01/07/2021	18.40 – 20.10
8	Tue 06/07/2021	18.40 – 20.10

## Syllabus of the course

Lecture	Topics	Book reference
1	<b>Introduction to Anaconda</b> <ul style="list-style-type: none"> <li>- Preliminaries</li> <li>- Install Anaconda</li> <li>- Walk through the development environment</li> <li>- Built-in data structures and sequences.</li> </ul> <i>Exercises</i>	Ch. 1, 2, and 3
2	<b>Arrays and vectorized computation</b> <ul style="list-style-type: none"> <li>- NumPy basics</li> <li>- Working with multidimensional array objects</li> <li>- Indexing, slicing, and transposing arrays</li> <li>- Array-Oriented Programming</li> <li>- Mathematical and statistical methods.</li> </ul>	Ch. 4

Lecture	Topics	Book reference
3	<b>Plotting and visualization</b> <ul style="list-style-type: none"> <li>- Data visualization using matplotlib</li> <li>- Figures and Axes</li> <li>- Saving figures to file</li> <li>- Sub-plots</li> <li>- Multiple line plots</li> <li>- Colors, line styles, axes limits, labels plot title, legend and other chart elements</li> <li>- Histograms.</li> </ul>	Ch. 9
4	<b>Data manipulation with pandas</b> <ul style="list-style-type: none"> <li>- Pandas basics</li> <li>- Introduction to Series, DataFrame, Index objects</li> <li>- Essential functionalities of pandas library</li> <li>- Summary statistics methods</li> <li>- Data visualization using pandas.</li> </ul> <p><i>Exercises</i></p>	Ch. 5
5	<b>Problem requiring data analysis</b> <ul style="list-style-type: none"> <li>- Data loading, storage and file formats</li> <li>- Dataset analysis</li> <li>- Reading and writing data in text format</li> <li>- Interacting with Web APIs</li> <li>- Interacting with Databases via pyodbc.</li> </ul> <p><i>Exercises</i></p>	Ch. 6
6	<b>Data Cleaning and Preparation</b> <ul style="list-style-type: none"> <li>- Handling missing data</li> <li>- Data formatting and string manipulation</li> <li>- Data transformation (normalization and binning)</li> <li>- Categorical values</li> </ul> <p><i>Exercises</i></p>	Ch. 7
7	<b>Exploratory Data Analysis</b> <ul style="list-style-type: none"> <li>- Descriptive statistics</li> <li>- GroupBy mechanics</li> <li>- The analysis of variance</li> <li>- Correlation between different variables</li> <li>- Pearson correlation and correlation heatmaps.</li> </ul> <p><i>Exercises</i></p>	Ch. 8, 10, 12
8	<b>Final Exam</b>	

## Software used

Anaconda3 2020.11

Anaconda Individual Edition is free version for solo practitioners, students and researchers, and supports Python version 3.8.

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

## 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 **60** participants. Registrations cannot be carried out once this number has been reached or after closing of the registration period.