Python for data science

ECTS : 3

Description du contenu de l'enseignement :

The course is organised as follows.

1 - Introduction to Python Programming

This first part introduces the fundamentals of Python programming. It covers topics such as working with basic built-in types (numbers, strings, booleans, ...), control flow statements, writing reusable code (functions), handling errors and exception that can occur during the execution of Python code, advanced data structures (lists, sets, dictionaries, ...), ...

2 - Scientific Computing With NumPy

This part focuses on using NumPy, a scientific computing package that provides a wide assortment of useful and highly-optimized routines for working with multi-dimensional arrays (matrices, tensors, ...), linear algebra, statistics and random simulation, and much more.

3 - Processing Tabular Data With pandas

The third part of the course is dedicated to pandas, a fundamental Python package when it comes to data science and data analysis. pandas provides functionalities for efficient manipulation of data frames, i.e., tabular data (stored in csv files, Excel sheets, ...). With the help of pandas, you can easily conduct tasks such as data cleaning (filling missing data, replacing outliers, ...), reshaping, merging, ...

4 - Visualizing Data With Matplotlib and seaborn

The last part of the course is a quick introduction to data visualization functionalities in Python using the Matplotlib and seaborn packages. Data visualization is a very powerful tool for making sens of large volumes of data, identifying patterns, and extracting useful insights that can help understand and solve real-world business cases.

Compétence à acquérir :

By the end of this course, you will be able to

- Write and understand entry-level to intermediate-level code in the Python programming language
- Use NumPy for scientific computing and efficient manipulation of multi-dimensional arrays and matrices
- Use pandas to load, manipulate, and analyze tabular data
- Use Matplotlib and seaborn to visualize data

Mode de contrôle des connaissances :

You will be evaluated based on a team project (conducted in pairs) in which you will apply the knowledge and skills you acquired during the course. The project takes the form of an exploratory data analysis in which you will work on a tabular data set in order to extract valuable insights that can help solve a business problem.
The expected deliverables of the project are:

- A 5–10 pages report;
- The source code (Jupyter notebooks or Python scripts) of your work, either in a Github repository or as a zip file.

You are expected to present your main findings during a 10-minutes presentation, which will be followed by approximatively 5 minutes of questions.

Document susceptible de mise à jour - 19/02/2026
Université Paris Dauphine - PSL - Place du Maréchal de Lattre de Tassigny - 75775 PARIS Cedex 16

1/1