

## DATA SCIENCE WITH PYTHON

Course Duration: 3 days; Instructor-led

### WHAT YOU WILL LEARN

In the past decade, the demand for data has increased exponentially. The industry has begun to realize the potential goldmine of summarized 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 analyzed 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 organizations are experiencing by merging disciplines to develop a teaching programmed 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 introduces data manipulation and cleaning techniques using the popular python pandas data science library and introduce the abstraction of the Series and Data Frame 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

This course "Data Science with Python" is intended for learners who have basic python knowledge and 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 python knowledge with a little programming background is preferred.

### METHODOLOGY

This program will be conducted with interactive lectures, PowerPoint presentation, discussion and practical exercise.

### COURSE OBJECTIVES

After completing this course, you should be able to:

- Explore Python fundamentals, including basic syntax, variables, and types.
- Create and manipulate regular Python lists.
- Use functions and import packages.
- Build Numpy arrays and perform interesting calculations.
- Create and customize plots on real data.
- Supercharge with control flow, and get to know the Pandas DataFrame
- Use Python to read and write files.
- Illustrate Supervised Learning Algorithms
- Identify and recognize machine learning algorithms around us

### COURSE OUTLINES

#### Module 1: Python Crash Course

- Introduction to the Course
- Environment Set-Up
- Virtual Environments
- Data types and Operators
- Integers, Floats, Strings, Bytes, Tuples and Lists
- Dictionaries and Ordered Dictionaries
- Sets and frozen sets.
- Flow control - if, elif statements
- Flow control - while loops
- Creating and using functions
- Creating modules and packages

- Distributing code to repositories

### Module 2: Python Object Oriented

- Creating Classes
- Creating Objects and Instances
- Data Encapsulation
- Class Inheritance
- Multiple Inheritance
- Decorators

### Module 3: Error Handling and Testing

- Handling Exception
- Raising exceptions
- Writing tests cases
- Executing tests
- Checking code coverage by tests

### Module 4: Working with Files and Directories

- Accessing different types of files
- File handling principles
- Creating and reading Files
- Updating Files
- Deleting files
- Text Files
- CSV Files
- Microsoft Word
- Microsoft Excel
- Regular Expressions
- Extracting data from text files using Regular Expressions
- Creating and deleting directories
- Listing and searching for files

### Module 5: Accessing RDBMS Databases

- Introduction to RDBMS Databases and Concepts
- Types of SQL Statements (DDL vs DML)
- How to Access Databases Using Python
- CREATE TABLE
- ALTER, DROP and TRUNCATE
- SELECT statement
- COUNT, DISTINCT, LIMIT
- INSERT statement
- UPDATE AND DELETE statements
- Using String Patterns and Ranges
- Sorting Result Sets
- Grouping Result Sets
- Built-in Database Functions
- Date and Time Built-in Functions
- Sub-Queries and Nested Selects

- Working with Multiple Tables

### Module 6: Accessing RDBMS Databases using SQLAlchemy

- Object Relational Mapper (SQLAlchemy)
- Introduction and Architecture
- Introduction to SQLAlchemy ORM
- Database Models
- Relationships
- Queries
- Inserting Data
- Updating Data
- Deleting Data

### Module 7: Accessing NOSQL Databases

- Introduction
- Differences between MongoDB and MySQL
- Setting up MongoDB
- Overview of MongoDB Features and Architecture
- Mapping between a relational database and MongoDB
- Connecting to MongoDB
- Starting a Python + MongoDB application
- Understanding the MongoDB Data Processing Pipeline
- Reading and Writing to the database
- Creating a New Database
- Understanding Availability in MongoDB
- Summary and Conclusion

### Module 8: Website Crawling and Scraping

- Introduction
- Setting up the Development Environment
- Python Primer: Data Structures, Conditionals, File Handling, etc.
- Python Packages for Web Scraping: Scrapy and BeautifulSoup
- How a Website Works
- How HTML is Structured
- Making a Web Request
- Scraping an HTML Page
- Working with XPath and CSS
- Filtering Data Using Regular Expressions
- Creating a Web Crawler
- Crawling AJAX and JavaScript Pages with Selenium.
- Web Scraping Best Practices
- Troubleshooting
- Summary and Conclusion



### Module 9: 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

- Matrix Plots
- Grids
- Regression Plots
- Pandas Built-in Data Visualization
- Plotly and Cufflinks
- Geographical Plotting
- Choropleth Maps

### Module 10: 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

### Module 11: Python for Data Analysis - Pandas

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

### Module 12: Python for Data Visualization

- Matplotlib
- Seaborn
- Distribution Plots
- Categorical Plots