

Data Science with Python

Course Contents

Lesson 00 - Course Overview

- Course Overview

Lesson 01 - Data Science Overview

- Introduction to Data Science
- Different Sectors Using Data Science
- Purpose and Components of Python

Lesson 02 - Data Analytics Overview

- Data Analytics Process
- Knowledge Check
- Exploratory Data Analysis (EDA)
- EDA-Quantitative Technique
- EDA - Graphical Technique
- Data Analytics Conclusion or Predictions
- Data Analytics Communication
- Data Types for Plotting
- Data Types and Plotting

Lesson 03 - Statistical Analysis and Business Applications

- Introduction to Statistics
- Statistical and Non-statistical Analysis
- Major Categories of Statistics
- Statistical Analysis Considerations
- Population and Sample
- Statistical Analysis Process
- Data Distribution
- Dispersion
- Histogram
- Testing
- Correlation and Inferential Statistics

Lesson 04 - Python Environment Setup and Essentials

- Anaconda
- Installation of Anaconda Python Distribution (contd.)
- Data Types with Python
- Basic Operators and Functions

Lesson 05 - Mathematical Computing with Python (NumPy)

- Introduction to Numpy
- Activity-Sequence it Right
- Demo 01-Creating and Printing an ndarray
- Knowledge Check
- Class and Attributes of ndarray
- Basic Operations
- Activity-Slice It
- Copy and Views
- Mathematical Functions of Numpy

Lesson 06 - Scientific computing with Python (SciPy)

- Introduction to SciPy
- SciPy Sub Package - Integration and Optimization
- Knowledge Check
- SciPy sub package
- Demo - Calculate Eigenvalues and Eigenvector
- Knowledge Check
- SciPy Sub Package - Statistics, Weave and IO

Lesson 07 - Data Manipulation with Pandas

- Introduction to Pandas
- Knowledge Check
- Understanding DataFrame
- View and Select Data Demo
- Missing Values
- Data Operations
- Knowledge Check
- File Read and Write Support
- Knowledge Check-Sequence it Right
- Pandas Sql Operation

Lesson 08 - Machine Learning with Scikit-Learn

- Machine Learning Approach
- Understand data sets and extract its features
- Identifying problem type and learning model
- How it Works
- Train, test and optimizing the model
- Supervised Learning Model Considerations
- Knowledge Check
- Scikit-Learn
- Knowledge Check
- Supervised Learning Models - Linear Regression

- Supervised Learning Models - Logistic Regression
- Unsupervised Learning Models
- Pipeline
- Model Persistence and Evaluation

Lesson 09 - Natural Language Processing with Scikit Lear

- NLP Overview
- NLP Applications
- Knowledge Check
- NLP Libraries-Scikit
- Extraction Considerations
- Scikit Learn-Model Training and Grid Search

Lesson 10 - Data Visualization in Python using matplotlib

- Introduction to Data Visualization
- Knowledge Check
- Line Properties
- (x,y) Plot and Subplots
- Knowledge Check
- Types of Plots

Lesson 11 - Web Scraping with BeautifulSoup

- Web Scraping and Parsing
- Knowledge Check
- Understanding and Searching the Tree
- Navigating options
- Demo 3 Navigating a Tree
- Knowledge Check
- Modifying the Tree
- Parsing and Printing the Document

Lesson 12 - Python integration with Hadoop MapReduce and Spark

- Why Big Data Solutions are Provided for Python
- Hadoop Core Components
- Python Integration with HDFS using Hadoop Streaming
- Demo 01 - Using Hadoop Streaming for Calculating Word Count
- Knowledge Check
- Python Integration with Spark using PySpark
- Demo 02 - Using PySpark to Determine Word Count