SoC Softech

Technologies for Innovation

Python with Data Science Course Curriculum

Duration : 90 hours **Fee : ------**

Website : www.socsoftech.com

Phone no: +91 9640987960 , 7801036796

Mail : info@socsoftech.com

socsoftech@gmail.com

Module 1 : Python Programming Language

Part A : Python Basic Concepts

1. Introduction to Python and its involvement with Data Science
2. Understanding Object Orientation Programming
3. Installation: Python 3.6 or later version, pip, iPython, Sublime Text Editor, Anaconda(Jupyter and Spyder)
4. Python Identifiers, Naming Conventions, Variables and Types
5. Defining Functions, Classes and Methods
6. Understanding Indentation
7. Executing sample programs in all Editors
8. Difference Between Functions and Methods
9. How to use Python Functions and Methods
10. Decision making through conditions and Loops
11. Declaring instances and Workout its accessibility
12. Understanding global and local variables in python
13. Instantiating Classes and flow of execution
14. Accessing Methods, Variables, Global variables and Functions
15. Working with self and super keywords
16. Object String representation through __str__ and __repr__
17. Constructors; Initialization; object: a base class
18. Inheritance Concept; Overriding and Overloading concept
19. Constructors with respect to inheritance
20. Understanding __name__ == '__main__'

21. Exceptions:
    a. Overview of exception
    b. Raising common causing exceptions
    c. Exception Hierarchy
    d. Raising exception at calling method
    e. Handling exceptions through try, except, else and finally
    f. Exception propagation
    g. Customized Exceptions

Part B: Data Structures:

a. List: Creating, Accessing, Slicing, Manipulating lists, Built-in Functions & Methods in list, Iterating & Enumerating list data and Working with Nested lists.
b. Tuple, Set and Dictionaries (same above all operations)
c. Handling conversions of sample data with Data Structures

Part C: Regular Expressions in Python

a. Patterns, searching, Modifiers, flags
b. Working with examples to find specific strings, phone numbers, email addresses and filtering html data with regular expressions

File I/O

c. Working with text files and .csv
d. Reading and Writing data to the files
e. Importing required packages to work with .csv

Module2 : Statistics - Probabilities and Linear Algebra

a. Statistical thinking in Python and approach of Data Analysis
b. Fundamental statistics terms and its definitions
c. Applying basic statistics in Python with *NumPy*
    i. Cumulative Distribution functions
    ii. Modelling Distributions
d. Graphical exploratory data analysis with Python
e. Probability theories:
    i. Ranges, Mean, Variance, Standard Deviation and various distributions
    ii. Mass and Density functions
    iii. Kernel density estimation
    iv. Understanding Bayes theorem and predictions*
f. Estimation
    i. Sampling distributions, bias and Exponential distributions
g. Hypothesis testing
    i. Hypothesis Test
    ii. Testing Correlation and Proportions
    iii. Chi-Squared Tests
    iv. Errors, Power and Replication

b. NumPy: N-dimensional array operations
    a. Array creations, conversions, dimensional understandings, shaping, reshaping, generating sample large datasets, Linear algebra functionalities and numerical operations etc…

c. SciPy: High-level Scientific Computing
    a. Linear Algebra operations
    b. Interpolation
    c. Optimization and fit
    d. Statistics and random numbers
    e. Numerical Integration
    f. Fast Fourier transforms
    g. Signal processing and image manipulation

Module3 : Data Mining & Data Analytics (Data Harvesting, Cleansing, Analyzing and Visualizing)

Part A :Pandas and NumPy Functionalities:

a. Introduction
b. Pandas DataFrame basics
    a. Understanding data, looking at columns, rows and cells
    b. Subsetting Columns, Rows with methods
    c. Grouped and Aggregated Calculations
        i. Frequency Means and Counts
    d. Basic plot
c. Pandas Data Structures
    a. Creating your own data (Series and DataFrame)
    b. Series (also called as Vector) Object operations
        i. Broadcasting and Scalar operations
    c. DataFrame Broadcasting (Vectorized)
    d. Making changes to Series and DataFrame
        i. Adding additional Columns
        ii. Dropping values
    e. Exporting and Importing Data

Part B : Introduction to Plotting:

a. Introduction
b. Matplotlib

c. Statistical Graphics using *matplotlib*
    a. Univariate
    b. Bivariate
    c. Multivariate Data
d. *Seaborn Library* Plotting methodology
    a. Univariate, Bivariate and Multivariate
e. Pandas Objects Plotting
    a. Histogram, Density Plot, Scatterplot, Hexbin Plot and Boxplot
f. Seaborn Themes and Styles

Part C : Data Manipulation:

a. Data Assembly
    a. Concatenations and Merging Multiple datasets
b. Missing Data:
    a. Introduction
    b. What is a *NaN* Value
    c. Working with merged data, user input values and Re-indexing
    d. Working with missing data
        a. Finding and Counting missing data
        b. Cleansing missing data
        c. Calculations with missing data
        d. Conclusion Understanding Multiple Observations (Normalization)

Part D : Data Munging:

a. Understanding Data Types
b. Converting types
c. Categorical Data
    a. Convert to Category
    b. Manipulating Categorical Data
d. Strings and Text Data
    a. String Subsettings
    b. String Methods
    c. String Formatting
e. Apply and Groupby Operations:
    a. Introduction
    b. Functions
    c. Apply over a Series and DataFrame
    d. Apply- Column-wise and Row-wise operations
    e. Groupby Operation:
    f. Aggregate Methods and Functions
f. The datetime Data Type:
    a. Python's datetime Object
    b. Loading, Converting, Extracting Date components
    c. Date Calculations
    d. Datetime Methods
    e. Subsetting datetime, Date Ranges, Shifting Values, TimeZones

Module 4 : Machine Learning (Data Modelling)

a. Linear Models
    a. Linear and Multiple Regressions using *statsmodels* and *sklearn*
b. Generalized Linear Models
    a. Logistic and Poisson Regressions using *statsmodels* and *sklearn*
    b. Survival Analysis
c. Model diagnostics
    a. Residuals
    b. Comparing Multiple Models
    c. *k*-Fold Cross-Validation
d. Regularization
e. Clustering
    a. *k*-Means, Dimension Reduction with PCA (Principal Component Analysis)
    b. Hierarchical Clusterings
    c. Conclusions

Practical Data Analysis and Understandings

Data Science Interview Questions Discussions (2 sessions)

Note: Keeping main objective as "Understanding" All the above topics are covered with logical and programmatic approach in Python. Also please note that Content order is NOT compulsorily followed at the time of delivering subject and knowledge