Benefits of Self-Learning Data Science Course

How will this data science self-learned beginner course benefit me?

Prepare yourself for a career as a Data Analyst and Data Scientist.

- Live online faculty-led training
- Learn NumPy - foundation library for Data Science in Python
- Learn SciPy - key algorithms core to Python's scientific computing
- Learn Pandas - library for data analysis and manipulation
- Learn Matplotlib - python module for visualization to make graphs, pie charts
- Learn SciKit - python module for machine learning

Data Science Course Curriculum

Module 1

Introduction to Python Programming

- Introduction to Data Science
- Introduction to Python
- Basic Operations in Python
- Variable Assignment
- Functions: in-built functions, user-defined functions
- Condition: if, if-else, nested if-else, else-if

Module 2

Data Structure - Introduction

- List: Different Data Types in a List, List in a List
- Operations on a list: Slicing, Splicing, Sub-setting
- Condition(true/false) on a List
- Applying functions on a List
- Dictionary: Index, Value
- Operation on a Dictionary: Slicing, Splicing, Sub-setting
- Condition(true/false) on a Dictionary
- Applying functions on a Dictionary
- Numpy Array: Data Types in an Array, Dimensions of an Array
- Operations on Array: Slicing, Splicing, Sub-setting
- Conditional(T/F) on an Array
- Loops: For, While
- Shorthand for *For*
- Conditions in shorthand for *For*

Module 3

Basics of Statistics

- Statistics & Plotting
- Seabourn & Matplotlib - Introduction
- Univariate Analysis of a Data
- Plot the Data - Histogram plot
- Find the distribution
- Find mean, median, and mode of the Data
- Take multiple data with the same mean but different sd, same mean and sd but different kurtosis: find mean, sd, plot
- Multiple data with different distributions
- Bootstrapping and sub-setting
- Making samples from the Data
- Making stratified samples - covered in bivariate analysis
- Find the mean of the sample
- Central limit theorem
- Plotting
- Hypothesis testing + DOE
- Bivariate analysis
- Correlation
- Scatter plots

- Making stratified samples
- Categorical variables
- Class variable

Module 4

Use of Pandas

- File I/O
- Series: Data Types in series, Index
- Data Frame
- Series to Data Frame
- Re-indexing
- Operations on Data Frame: Slicing, Splicing (also Alternate), Sub-setting
- Pandas
- Stat operations on Data Frame
- Reading from different sources
- Missing data treatment
- Merge, join
- Options for look and feel of the data frame
- Writing to file
- DB operations

Module 5

Data Manipulation & Visualization

- Data Aggregation, Filtering, and Transforming
- Lambda Functions
- Apply, Group-by
- Map, Filter, and Reduce
- Visualization
- Matplotlib, pyplot
- Seaborn
- Scatter plot, histogram, density, heat-map, bar charts

Module 6

Linear Regression

- Regression - Introduction
- Linear Regression: Lasso, Ridge
- Variable Selection
- Forward & Backward Regression

Module 7

Logistic Regression

- Logistic Regression: Lasso, Ridge
- Naive Bayes

Module 8

Unsupervised Learning

- Unsupervised Learning - Introduction
- Distance Concepts
- Classification
- k nearest
- Clustering
- k means
- Multidimensional Scaling
- PCA

Module 9

Random Forest

- Decision trees
- Cart C4.5
- Random Forest
- Boosted Trees
- Gradient Boosting

Module 10

SVM

- SVM - Introduction
- Hyper-plane
- Hyper-plane to segregate to classes
- Gamma

Data Science Beginner Projects

- Walmart Store Sales Forecast using Linear Regression Models
- Predict the Survival of passengers on Titanic using Logistic Regression
- Bike Sharing Demand Problem
- Clustering of MNIST Digit Image Data
- Predict the Driver Alertness using Tree-Based Classification Algorithm