

**Python for Data Analysis**

**Course Number:** PYTH-130
**Duration:** 3 days

**Overview**

Accelebrate's Python for Data Analysis training course teaches data analysts how to search, manipulate, and analyze data using the powerful Python programming language.

**Prerequisites**

All attendees should have basic Python programming skills.

**Materials**

All Python training attendees receive comprehensive courseware.

**Software Needed on Each Student PC**

* Any Windows, Linux, or macOS operating system
* Anaconda Python 3.5 or later
* A text editor or IDE (PyCharm Community Edition recommended)

**Objectives**

* Extract data from binary files or other binary data streams
* Create data structures using classes and named tuples
* Search and replace text with regular expressions
* Read and write CSV and other data formats
* Serialize data to pickle files, JSON, and XML
* Consume and process data from the Web
* Deal with missing data
* Share data with Excel spreadsheets
* Analyze data with SciPy/NumPy

**Outline**

* Introduction
* File I/O
	+ Opening a file
	+ Iterating over lines
	+ Reading characters or bytes
	+ Reading all lines
	+ Formatted output
	+ Using fileinput
* Classes
	+ Defining classes
	+ Constructors
	+ Instance methods and data
	+ Class/static methods and data
* Generators and Other Iterables
	+ Iterables
	+ Saving memory with generators
	+ Generator expressions
	+ Generator functions
	+ Generator classes
	+ Stacking generators
* Data Structures
	+ How to store data
	+ The basics: lists and tuples
	+ Named access with dictionaries
	+ Named tuples: best of both worlds
	+ Using classes as data structures
* Serializing Data
	+ Pickle
	+ JSON
	+ CSV
	+ XML
* Consuming Data from the Web
	+ Web data sources
	+ Data via URL
	+ RESTful data
	+ Screen-scraping
* Excel Spreadsheets
	+ The xlrd, xlwr, and xlutil modules
	+ Reading an existing spreadsheet
	+ Creating a spreadsheet from scratch
	+ Modifying an existing spreadsheet
* Dates and Times
	+ Python date and time objects
	+ The time module
	+ Using calendars
	+ Converting between formats
	+ Parsing and printing
	+ Time zones
* Regular Expressions
	+ RE syntax overview
	+ Basic patterns
	+ RE objects
	+ Searching and matching
	+ Compilation flags
	+ Grouping
	+ Replacing text
	+ Splitting a string
* Working with Binary Data
	+ Isn't all data binary?
	+ Binary file handling
	+ Parsing raw data
	+ Writing a binary stream
* Analyzing Datasets
	+ Sorting data
	+ Filtering values
	+ Basic statistics
	+ Leveraging SciPy/NumPy
	+ Using pandas
* Bigger Data - Working with PyTables
	+ About HDF5 data
	+ Using PyTables
	+ Reading a dataset
	+ Pulling data
	+ Updating the dataset
	+ Writing to HDF5
* Conclusion