



## Python Course Content

### Introduction to Python

What is Python Language and features, Why Python and why it is different from other languages, Installation of Python, Anaconda Python distribution for Windows, Mac, Linux. Run a sample python script, working with Python IDE's. Running basic python commands – Data types, Variables, Keywords, etc

**Hands-on Exercise** – Install Anaconda Python distribution for your OS (Windows/Linux/Mac)

### Basic constructs of Python language

Indentation (Tabs and Spaces) and Code Comments (Pound # character); Variables and Names; Built-in Data Types in Python – Numeric: int, float, complex – Containers: list, tuple, set, dict – Text Sequence: Str (String) – Others: Modules, Classes, Instances, Exceptions, Null Object, Ellipsis Object – Constants: False, True, None, NotImplemented, Ellipsis, \_\_debug\_\_; Basic Operators: Arithmetic, Comparison, Assignment, Logical, Bitwise, Membership, Identity; Slicing and The Slice Operator [n:m]; Control and Loop Statements: if, for, while, range(), break, continue, else;

**Hands-on Exercise** – Write your first Python program Write a Python Function (with and without parameters) Use Lambda expression Write a class, create a member function and a variable, Create an object Write a for loop to print all odd numbers

### Writing Object Oriented Program in Python and connecting with Database

Classes – classes and objects, access modifiers, instance and class members OOPS paradigm – Inheritance, Polymorphism and Encapsulation in Python. Functions: Parameters and Return Types; Lambda Expressions, Making connection with Database for pulling data.

### File Handling, Exception Handling in Python

Open a File, Read from a File, Write into a File; Resetting the current position in a File; The Pickle (Serialize and Deserialize Python Objects); The Shelve (Overcome the limitation of Pickle); What is an Exception; Raising an Exception; Catching an Exception;

**Hands-on Exercise** – Open a text file and read the contents, Write a new line in the opened file, Use pickle to serialize a python object, deserialize the object, Raise an exception and catch it

### Mathematical Computing with Python (NumPy)

Arrays and Matrices, ND-array object, Array indexing, Datatypes, Array math Broadcasting, Std Deviation, Conditional Prob, Covariance and Correlation.

**Hands-on Exercise** – Import numpy module, Create an array using ND-array, Calculate std deviation on an array of numbers, Calculate correlation between two variables

## Scientific Computing with Python (SciPy)

Builds on top of NumPy, SciPy and its characteristics, subpackages: cluster, fftpack, linalg, signal, integrate, optimize, stats; Bayes Theorem using SciPy

**Hands-on Exercise** – Import SciPy, Apply Bayes theorem using SciPy on the given dataset

## Data Visualization (Matplotlib)

Plotting Graphs and Charts (Line, Pie, Bar, Scatter, Histogram, 3-D); Subplots; The Matplotlib API

**Hands-on Exercise** – Plot Line, Pie, Scatter, Histogram and other charts using Matplotlib

## Data Analysis and Machine Learning (Pandas) OR Data Manipulation with Python

Dataframes, NumPy array to a dataframe; Import Data (csv, json, excel, sql database); Data operations: View, Select, Filter, Sort, Groupby, Cleaning, Join/Combine, Handling Missing Values; Introduction to Machine Learning(ML); Linear Regression; Time Series

**Hands-on Exercise** – Import Pandas, Use it to import data from a json file, Select records by a group and apply filter on top of that, View the records, Perform Linear Regression analysis, Create a Time Series

## Natural Language Processing, Machine Learning (Scikit-Learn)

Introduction to Natural Language Processing (NLP); NLP approach for Text Data; Environment Setup (Jupyter Notebook); Sentence Analysis; ML Algorithms in Scikit-Learn; What is Bag of Words Model; Feature Extraction from Text; Model Training; Search Grid; Multiple Parameters; Build a Pipeline

**Hands-on Exercise** – Setup Jupyter Notebook environment, Load a dataset in Jupyter, Use algorithm in Scikit-Learn package to perform ML techniques, Train a model Create a search grid

## Web Scraping for Data Science

What is Web Scraping; Web Scraping Libraries (Beautifulsoup, Scrapy); Installation of BeautifulSoup; Install lxml Python Parser; Making a Soup Object using an input html; Navigating Py Objects in the Soup Tree; Searching the Tree; Output Print; Parsing Full or Partial

**Hands-on Exercise** – Install BeautifulSoup and lxml Python parser, Make a Soup object using an input html file, Navigate Py objects in the soup tree, Search tree, Print output

## Python on Hadoop

Understanding Hadoop and its various components; Hadoop ecosystem and Hadoop common; HDFS and MapReduce Architecture; Python scripting for MapReduce Jobs on Hadoop framework

**Hands-on Exercise** – Write a basic MapReduce Job in Python and connect with Hadoop Framework to perform the task

## Writing Spark code using Python

What is Spark, understanding RDDs, Spark Libs, writing Spark code using python, Spark Machine Libraries Mlib, Regression, Classification and Clustering using Spark MLlib

**Hands-on Exercise** – Implement sandbox, Run a python code in sandbox, Work with HDFS file system from sandbox

## Python Projects

**Project 1:** – Python Web Scraping for Data Science

In this project you will be introduced to the process of web scraping using Python. It involves installation of BeautifulSoup, web scraping libraries, working on common data and page format on the web, learning the important kinds of objects, Navigable String, deploying the searching tree, navigation options, parser, search tree, searching by CSS class, list, function and keyword argument.

**Project 2** :-Create a password generator

**Objective** – To generate a password using Python code which would be tough to guess

**Requirements:**–

- To generate a password that is 8-12 characters long
- Password contains at least two special characters
- Password doesn't start with a special character

**Project 3:**– Impact of pre-paid plans on the preferences of investors

**Domain** – Finance

**Objective** – The project aims to find the most impacting factors in preferences of pre-paid model, also identifies which are all the variables highly correlated with impacting factors

**Requirements:**–

- To identify the various reasons for Pre-paid model preference and non-preference among the investors. And also understand the penetration of the Pre-paid model in the brokerage firms
- To identify the Pre-paid scheme advantages and disadvantages and also identify brand wise market share
- In addition to this, the project also looks to identify various insights that would help a newly established brand to foray deeper into the market on a large scale

**Project 4:**– Machine Learning – Prediction of stock prices

**Domain** – Stock Market

**Objective** – This project focuses on Machine Learning by creating predictive data model to predict future stock prices

**Requirements:**–

- Quantitative Value Investing: Predict 6-month price movements based fundamental indicators from companies' quarterly reports
- Forecasting: Build time series models on the delta between implied and actual volatility
- Predict 6-month price movements based fundamental indicators from companies' quarterly reports
- Build time series models on the delta between implied and actual volatility?

**Project 5:** Server logs/Firewall logs

**Objective** – This includes the process of loading the server logs into the cluster using Flume. It can then be refined using Pig Script, Ambari and HCatalog. You can then visualize it using elastic search and excel.

This project task includes:

- Server logs

- Potential uses of server log data
- Pig script
- Firewall logs
- Work flow editor