

Updated on 10/17/2024

Sign up

Python Training: Data Analysis

3 days (21 hours)

Presentation

Python is one of the most flexible and versatile languages. Used in countless fields, Python is the language of the year 2022 according to the [TIOBE index](#). Its simple syntax makes it easy to learn, making it [the most widely used](#) programming language in Big Data.

With the rise in popularity of Big Data in recent years, Python has become an essential language. Its ecosystem of APIs enables it to process different types of data (SQL and NoSQL) and drive advanced processing tools.

Our Python: data analysis training course will teach you how to solve problems using data analysis. You'll learn the principles of statistical modeling and processing tools. You'll also see how to extract data from a file and set up a simple learning model.

At the end of this course, you'll be able to manipulate statistical data using the Python language. We will introduce you to the latest version of [Python 3.13](#).

Objectives

- Understand the principle of statistical modeling
- Master data processing and analysis tools for Python
- Extract data from a file
- Apply best practices for data cleansing and preparation
- Set up a simple learning model
- Select classification and regression according to data type
- Optimizing the predictive performance of an algorithm
- Create rankings and selections from large volumes of data

Target audience

- Python developers
- Data center managers
- Software developers
- Programmers
- Data Analysts
- Data Scientists

Prerequisites

- Basic knowledge of Python
- Basic knowledge of statistics

Software requirements

Latest versions of Anaconda and Python installed.

Big Data training program: Python for Data Analysis

Bookstore presentation

- Pandas
- NumPy and SciPy
- Matplotlib
- Scikit-Learn
- Using library functions
- Managing modules and libraries

Data preparation

- The importance of data integrity and preparation
- Reading and writing CSV files
- Import data sets
- Clean and prepare data
- Data formatting
- Building data pipelines

Model your data with Pandas

- Handling your data

- Create columns
- Transforming columns
- Querying, sorting and filtering
- Grouping and aggregation for data synthesis
- Merge and join data from multiple sources
- Concatenation
- Select items
- Replace values

Data visualization with Pandas

- Make comparisons
- Creating graphics :
 - Histogram
 - Pie chart
 - Kernel density graph
 - Hexagonal graph
 - Scatter chart
 - Line chart
 - Boxplot
- Practical work: Finding insights from visualizations

Numerical calculation with NumPY

- Switching from Python lists to NumPy arrays
- Working with multidimensional tables
- Array operations, cutting and distribution
- Linear and matrix algebra functions
- The main mathematical functions

Statistical analysis with SciPy

- Best practices for successful data analysis
- SciPy's 3 statistical classes
- Use descriptive statistics functions
- Calculate z-score and t-score
- Practical work: Analysis of several different dataframes

Development of machine learning models

- The strengths and limitations of machine learning
- Choose your models with the help of visualization
- Optimization functions and preprocessing with SciPy
- Different learning models with SciKit-Learn

Regression with Scikit-learn

- Simple linear regression
- Multiple linear regression
- Logistic regression
- Polynomial regression
- R-squared and EQM to evaluate your model
- Prediction and decision-making

Classification with Scikit-learn

- SVM
- Nearest neighbors
- Random forest
- K-Means

Data algorithms

- Discover algorithms for data analysis
- Implement data analysis algorithms
- Evaluating predictive performance

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.