
Beginning Data Science with Python and Jupyter

Duration: 1 Day Course Code: LO035412 Delivery Method: Virtual Learning

Overview:

This fast-paced practical single-day course focuses on solving challenges presented by data science in a manner that is simple to conceptualize and easy to implement.

This is achieved by leveraging Python libraries that offer abstractions to complicated underlying algorithms. The result is that data science becomes very approachable for beginners, a fact which is reflected by the strength and growing popularity of the Python ecosystem. In this course, we will cover every aspect of the standard data workflow process, including collecting, cleaning, investigating, visualizing, and modeling data.

The Jupyter Notebook acts as an add-on tool - a virtual playground - that allows you to create and share live code, equations, visualizations, and text.

By touching on a variety of topics within the discipline, you'll be exposed to many interesting examples with real-world data.

Company Events

These events can be delivered exclusively for your company at our locations or yours, specifically for your delegates and your needs. The Company Events can be tailored or standard course deliveries.

Virtual Learning

This interactive training can be taken from any location, your office or home and is delivered by a trainer. This training does not have any delegates in the class with the instructor, since all delegates are virtually connected. Virtual delegates do not travel to this course, Global Knowledge will send you all the information needed before the start of the course and you can test the logins.

Target Audience:

If you're a Python programmer stepping out into the hugely popular world of data science, opting for this course is the right way to get started.

For the best experience in this course, you should have knowledge of programming fundamentals and some experience with Python. In particular, having some familiarity with the Python libraries Pandas, Matplotlib, and scikit-learn will be useful.

Objectives:

- This course focuses on creating reproducible data analyses using Python and Jupyter, and is intended for an audience with a background in Python. As such, we do not cover the basics of Python in this course. However, we will take a brief tour of the Jupyter interface.

Prerequisites:

Hardware

This course will require a computer system for the instructor and one for each student. The minimum hardware requirements are as follows:

- Processor: i5
- Memory: 8 GB RAM
- Hard disk: 10 GB

- An internet connection
- Software

For this course, we will use the following software:

- Anaconda 4.3+ and Python 3.5+
 - Python libraries included with Anaconda installation:
 - matplotlib 2.1.0+
 - ipython 6.1.0+
 - requests 2.18.4+
 - beautifulsoup4 4.6.0+
 - numpy 1.13.1+
 - pandas 0.20.3+
 - scikit-learn 0.19.0+
 - seaborn 0.8.0+
 - bokeh 0.12.10+
- Python libraries that require manual installation:

- mlxtend
- version_information
- ipython-sql
- pdir2
- graphviz
- Download and install all the required Python libraries

Content:

Lesson 1: Jupyter Fundamentals

- Basic Functionality and Features
- Our First Analysis - The Boston Housing Dataset

Lesson 2: Data Cleaning and Advanced Machine Learning

- Preparing to Train a Predictive Model
- Training Classification Models

Lesson 3: Web Scraping and Interactive Visualizations

- Scraping Web Page Data
 - Interactive Visualizations
-

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK