

Certificate Course Offered by Department of Information Technology

Name of the course: Python for Data Science

Course Code: 400004/IT100B

Objectives

1. To gain hands-on experience and practice using Python to solve real data science challenges.
2. To familiarize the students to practice Python programming and coding for modeling, statistics, and storytelling.
3. To utilize popular libraries such as Pandas, numPy, matplotlib, and SKLearn.
4. To enable the students to get hands-on experience creating analytics models and apply those models to real-world problems.

Course Outcomes

Sl. No.	Description
CO 1	Identify the need for data science and solve basic problems using Python built-in data types and their methods.
CO 2	Design an application with user-defined modules and packages using OOP concept.
CO 3	Exemplify the numerical computation with “Numpy” library.
CO4	Apply the data transformation and data manipulation operations using “pandas”.
CO5	Analyze nature of data with help of different tools and visualization.
CO6	Implement statistical analysis techniques for solving practical problems.

Syllabus

1. Overview of Python and Data Structures

Introduction to Data Science - Why Python? - Essential Python libraries - Python Introduction- Features, Identifiers, Reserved words, Indentation, Comments, Built-in Data types and their Methods: Strings, List, Tuples, Dictionary, Set - Type Conversion- Operators. Decision Making- Looping- Loop Control statement- Math and Random number functions. User defined functions - function arguments & its types.

2. File, Exception Handling and OOP

User defined Modules and Packages in Python- Files: File manipulations, File and Directory related methods - Python Exception Handling. OOPs Concepts -Class and Objects, Constructors – Data hiding- Data Abstraction- Inheritance.

3. NumPy for Simulation Modeling

Introduction to NumPy - Basics of NumPy Arrays, Computation on NumPy Arrays- indexing, slicing, reshaping. Universal Functions, Aggregations.Computation on Arrays – broadcasting, comparisons, Fancy indexing, Sorting Arrays, Structured Arrays.

4. Data wrangling, Reshaping and Summarizing with pandas

Introducing Pandas Objects – series, data frames, index, Processing CSV, JSON, XLS data, Operations on Pandas Objects – indexing and selection, universal functions, missing data, hierarchical indexing, Combining Dataset – concat and append, merge and join. Aggregation and grouping, Pivot tables, Vectorized string operations, Working with time series, High performance Pandas – eval(), query().

5. Data Visualization using Matplotlib

General Matplotlib, Simple Line Plots, Simple Scatter Plots, Density and Contour Plots, Histograms, Binnin, and Density, Customizing Plot Legends, Customizing Colorbars, Text and Annotation, Three-Dimensional Plotting in Matplotlib, Geographic Data with Basemap, Visualization with Seaborn.

6. Statistical Modeling using Python

Introduction to probability - Probability distributions, Sampling and sampling distribution, Hypothesis Hypothesis testing - Two sample testing, Introduction to ANOVA, Two way ANOVA, Regression - Linear regression, Multiple regression, Clustering analysis, Classification and Regression Trees (CART).

Practical Sessions:

1. Perform Creation, indexing, slicing, concatenation and repetition operations on Python built-in data types: Strings, List, Tuples, Dictionary, Set
2. Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem.
3. Handle numerical operations using math and random number functions.
4. Create user-defined functions with different types of function arguments.
5. Perform File manipulations- open, close, read, write, append and copy from one file to another.
6. Write a program to implement OOP concepts like Data hiding and Data Abstraction.
7. Create NumPy arrays from Python Data Structures, Intrinsic NumPy objects and Random Functions.
8. Manipulation of NumPy arrays- Indexing, Slicing, Reshaping, Joining and Splitting.
9. Computation on NumPy arrays using Universal Functions and Mathematical methods.
10. Load an image file and do crop and flip operation using NumPy Indexing.
11. Create Pandas Series and Data Frame from various inputs.
12. Import any CSV file to Pandas Data Frame and perform the following:
 - (a) Visualize the first and last 10 records
 - (b) Get the shape, index and column details
 - (c) Select/Delete the records (rows)/columns based on conditions.
 - (d) Perform ranking and sorting operations.
 - (e) Do required statistical operations on the given columns.
 - (f) Find the count and uniqueness of the given categorical values.
 - (g) Rename single/multiple columns
13. Import any CSV file to Pandas Data Frame and perform the following:
 - (a) Handle missing data by detecting and dropping/ filling missing values.
 - (b) Transform data using apply () and map() method.
 - (c) Detect and filter outliers.
 - (d) Perform Vectorized String operations on Pandas Series.
14. Visualize data using Line Plots, Bar Plots, Histograms, Density Plots and Scatter Plots using Matplotlib.
15. Statistical Analysis using Python.

References

1. Wesley J. Chun, “Core Python Programming”, Prentice Hall,2006.
2. Mark Lutz, “Learning Python”, O’Reilly, 4th Edition, 2009.
3. Y. Daniel Liang, “Introduction to Programming using Python”, Pearson,2012.

4. Wes McKinney, “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython”, O’Reilly, 2nd Edition,2018.
5. Jake VanderPlas, “Python Data Science Handbook: Essential Tools for Working with Data”, O’Reilly, 2017.
6. Stefanie Molin,” Hands-On Data Analysis with Pandas”, Packt Publishing Ltd,2019.
7. Allan Visochek,” Practical Data Wrangling”, Packt Publishing Ltd,2017.
8. Andreas C. Muller, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, O’Reilly,2016.
9. McKinney, W., "Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc., 2012.

Resource Persons

1. Dr. Neeba E A

Associate Professor & HoD
 Department of Information Technology
 Email: neebaea@rajagiritech.edu.in

2. Dr. Ranju S Kartha

Assistant Professor
 Department of Information Technology
 Email: ranjusk@rajagiritech.edu.in

Online Division of Teaching & Learning

Topic		Resource Person	Total Hours
1	Overview of Python and Data Structures	Dr. Ranju S Kartha	8
2	File, Exception Handling and OOP	Dr. Ranju S Kartha	8
3	NumPy for Simulation Modeling	Dr. Neeba E A	8
4	Data wrangling, Reshaping and Summarizing with pandas	Dr. Neeba E A	9

5	Data Visualization using Matplotlib	Dr. Ranju S Kartha	9
6	Statistical Modeling using Python	Dr. Neeba E A	8

Mode of Delivery: Online/ Offline

Duration: 22 Hrs. of Theory & 28 Hrs. of Practical Session

Fee Structure

The registration fee for inhouse candidates - **Rs 500/-**

The registration fee for external candidates - **Rs 1000/-**

Eligibility Criteria

This certificate course is mainly for the students pursuing B.Tech. in Computer Science, Information Technology, Electronics and Communication, Applied Electronics and Instrumentation, B.Sc. Computer Science & Electronics. Those who are completed plus two are also eligible for this course.

Evaluation scheme

Assignment/Quiz: Total Marks: 40

Assignments will be provided after the completion of each module.

Exam : (Marks: 60)

The exam will be conducted after the completion of the entire course.

Cut off mark

Those students who acquire a minimum of **60** marks from both the assignments and the exam will be eligible to get the certificate

Project

The interested inhouse students will get a chance to carry out a project after the successful completion of the course.