**Python for Data Science**

**Title: Python for Data Science**

**Duration:** 1 day

**Pre-requisite:** Basic Python Programming

**Short Description:**

Python is the language of data science and this Python for Data Science class will expose you to the most important libraries (i.e., NumPy, Pandas, Matplotlib, and Scikit-learn) that will enable you to effectively do data science using Python.

**Long Description:**

Python is the language of data science and this Python for Data Science course will expose you to the most important libraries (i.e., NumPy, Pandas, Matplotlib and Scikit-learn) that will enable you to effectively do data science using Python.

We will begin by installing Anaconda, which provides the libraries required for most data problems. We will discuss the focus and strengths of the most important libraries and how they enable data analysis and the application of machine learning to defined data problems. We will then use these libraries to perform data exploration, visualization, analysis and modeling on a variety of datasets as we work through the data science process.

**Learning Objectives:**

After this course, you will be able to:

* Install Anaconda on a personal computer.
* Understand the various options for performing data science.
* Understand the reasons for Python’s popularity in data science.
* Learn the primary libraries for data science in Python including NumPy, Pandas, Matplotlib and Scikit-learn.
* Perform exploratory data analysis using Pandas.
* Use Matplotlib and Seaborn to perform data visualization.
* Prepare data for machine learning
* Apply machine learning on a variety of datasets
* Understand the data science process
* Understand the big picture and the importance of data science in business, industry, and technology.

**Topic Outline:**

* Course Introduction
* Overview of data science
* Understand the reasons for Python’s popularity in data science
* Installing Anaconda
* Milestone 1: Learn how to use Jupyter Notebooks
* The data science process
* Essential Python data science libraries
  + NumPy
  + Pandas
  + Matplotlib
  + Scikit-learn
* Data Visualization
* Line Chart
* Scatterplot
* Pairplot
* Histogram
* Density Plot
* Bar Chart
* Boxplot
* Customizing Charts
* Prepare data for machine learning
* Milestone 2: Perform exploratory data analysis using Pandas
* Milestone 3: Apply machine learning algorithms using Scikit-learn
* Conclusion: Data Science in the real world, next steps

**Structured Activity/Exercises/Case Studies:**

* Milestone Project 1: Install and setup Anaconda/Jupyter Notebooks
* Milestone Project 2: Perform exploratory data analysis using Pandas
* Milestone Project 3: Apply machine learning algorithms using Scikit-learn

**Training material provided:** Yes (Digital format)