

# Python for Data Analytics (3 Days)

## Overview

The Python for Data Analytics course is intended to develop data analysts capable of driving data-driven decisions in any organization. The course begins with a refresher on the Python programming language and builds on this knowledge to introduce students to the world of data analytics in Python. Students will learn to use advanced data structures in Python, access data from external sources, use descriptive analytics, perform statistical analysis, and visualize their data in meaningful ways.

## Audience

This course is appropriate for individuals and organizations seeking to learn the basics of data analytics in Python. Students are expected to begin the course with at least some programming knowledge. Pre-work can be provided to students with no programming experience.

## Prerequisites

Little to some programming knowledge in any programming language, preferably Python.

## Course Objectives

Upon completion of this course, participants should:

- Write Python scripts that take in user input and output results
- Use Anaconda to simplify package management
- Create and use advanced data structures like vectors, matrices, and data frames
- Access data from external sources and manipulate it with Python data structures
- Use descriptive analytics and statistical analysis techniques to derive useful insight from data
- Visualize and format data in common charts using matplotlib
- Drive data-driven decisions in your organization

## Course Outline

### Part 1: Python Environment, Conventions, and Fundamentals Refresher

- What is Python?
- Introduction to Anaconda
- Writing Python Scripts
- Basic Data Types
- Functions
- Flow Control
- Python conventions
- Python best practices

## Part 2: Introduction to Data Analytics

- The Data Analysis Process
- Descriptive Data Analytics
- Diagnostic Data Analytics
- Predictive Data Analytics
- Prescriptive Data Analytics

## Part 3: Advanced Data Structures

- What are Data Structures?
  - Why Do We Need Different Kinds of Data Structures?
- Lists
- Arrays / Vectors
  - Matrices
- Factors (Categorical Data in pandas)
- pandas DataFrame
- CSV Files

## Part 4: Accessing Data

- CSV Files in Python
  - Reading CSV Files
  - Parsing CSV Files
  - Loading Data into Python Data Structures
  - Outputting CSV Files
- Connecting to Databases in Python
  - Reading Database Tables
  - Accessing Data in Databases
  - Loading Data into Python Data Structures
  - Creating Data Tables
  - Updating Data in Databases

## Part 5: Data Wrangling

- Filter Out Bad or Irrelevant Data
- Derive Composite Fields
- Aggregating Data from Multiple Sources
- Standardization
- Normalization

## **Part 6: Introduction to Exploratory Data Analytics**

- Explore the Data
- Answer the Question, “What Happened?”
- Identify Trends and Patterns
  - Summary Statistics
  - Clustering
  - Pattern Tracking
  - Regression Analysis
- Generate Primitive Reports

## **Part 7: Introduction to Statistical Analytics**

- Statistical Analysis Basics
  - Mean
  - Median
  - Mode
  - Range
- Standard Deviation
- Linear Regression
- Hypothesis Testing
- Sample Size Determination
- How Do We Apply This Data to Business Decisions?

## **Part 8: Visualizing Data**

- Charts
  - Bar Charts
  - Box Plots
  - Histograms
  - Line Plots
  - Pie Charts
- Generate Business Reports
- Tableau, Power BI, and Other Tools