Data Science with Python Certification Training
Course Agenda

Lesson 1: Data Science Overview

- Data Science
- Data Scientists
- Examples of Data Science
- Python for Data Science

Lesson 2: Data Analytics Overview

- Introduction to Data Visualization
- Processes in Data Science
- Data Wrangling, Data Exploration, and Model Selection
- Exploratory Data Analysis or EDA
- Data Visualization
- Plotting
- Hypothesis Building and Testing

Lesson 3: Statistical Analysis and Business Applications

- Introduction to Statistics
- Statistical and Non-Statistical Analysis
- Some Common Terms Used in Statistics
• Data Distribution: Central Tendency, Percentiles, Dispersion

• Histogram

• Bell Curve

• Hypothesis Testing

• Chi-Square Test

• Correlation Matrix

• Inferential Statistics

Lesson 4: Python: Environment Setup and Essentials

• Introduction to Anaconda

• Installation of Anaconda Python Distribution - For Windows, Mac OS, and Linux

• Jupyter Notebook Installation

• Jupyter Notebook Introduction

• Variable Assignment

• Basic Data Types: Integer, Float, String, None, and Boolean; Typecasting

• Creating, accessing, and slicing tuples

• Creating, accessing, and slicing lists

• Creating, viewing, accessing, and modifying dicts

• Creating and using operations on sets

• Basic Operators: 'in', '+', '*'

• Functions
Lesson 5: Mathematical Computing with Python (NumPy)

- NumPy Overview
- Properties, Purpose, and Types of ndarray
- Class and Attributes of ndarray Object
- Basic Operations: Concept and Examples
- Accessing Array Elements: Indexing, Slicing, Iteration, Indexing with Boolean Arrays
- Copy and Views
- Universal Functions (ufunc)
- Shape Manipulation
- Broadcasting
- Linear Algebra

Lesson 6: Scientific computing with Python (Scipy)

- SciPy and its Characteristics
- SciPy sub-packages
- SciPy sub-packages – Integration
- SciPy sub-packages – Optimize
- Linear Algebra
- SciPy sub-packages – Statistics
- SciPy sub-packages – Weave
Lesson 7: Data Manipulation with Python (Pandas)

- Introduction to Pandas
- Data Structures
- Series
- DataFrame
- Missing Values
- Data Operations
- Data Standardization
- Pandas File Read and Write Support
- SQL Operation

Lesson 8: Machine Learning with Python (Scikit-Learn)

- Introduction to Machine Learning
- Machine Learning Approach
- How Supervised and Unsupervised Learning Models Work
- Scikit-Learn
- Supervised Learning Models - Linear Regression
- Supervised Learning Models: Logistic Regression
- K Nearest Neighbors (K-NN) Model
- Unsupervised Learning Models: Clustering
• Unsupervised Learning Models: Dimensionality Reduction
• Pipeline
• Model Persistence
• Model Evaluation - Metric Functions

Lesson 9: Natural Language Processing with Scikit-Learn

• NLP Overview
• NLP Approach for Text Data
• NLP Environment Setup
• NLP Sentence analysis
• NLP Applications
• Major NLP Libraries
• Scikit-Learn Approach
• Scikit - Learn Approach Built - in Modules
• Scikit - Learn Approach Feature Extraction
• Bag of Words
• Extraction Considerations
• Scikit - Learn Approach Model Training
• Scikit - Learn Grid Search and Multiple Parameters
• Pipeline
Lesson 10: Data Visualization in Python using Matplotlib

- Introduction to Data Visualization
- Python Libraries
- Plots
- Matplotlib Features:
  - Line Properties Plot with (x, y)
  - Controlling Line Patterns and Colors
  - Set Axis, Labels, and Legend Properties
  - Alpha and Annotation
  - Multiple Plots
  - Subplots
- Types of Plots and Seaborn

Lesson 11: Data Science with Python Web Scraping

- Web Scraping
- Common Data/Page Formats on The Web
- The Parser
- Importance of Objects
- Understanding the Tree
- Searching the Tree
- Navigating options
• Modifying the Tree
• Parsing Only Part of the Document
• Printing and Formatting
• Encoding

Lesson 12: Python integration with Hadoop, MapReduce and Spark

• Need for Integrating Python with Hadoop
• Big Data Hadoop Architecture
• MapReduce
• Cloudera QuickStart VM Set Up
• Apache Spark
• Resilient Distributed Systems (RDD)
• PySpark
• Spark Tools
• PySpark Integration with Jupyter Notebook

For information on the course, visit: http://www.simplilearn.com/big-data-and-analytics/python-for-data-science-training