

Python Programming

Duration: 5 days; Instructor-led

WHAT YOU WILL LEARN

In the past decade, the demand for data has increased exponentially. The industry has begun to realise the potential goldmine of summarised information collected online. The various processes in data science are collect, collate and disseminate. The industry is also investigating on various applications that can streamline the valuable information for analytics processes and making the data collection simple and efficient. The industry is expected to be worth over \$128 billion by 2022, a predicted 36 per cent growth from 2016. With the Data Analytics Industry becoming dynamic, the prospects for someone looking to make Data Science as their career are high.

Although the amount of collected data is impressive, the data is useless without it is being analysed and insights leading transformation. Without enough manpower to work out on the information, it is pointless collecting the data in the first place. Businesses are also starting to react to the data scientist shortage and are collaborating with other firms and educational establishments to close the gap before it becomes too large. Through this course, we have focused on the practical challenges that organisations are experiencing by merging disciplines to develop a teaching programme that makes the link between business, management and data analytics.

This course includes the fundamental python programming techniques such as lambdas, reading and manipulating csv files, and the numpy library. The course also introduce data manipulation and cleaning techniques using the popular python pandas data science library and introduce the abstraction of the Series and DataFrame as the central data structures for data analysis, along with tutorials on how to use functions such as group by, merge, and pivot tables effectively. By the end of this course, participants will be able to take tabular data, clean it, manipulate it, and run basic inferential statistical analyses.

AUDIENCE

- Anyone who needs to learn how to write programs in Python
- Anyone who wants a solid exposure to Python as their first programming language
- Anyone who wants to apply statistics, machine learning, information visualization, social network analysis, and text analysis techniques to gain new insight into data.

PREREQUISITES

There are no prerequisites for this course but some programming background is preferred.

COURSE OBJECTIVES

- Master the fundamentals of writing Python scripts
- Learn core Python scripting elements such as variables and flow control structures
- Discover how to work with lists and sequence data
- Write Python functions to facilitate code reuse
- Use Python to read and write files
- Make their code robust by handling errors and exceptions properly
- Work with the Python standard library
- Explore Python's object-oriented features
- Search text using regular expressions

OUTLINES

Module 1: An Overview of Python

- What is Python?
- Interpreted languages
- Advantages and disadvantages
- Downloading and installing
- Which version of Python
- Where to find documentation
- Python Comments
- Output to the screen
- Running Python Scripts
- Structure of a Python script
- Using the interpreter interactively

Module 2: Getting Started

- Using variables
- Assigning value to multiple variables
- Expression
- Math operators
- String types: normal, raw and Unicode
- String operators
- Command line parameters
- Reading from the keyboard

Module 3: Decision & Flow Control

- About flow control
- Indenting is significant
- The if statements
- The nested if statements
- The elif statements
- The for loops
- The while loops
- Loop Controls - break and continue
- The range() function
- Arrays

Module 4: Defining Functions

- Syntax of function definition
- Formal parameters
- Global versus local variables
- Passing parameters and returning values
- Passing list of parameters
- Variable length arguments
- Lambda functions
- Passing function to another function
- Returning function
- Inner functions

Module 5: Working with Files

- Text file I/O overview
- Opening a text file
- Reading text files
- Raw (binary) data
- Writing to a text file
- Opening Excel File
- Reading from Excel File
- Writing data into Excel File

Module 6: Sequence

- List overview
- List methods
- Tuple overview
- Tuple methods
- Dictionary overview
- Dictionary methods
- Set overview
- Set methods

- Fetching values
- Fetching keys
- Testing for existence of elements
- Deleting elements
- Set Operators

Module 7: Python Classes

- About o-o programming
- Defining classes
- Class methods and data
- Constructors
- Objects
- Instance methods
- Instance data
- Destructors
- Interfaces
- Inheritances

Module 8: Errors and Exception Handling

- Dealing with syntax errors
- Exceptions
- Handling exceptions with try/except
- Cleaning up with finally

Module 9: Using Modules

- What is a module?
- The import statement
- Function aliases
- Packages
- Installing Packages from PYPI
- Standard Modules – sys
- Standard Modules – math
- Standard Modules – time

Module 10: Regular Expressions

- RE Objects and Pattern matching
- Parsing data
- Subexpressions
- Complex substitutions
- RE tips and tricks

Module 11: Highlights of the Standard Library

- Working with the operating system
- Grabbing web pages
- Sending email
- Using glob for filename wildcards
- math and random
- Accessing dates and times with datetime
- Working with compressed files

Module 12: Accessing Databases

- Selecting Data
- Inserting and Updating Data
- Deleting data
- Generic database API based on MySQL
- Using the Object Relational Mapper (SQLAlchemy)
- Working with NoSQL databases

Module 13: Data Distribution

- Center
- Spread
- Shape – Symmetry, Number of peaks, Skewness, Uniform
- Unusual Features – Gaps, Outliers
- Measures of central tendency - Mean, Median, Mode, Midrange
- Measures of spread - Range, Variation, Standard deviation, Interquartile range
- Measures of shape - Empirical rule, Chebyshev's rule, Skewness, Kurtosis
- Measures of relative position – Quartiles, Percentiles, Midquartile

Module 14: Python Data Distribution

- Introduction to Series
- Introduction to Pandas
- DataFrames
- Read From CSV
- Methods: head, shape, info, mean, median mode
- Histogram
- Methods: min, max, range, sqrt
- Methods: sorted, std, hist, correlation, heatmap
- Methods: skew, kurt, cov, quantile

Module 15: Extract Data from Website - BeautifulSoup

- Installing BeautifulSoup
- Installing a parser
- Making the soup
- Kinds of objects
- Navigating the tree
- Managing the tree
- Searching the tree
- Append the tree
- Insert inside the tree
- Extract, decompose, replace with,
- wrap and unwrap
- Pretty-printing
- Non-pretty printing
- Output formatters
- Get Text
- Output Encoding
- Unicode

Module 16: Selenium IDE

- Selenium Overview
- Selenium IDE Introduction
- Downloading and Installing Selenium IDE
- Recording and Running a Simple Test
- Selenium IDE – Features
- Installing Useful Tools for Writing Tests
- Selenium Concepts

Module 17: Selenium Webdriver

- Introduction to selenium webdriver
- Advantages of webdriver
- Downloading and configuring Webdriver
- Converting Selenium IDE test to programming language (Python)
- Detailed discussion about webdriver commands
- Handling different browsers
- Create our own methods in Webdriver
- Using RC commands from webdriver project

Module 18: Python for Data Analysis – NumPy

- Introduction
- Nddarray Object
- Data Types
- Array Attributes
- Array Creation Routines
- Array from existing data
- Numerical ranges
- Array Indexing and Slicing
- Advanced Indexing
- Iterating over Array
- Array Manipulation
- Arithmetic Operators
- Binary Operators
- String Functions
- Mathematical Functions
- Statistical Functions

Module 19: Python for Data Analysis – Pandas

- Introduction to Pandas
- Series
- DataFrames
- Missing Data
- Group By
- Merging Joining and Concatenating
- Operations
- Data Input and Output

Module 20: Python for Data Visualization

- Matplotlib
- Seaborn
- Distribution Plots
- Categorical Plots
- Matrix Plots
- Grids
- Regression Plots
- Pandas Built-in Data Visualization
- Plotly
- Cufflinks
- Geographical Plotting
- Choropleth Maps

Module 21: Python for Data Analysis – SciPy

- Introduction
- Basic functions
- Special functions
- Integration
- Optimization
- Interpolation
- Fourier transforms
- Signal Processing
- Linear Algebra
- Sparse Eigenvalue Problems with ARPACK
- Compressed Sparse Graph Routines
- Spatial data structures and algorithms
- Statistics
- Multidimensional image processing
- File IO