Course : Big Data - Python for data analysis

Practical course - 3d - 21h00 - Ref. PBD
Price : 2360 CHF E.T.

★★★★⯪ **4,6 / 5**

The Python language provides a scientific ecosystem for statistical processing: from the construction of analysis models to their evaluation and representation. This training course will enable you to analyze data from a variety of sources using Python libraries.

Teaching objectives

At the end of the training, the participant will be able to:

- Understanding the principle of statistical modeling
- Know how to use the main data processing and analysis tools for Python
- Know how to apply best practices in data cleaning and preparation prior to analysis
- Choosing between regression and classification according to data type
- Learn how to set up a simple learning model
- Be able to extract data from a file

Intended audience
Python developers, data center managers, software developers, programmers, data analysts, data scientists.

Prerequisites
Mastery of Python programming. Basic knowledge of statistics or completion of the course "Statistics, mastering the fundamentals" (Ref. STA).

Practical details
Developing/performing analyses with Python, using Pandas, Numpy and SciPy modules.

Course schedule

PARTICIPANTS
Python developers, data center managers, software developers, programmers, data analysts, data scientists.

PREREQUISITES
Mastery of Python programming. Basic knowledge of statistics or completion of the course "Statistics, mastering the fundamentals" (Ref. STA).

TRAINER QUALIFICATIONS
The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS
The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.
Participants also complete a placement test before and after the course to measure the skills they've developed.

1 Introduction to the scientific Python ecosystem

- Overview of Python's scientific ecosystem: must-have libraries.
- Know where to find new bookstores and assess their sustainability.
- The main open source tools and software for data science.

Hands-on work
Installation of Python 3, Anaconda and Jupiter Notebook.

2 Working with data in Python

- Python's scientific foundation: the SciPy Stack.
- Best practices for getting your data science project off to a good start with Python.
- Scientific file formats and libraries for manipulating them.
- Pandas: analysis of tabular data (CSV files, Excel, etc.), statistics, pivots, filters, searches, etc.
- NumPy: numerical calculation and linear algebra (vectors, matrices, images).
- Data extraction, preparation, cleaning.

Hands-on work
Write Python scripts to work with data from files, to apply filters, formatting and cleaning processes.

3 Introduction to modeling

- Steps in building a model.
- Supervised and unsupervised algorithms.
- The choice between regression and classification.

Hands-on work
Integration of Python scripts in the installed environment, for analysis.

4 Model evaluation procedures

- Re-sampling techniques in training, validation and test games.
- Testing the representativeness of training data.
- Performance measurements for predictive models.
- Confusion and cost matrix, ROC and AUC curves.

Hands-on work
Set up dataset sampling. Perform evaluation tests on several supplied models.

5 Supervised algorithms

- The principle of univariate linear regression.
- Multivariate regression.
- Polynomial regression.
- Regularized regression.
- The Naive Bayes.
- Logistic regression.

Hands-on work
Implement regressions and classifications on several types of data.

TEACHING AIDS AND TECHNICAL RESOURCES

- The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
- At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
- A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

6 Unsupervised algorithms

- Hierarchical clustering.
- Non-hierarchical clustering.
- Mixed approaches.

Hands-on work
Unsupervised clustering on multiple datasets.

Dates and locations

REMOTE CLASS
2026 : 22 June, 21 Sep., 30 Nov.