Data Science with Python: Course Content

- Statistics
- Introduction to Statistics
- Categories of Statistics
- Descriptive
- Inferential
- Population and Sample
- Statistical Analysis Process
- Data Distributions
- Mean,Median,Mode
- Variance and Standard Deviation
- Covariance and correlation
- HypothesisTesting
  - Skewness.
  - Boxplot
  - Five Number Summary
  - Emperical Formula
  - Estimation
  - Confidence interval
  - Scales of Measurements and Data Types
  - Data Summarization
  - Visual Summarization
  - Numerical Summarization
  - Outliers & Summary

Module 1- Introduction to Data Science

Objectives:

This module introduces you to some of the important keywords in python like Business Intelligence, Business

· Analytics, Data and Information. You can also learn how python can play an important role in solving complex analytical problems.

· This module tells you what is python and how it is used by the giants like Google, Facebook, etc.

· Also, you will learn use of 'python' in the industry, this module also helps you compare python with other software

· in analytics, install python and its packages.

Topics:

Business Analytics, Data, Information

· Understanding Business Analytics and python

· Compare python with other software in analytics

· Install python

· Perform basic operations in python using command line

Module 2:

Introduction
Why do we need Python?
Program structure

Execution steps
Interactive Shell
Executable or script files
User Interface or IDE

Module 3:

Memory management and Garbage collections
Object creation and deletion
Object properties

Data Types and Operations
Numbers
Strings
List
Tuple
Dictionary
Other Core Types

Module 4:

Statements and Syntax
Assignments, Expressions and prints
If tests and Syntax Rules
While and For Loops
Iterations and Comprehensions

Module 5:

File Operations
Opening a file
Using Files
Other File tools

Module 6:

Functions
Function definition and call
Function Scope
Arguments
Function Objects
Anonymous Functions

Module 9-:

1. Python Modules for Data Science(8Hours)

- Python for Data Science :Mathematical Computing with Python (numpy)
  - Numpy Introduction (ndarray)
  - Numerical operations on Numpy
  - Numpy Overview
  - Basic operations, types of
  - Accessing elements
  - Shape Manipulation
  - Transpose
  - Slicing
  - Examples
  - Python for Data Science: Data Manipulation with Python(pandas)
  - Understanding Series
  - Understanding Data Frame
  - View and Select Data
  - Missing Values
  - Data Operations
  - Indexing, Selection and Filtering
  - Dropping entries from an axis
  - Concatenation
  - Handling categorical Data(Get Dummies)
  - Python for Data Science: Data Visualization with Python (Matplotlib, Seaborn)
  - Introduction to Matpotlib
  - Customization of Matplotlib
  - Plotting with Pandas
  - Barplots, Histograms plots, Density Plots
  - Introduction to Seaborn, Style Management
  - Plotting with Categorical Data
  - Visualizing Linear Relationships

Module 10- Basics of Statistics & Linear & Multiple Regression

- This module touches the base of Descriptive and Inferential Statistics and Probabilities & 'Regression Techniques'.
- Linear and logistic regression is explained from the basics with the examples and it is implemented in R using two case studies dedicated to each type of Regression discussed.
- Assessing the Accuracy of the Coefficient Estimates.
- Assessing the Accuracy of the Model.
- Estimating the Regression Coefficients.
- Some Important Questions
- Lab: Linear Regression.

i. Libraries .
ii. Simple Linear Regression
iii. Multiple Linear Regression
iv. Interaction Terms
v. Qualitative Predictors
vi. Writing Functions

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion

Module11 - Classification-:

- An Overview of Classification.
- Why Not Linear Regression?
- Logistic Regression
- The Logistic Model
- Estimating the Regression Coefficients
- Making Predictions
- Logistic Regression for >2 Response Classes
- Lab: Logistic Regression.
- The Stock Market Data
- Logistic Regression

NOTE-:

- Assignments with Different Datasets.

Business Scenerio/Group Discussion.

Module-12-Machine Learning vs Statistical Modeling & Supervised vs Unsupervised Learning

- Machine Learning Languages, Types, and Examples
- Machine Learning vs Statistical Modelling
- Supervised vs Unsupervised Learning
- Supervised Learning Classification
- Unsupervised Learning

Module1 3- Supervised Learning I

- K-Nearest Neighbors
- Decision Trees
- Random Forests
- Reliability of Random Forests
- Advantages & Disadvantages of Decision Trees

Module14 - Supervised Learning II

- Regression Algorithms
- Model Evaluation
- Model Evaluation: Overfitting & Underfitting
- Understanding Different Evaluation Models

Module 15 - Unsupervised Learning

- K-Means Clustering plus Advantages & Disadvantages
- Hierarchical Clustering plus Advantages & Disadvantages
- Measuring the Distances Between Clusters - Single Linkage Clustering
- Measuring the Distances Between Clusters - Algorithms for Hierarchy Clustering
- Density-Based Clustering

Module 16- Tree-Based Methods-:

- The Basics of Decision Trees
- Regression Trees
- Classification Trees
- Trees Versus Linear Models
- Advantages and Disadvantages of Trees
- Bagging, Random Forests, Boosting
- Bagging
- Random Forests
- Lab: Decision Trees
- Fitting Classification Trees
- Fitting Regression Trees

NOTE-:

- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 17- Support Vector Machine(SVM):

- Linear Classifiers
- Margin of SVM's
- SVM optimization
- SVM for Data which is not linear separable
- Learning non-linear patterns
- Kernel Trick
- SVM Parameter Tuning
- Linear SVM using Python