

# Python for Data Analysis

## Duration

24h (4 days x 6h)

## Description

Python is the go-to language in many organizations and research institutes when dealing with data. This course aims to provide you with the right tools for the job and give you a new outlook on “think about data” in your data challenges.

## Audience

Everyone is looking for multiple ways to work with and visualize data: importing, extracting, joining, and manipulating data to fit their needs.

## Prerequisites

- Experience in Python programming or completion of our *Python Fundamentals* course.

## Objectives

After completing the course, the students should:

- know how to bring in data from multiple sources into python
- know how to organize that data before analysis
- know how to manipulate, clean, and enrich data
- know how to aggregate, combine and measure data
- know how to build data visualizations

## Course Outline

1. Welcome
  - a) Course Overview
  - b) Installation and Setup
2. New Data Ecosystem
  - a) The Big Data paradigm
  - b) Use Cases for Data Science
3. Revisiting Python fundamentals
4. Statistics in Python
5. Essential data science libraries
  - a) Pandas

- b) Numpy
- 6. Data visualization packages
- 7. Exploratory Data Analysis (EDA) packages
- 8. Data Engineering
  - a) ETL process (Extract, Transform, Load)
  - b) Extracting from multiple sources and file formats
    - i. Excel
    - ii. CSV
    - iii. REST API's (JSON)
    - iv. Web Scraping
    - v. FTP
    - vi. Google Drive
    - vii. Databases
  - c) SQL has the universal language for data engineering
  - d) Data lake
  - e) Unique data models
  - f) Data warehousing
- 9. Capstone Project: BYOD (Bring Your Own Data)
- 10. Goodbye
  - a) Course Recap
  - b) Course evaluation
  - c) Q&A