simplilearn

Data Science with Python

This Data Science with Python course will establish your mastery of data science and analytics techniques using Python. With this Python for Data Science Course, you'll learn the essential concepts of Python programming and gain deep knowledge in data analytics, machine learning, data visualization, web scraping and natural language processing. Python is a required skill for many data science positions, so jumpstart your career with this interactive, hands-on course.

About the Program

The Data Science with Python course will furnish you with in-depth knowledge of the various libraries and packages required to perform data analysis, data visualization, web scraping, machine learning and natural language processing using Python.

Python has surpassed Java as the top language used to introduce US students to programming and computer science, and 46 percent of data science jobs list Python as a required skill.

Who Can Enroll For This Program?

There is a booming demand for skilled data scientists across all industries that make this course suited for participants at all levels of experience. We recommend this Data Science with Python training particularly for the following professionals:

IT professionals

Analytics professionals

Software Developers

Program features:

68
hours of in-depth learning

4
real-life industry-based projects

Lifetime access
to self-paced learning

Dedicated mentoring session
from our faculty of industry experts

Data Science with Python Outcomes:

This Python for Data Science training course will enable you to:

1. Gain an in-depth understanding of data science processes, data wrangling, data exploration, data visualization, hypothesis building, and testing. You will also learn the basics of statistics
2. Install the required Python environment and other auxiliary tools and libraries
3. Understand the essential concepts of Python programming such as data types, tuples, lists, dicts, basic operators and functions
4. Perform high-level mathematical computing using the NumPy package and its large library of mathematical functions
5. Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO and Weave
6. Perform data analysis and manipulation using data structures and tools provided in the Pandas package
7. Gain expertise in machine learning using the Scikit-Learn package
8. Gain an in-depth understanding of supervised learning and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline
9. Use the Scikit-Learn package for natural language processing
10. Use the matplotlib library of Python for data visualization
11. Extract useful data from websites by performing web scrapping using Python
12. Integrate Python with Hadoop, Spark and MapReduce

Chapter level details:

Lesson 1: Data Science Overview

Lesson Objective: This introductory lesson gives an overview of data science and where it is being used. It also explains the components and purpose of Python.

Topics:

1. Introduction to Data Science
2. Different Sectors Using Data Science
3. Purpose and Components of Python
4. Quiz
5. Key Takeaways

Lesson 2: Data Analytics Overview

Lesson Objective: This lesson on Data Analytics overview explains the data analytics process in detail and also covers data types for plotting. Exploratory Data Analysis(EDA) techniques are also covered.

Topics:

1. Data Analytics Process
2. Knowledge Check
3. Exploratory Data Analysis(EDA)
4. EDA-Quantitative Technique
5. EDA - Graphical Technique
6. Data Analytics Conclusion or Predictions
7. Data Analytics Communication
8. Data Types for Plotting
9. Data Types and Plotting
10. Knowledge Check
11. Quiz
12. Key Takeaways

Lesson 3: Statistical Analysis and Business Applications

Lesson Objective: This lesson introduces you to statistics and statistical analysis process, data distribution, dispersion, histogram, and testing.

Topics:

1. Introduction to Statistics
2. Statistical and Non-statistical Analysis
3. Major Categories of Statistics
4. Statistical Analysis Considerations
5. Population and Sample
6. Statistical Analysis Process
7. Data Distribution
8. Dispersion
9. Knowledge Check
10. Histogram
11. Knowledge Check
12. Testing
13. Knowledge Check
14. Correlation and Inferential Statistics
15. Quiz
16. Key Takeaways

Lesson 4: Python Environment Setup and Essentials

Lesson Objective: This lesson will teach you how to install Anaconda, what are the different data types, operators, and functions in Python.

Topics:

1. Anaconda
2. Installation of Anaconda Python Distribution (contd.)
3. Data Types with Python
4. Basic Operators and Functions
5. Quiz
6. Key Takeaways

Lesson 5: Mathematical Computing with Python (NumPy)

Lesson Objective: This lesson begins with an introduction to Numpy, it then progresses to ndarray and mathematical functions of Numpy.

Topics:

1. Introduction to Numpy
2. Activity-Sequence it Right
3. Demo 01-Creating and Printing an ndarray
4. Knowledge Check
5. Class and Attributes of ndarray
6. Basic Operations
7. Activity-Slice It
8. Copy and Views
9. Mathematical Functions of Numpy
10. Assignment 01: Evaluate the datasets containing GDPs of different countries
11. Demo: Assignment 01
12. Assignment 02: Evaluate the datasets of Summer Olympics, 2012
13. Demo: Assignment 02
14. Quiz
15. Key Takeaways

Lesson 06 - Scientific computing with Python (Scipy)

Lesson Objective: This lesson will give you a detailed overview of SciPy and its sub packages.

Topics:

1. Introduction to SciPy
2. SciPy Sub Package - Integration and Optimization
3. Knowledge Check
4. SciPy sub package
5. Demo - Calculate Eigenvalues and Eigenvector
6. Knowledge Check
7. SciPy Sub Package - Statistics, Weave and IO
8. Assignment 01: Use SciPy to solve a linear algebra problem
9. Demo: Assignment 01
10. Assignment 02: Use SciPy to define 20 random variables for random values
11. Demo: Assignment 02
12. Quiz
13. Key Takeaways

Lesson 07 - Data Manipulation with Pandas

Lesson Objective: You will learn about Data Manipulation with Pandas in this lesson. Data frames, data demos, data operations, read and write supports, sql operation are all covered.

Topics:

1. Introduction to Pandas
2. Knowledge Check
3. Understanding DataFrame
4. View and Select Data Demo
5. Missing Values
6. Data Operations
7. Knowledge Check
8. File Read and Write Support
9. Knowledge Check-Sequence it Right
10. Pandas Sql Operation
11. Assignment 01: Analyze the Federal Aviation Authority(FAA) dataset using Pandas
12. Demo: Assignment 01
13. Assignment 02: Analyze the dataset in csv format given for fire department
14. Demo: Assignment 02
15. Quiz
16. Key Takeaways

Lesson 08 - Machine Learning with Scikit–Learn

Lesson Objective: This lesson covers the Machine Learning approach and how it works, supervised and unsupervised learning models.

Topics:

1. Machine Learning Approach
2. Steps 1 and 2: Understand the dataset and extract its features
3. Steps 3 and 4: Identify the problem type and learning model
4. How it Works
5. Steps 5 and 6: Train, test, and optimize the models
6. Supervised Learning Model Considerations
7. Knowledge Check
8. Scikit-Learn
9. Knowledge Check
10. Supervised Learning Models - Linear Regression
11. Supervised Learning Models - Logistic Regression
12. Unsupervised Learning Models
13. Pipeline
14. Model Persistence and Evaluation
15. Knowledge Check
16. Assignment 01: Evaluate a dataset to find the features or media channels used by a firm and sales figures for each channel
17. Demo: Assignment 01
18. Assignment 02: Analyze a dataset to find the features and response label of it
19. Demo: Assignment 02
20. Quiz
21. Key Takeaways

Lesson 09 - Natural Language Processing with Scikit Learn

Lesson Objective: Natural Language Processing is covered in this lesson. Overview, Applications, Libraries, Scikit Learn-Model Training are covered in detail.

Topics:

1. NLP Overview
2. NLP Applications
3. Knowledge check
4. NLP Libraries-Scikit
5. Extraction Considerations
6. Scikit Learn-Model Training and Grid Search
7. Assignment 01: Analyze a given spam collection dataset
8. Demo: Assignment 01
9. Assignment 02: Analyze the sentiment dataset using NLP
10. Demo: Assignment 02
11. Quiz
12. Key Takeaway

Lesson 10 - Data Visualization in Python using matplotlib

Lesson Objective: This lesson teaches you to visualize data in python using matplotlib and plot them.

Topics:

1. Introduction to Data Visualization
2. Knowledge Check
3. Line Properties
4. (x,y) Plot and Subplots
5. Knowledge Check
6. Types of Plots
7. Assignment 01: Analyze the “auto mpg data” and draw a pairplot using seaborn library for mpg, weight, and origin
8. Demo : Assignment 01
9. Assignment 02: Draw a pie chart to visualize a dataset
10. Demo: Assignment 02
11. Quiz
12. Key Takeaways

Lesson 11 - Web Scraping with BeautifulSoup

Lesson Objective: Web Scraping and Parsing, how to search, navigate, modify a tree are part of this lesson.

Topics:

1. Web Scraping and Parsing
2. Knowledge Check
3. Understanding and Searching the Tree
4. Navigating options
5. Demo3 Navigating a Tree
6. Knowledge Check
7. Modifying the Tree
8. Parsing and Printing the Document
9. Assignment 01: Scrape the Simplilearn website page to perform some tasks
10. Demo: Assignment 01
11. Assignment 02: Scrape the Simplilearn website page to perform some tasks
12. Demo: Assignment 02
13. Quiz
14. Key takeaways

Lesson 12 - Python integration with Hadoop MapReduce and Spark

Lesson Objective:

Topics:

1. Why Big Data Solutions are Provided for Python
2. Hadoop Core Components
3. Python Integration with HDFS using Hadoop Streaming
4. Demo 01 - Using Hadoop Streaming for Calculating Word Count
5. Knowledge Check
6. Python Integration with Spark using PySpark
7. Demo 02 - Using PySpark to Determine Word Count
8. Knowledge Check
9. Assignment 01: Determine the word count for Amazon dataset
10. Demo: Assignment 01
11. Assignment 02: Count and display all airports present in New York using PySpark
12. Demo : Assignment 02
13. Quiz
14. Key takeaways

Projects Covered:

The course includes four real-world, industry-based projects. Successful evaluation of one of the following projects is a part of the certification eligibility criteria:

Project 1: Products rating prediction for Amazon

Amazon, one of the leading US-based e-commerce companies, recommends products within the same category to customers based on their activity and reviews on other similar products. Amazon would like to improve this recommendation engine by predicting ratings for the non-rated products and add them to recommendations accordingly.

Domain: E-commerce

Project 2: Demand Forecasting for Walmart

Predict accurate sales for 45 stores of Walmart, one of the US-based leading retail stores, considering the impact of promotional markdown events. Check if macroeconomic factors like CPI, unemployment rate, etc. have an impact on sales.

Domain: Retail

Project 3: Improving customer experience for Comcast

Comcast, one of the US-based global telecommunication companies wants to improve customer experience by identifying and acting on problem areas that lower customer satisfaction if any. The company is also looking for key recommendations that can be implemented to deliver the best customer experience.

Domain: Telecom

Project 4: Attrition Analysis for IBM

IBM, one of the leading US-based IT companies, would like to identify the factors that influence attrition of employees. Based on the parameters identified, the company would also like to build a logistics regression model that can help predict if an employee will churn or not.

Domain: Workforce Analytics

Project 5: NYC 311 Service Request Analysis

Perform a service request data analysis of New York City 311 calls. You will focus on data wrangling techniques to understand patterns in the data and visualize the major complaint types.

Domain: Telecommunication

Project 6: MovieLens Dataset Analysis

The GroupLens Research Project is a research group in the Department of Computer Science and Engineering at the University of Minnesota. The researchers of this group are involved in several research projects in the fields of information filtering, collaborative filtering and recommender systems. Here, we ask you to perform an analysis using the Exploratory Data Analysis technique for user datasets.

Domain: Engineering

Project 7: Stock Market Data Analysis

As a part of this project, you will import data using Yahoo data reader from the following companies: Yahoo, Apple, Amazon, Microsoft and Google. You will perform fundamental analytics, including plotting, closing price, plotting stock trade by volume, performing daily return analysis, and using pair plot to show the correlation between all of the stocks.

Domain: Stock Market

Project 8: Titanic Dataset Analysis

On April 15, 1912, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This tragedy shocked the world and led to better safety regulations for ships. Here, we ask you to perform an analysis using the exploratory data analysis technique, in particular applying machine learning tools to predict which passengers survived the tragedy.

Domain: Hazard