

Course Name : CERTIFIED BIG DATA PROFESSIONAL
Duration : 4 days
Level : Beginner
Mode : Physical Classroom or Online Live Instructor

COURSE DESCRIPTION:

In this project-based Big Data course, our certified trainer will teach the programming skills and refine the candidate's ability to work with messy, complex big datasets. The candidates will learn to manipulate and prepare data for analysis and create visualizations for data exploration. Finally, the candidates shall learn to use your data skills to tell a story with data.

WHAT WILL YOU LEARN?

The first *Python Basics for Data Science* course provides a beginner-friendly introduction to Python for Big data and Data Science. Practice through lab exercises, and the candidates will be ready to create your first Python scripts on your own!

The second *Analyzing Data with Python* course, the candidates will learn how to analyze data in Python using multi-dimensional arrays in Numpy, manipulate Data Frames in pandas, use SciPy library of mathematical routines, and perform machine learning using scikit-learn!

The third *Visualization Data with Python* course covers the graphical representation of data to convey insights interactively and efficiently to clients, customers, and stakeholders in general.

PREREQUISITE:

Beginners. No programming knowledge is required.

METHODOLGY:

This program will be conducted with interactive lectures, PowerPoint presentation, discussions, and practical exercise. This course can be conducted as instructor-led (ILT) or virtual instructor-led training (VILT).

JOB SCOPE:

After completing this course, the candidates can join for the following job roles,

- Big Data Engineer
- Big Data Analyst
- Data Engineer
- Data Solutions Engineer
- Data Analyst
- Data Migration Consultant

COURSE I: PYTHON BASICS FOR DATA SCIENCE

MODULE 1: PYTHON BASICS

- Your first program
- Types
- Expressions and Variables
- String Operations

MODULE 2: PYTHON DATA STRUCTURES

- Lists and Tuples
- Sets
- Dictionaries

MODULE 3: PYTHON PROGRAMMING FUNDAMENTALS

- Conditions and Branching
- Loops
- Functions
- Objects and Classes

MODULE 4: WORKING WITH DATA IN PYTHON

- Reading files with open
- Writing files with open
- Loading data with Pandas
- Working with and Saving data with Pandas

MODULE 5: WORKING WITH NUMPY ARRAYS

- Numpy 1d Arrays
- Numpy 2d Arrays

COURSE II: ANALYZING DATA WITH PYTHON

MODULE 1: IMPORTING DATASETS

- Learning Objectives
- Understanding the Domain
- Understanding the Dataset
- Python package for data science
- Importing and Exporting Data in Python
- Basic Insights from Datasets

MODULE 2: CLEANING AND PREPARING THE DATA

- Identify and Handle Missing Values
- Data Formatting
- Data Normalization Sets
- Binning
- Indicator variables

MODULE 3 - SUMMARIZING THE DATA FRAME

- Descriptive Statistics
- Basic of Grouping
- ANOVA
- Correlation
- More on Correlation

MODULE 4 - MODEL DEVELOPMENT

- Simple and Multiple Linear Regression
- Model Evaluation Using Visualization
- Polynomial Regression and Pipelines
- R-squared and MSE for In-Sample Evaluation
- Prediction and Decision Making

MODULE 5 - MODEL EVALUATION

- Model Evaluation
- Over-fitting, Under-fitting, and Model Selection
- Ridge Regression
- Grid Search
- Model Refinement

COURSE III: VISUALIZING DATA WITH PYTHON

MODULE 1: INTRODUCTION TO VISUALIZATION TOOLS

- Introduction to Data Visualization
- Introduction to Matplotlib
- Basic Plotting with Matplotlib
- Dataset on Immigration to Canada
- Line Plots

MODULE 2: BASIC VISUALIZATION TOOLS

- Area Plots
- Histograms
- Bar Charts

MODULE 3: SPECIALIZED VISUALIZATION TOOLS

- Pie Charts
- Box Plots
- Scatter Plots
- Bubble Plots

MODULE 4: ADVANCED VISUALIZATION TOOLS

- Waffle Charts
- Word Clouds
- Seaborn and Regression Plots

MODULE 5: CREATING MAPS AND VISUALIZING GEOSPATIAL DATA

- Introduction to Folium
- Maps with Markers
- Choropleth Maps

CONCLUSION

- QA
- Useful References and Books
- Feedback