

Module 1: Introduction to Data Science

- What is Data Science?
- Why Python for data science?
- Versions relevant
- Relevance in industry and need of the hour
- How leading companies are harnessing the power of Data Science with Python?
- Different phases of a typical Analytics/Data Science projects and role of python
- Anaconda vs. Python



Module 2: Python Essentials (Core)

- Overview of Python- Starting with Python
- Introduction to installation of Python
- Introduction to Python Editors & IDE's(Canopy, pycharm, Jupyter, Rodeo, Ipython etc...)
- Understand Jupyter notebook & Customize Settings
- Concept of Packages/Libraries - Important packages(NumPy, SciPy, scikit-learn, Pandas, Matplotlib, etc)
- Installing & loading Packages & Name Spaces
- Data Types & Data objects/structures (strings, Tuples, Lists, Dictionaries)
- List and Dictionary Comprehensions
- Variable & Value Labels – Date & Time Values
- Basic Operations - Mathematical - string - date
- Reading and writing data
- Simple plotting
- Control flow & conditional statements
- Debugging & Code profiling
- How to create class and modules and how to call them?

Module 3: Scientific Distributions used in Python for Data Science

Numpy, scify, pandas, scikitlearn, statmodels, nltk etc

Module 4: Accessing / Importing and Exporting Data using Python Modules

- Importing Data from various sources (Csv, txt, excel, access etc)
- Database Input (Connecting to database)

- Viewing Data objects - subsetting, methods
- Exporting Data to various formats
- Important python modules: Pandas, beautifulsoup

Module 5: Data Manipulation – Cleansing – Munging using Python Modules

- Cleansing Data with Python
- Data Manipulation steps(Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc)
- Data manipulation tools(Operators, Functions, Packages, control structures, Loops, arrays etc)
- Python Built-in Functions (Text, numeric, date, utility functions)

Module 6: Data Analysis – Visualization using Python

- Introduction exploratory data analysis
- Descriptive statistics, Frequency Tables and summarization
- Univariate Analysis (Distribution of data & Graphical Analysis)
- Bivariate Analysis(Cross Tabs, Distributions & Relationships, Graphical Analysis)
- Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ density etc)

Module 7: Python: Machine Learning – Predictive Modelling – Basics

- Introduction to Machine Learning & Predictive Modeling
- Types of Business problems - Mapping of Techniques - Regression vs. classification vs. segmentation vs. Forecasting
- Major Classes of Learning Algorithms -Supervised vs Unsupervised Learning
- Different Phases of Predictive Modeling (Data Pre-processing, Sampling, Model Building, Validation)
- Overfitting (Bias-Variance Trade off) & Performance Metrics
- Feature engineering & dimension reduction

Module 8: Machine Learning Algorithms & Applications – Implementation in Python

- **Linear & Logistic Regression**
- **Segmentation - Cluster Analysis (K-Means)**
- **Decision Trees (CART/CD 5.0)**
- **Ensemble Learning (Random Forest, Bagging & boosting)**
- **Artificial Neural Networks(ANN)**
- **Support Vector Machines(SVM)**
- **Other Techniques (KNN, Naïve Bayes, PCA)**
- **Introduction to Text Mining using NLTK**
- **Introduction to Time Series Forecasting (Decomposition & ARIMA)**
- **Important python modules for Machine Learning (SciKit Learn, stats models, scipy, nltk etc)**

