SYLLABUS

Subject: Computer Science	**Semester: V**
Course Title: Python for Data Science	**Course Code: 20CSSEC32PD3**
No. of Hours: 45 LTP: 300	**Credits: 3**

Objectives

- To learn the core programming basics required for Data Science using Python language.
- To gain knowledge on Data Science modules NumPy, SciPy and Matplotlib.

Course Outcomes

CO1: Identify the need for data science and solve basic problems using Python built-in data types and their methods.

CO2: Design an application with user-defined modules and packages using OOP concept

CO3: Employ efficient storage and data operations using NumPy arrays.

CO4: Apply powerful data manipulations using Pandas.

CO5: Do data pre-processing and visualization using Pandas

UNIT-I (9 Hrs.)

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 – Programming Exercises.

UNIT-II (9 Hrs.)

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 – Programming Exercises.

UNIT-III (9 Hrs.)

NumPy Basics: Arrays and Vectorized Computation- The NumPy ndarray-Creating ndarrays- Data Types for ndarrays- Arithmetic with NumPy Arrays-Basic Indexing and Slicing - Boolean Indexing-Transposing Arrays and Swapping Axes.
Universal Functions: Fast Element-Wise Array Functions- Mathematical and Statistical Methods-Sorting- Unique and Other Set Logic – Programming Exercises.

UNIT-IV (9 Hrs.)

Introduction to pandas Data Structures: Series, Data Frame and Essential Functionality: Dropping Entries- Indexing, Selection, and Filtering- Function Application and Mapping- Sorting and Ranking.

Summarizing and Computing Descriptive Statistics- Unique Values, Value Counts, and Membership. Reading and Writing Data in Text Format – Programming Exercises.

UNIT-V (9 Hrs.)

Data Cleaning and Preparation: Handling Missing Data - Data Transformation: Removing Duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Detecting and Filtering Outliers- String Manipulation: Vectorized String Functions in pandas.

Plotting with pandas: Line Plots, Bar Plots, Histograms and Density Plots, Scatter or Point Plots – Programming Exercises.

Co-Curricular Activities

- Assignments on problem solving
- Group discussions
- Student presentations and seminars
- Online quizzes
- Project work

Prescribed Books

1. Y. Daniel Liang, "Introduction to Programming using Python", Pearson, 2012.
2. Wes McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly, 2nd Edition, 2018.
3. Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly, 2017.

Reference Books

1. Wesley J. Chun, "Core Python Programming", Prentice Hall, 2006.
2. Mark Lutz, "Learning Python", O'Reilly, 4th Edition, 2009.

Web resources:

a. https://www.edx.org/course/python-basics-for-data-science
b. https://www.edx.org/course/analyzing-data-with-python
c. https://www.coursera.org/learn/python-plotting?specialization=data-science- python
d. https://www.programmer-books.com/introducing-data-science-pdf/

Subject: Computer Science   **Semester: V**

Course Title: Python for Data Science   **Course Code: 20CSSEC32PD3**

Time: 3 Hrs.   **Max. Marks: 100**

SECTION – A

Answer **ALL** questions   **20 x 1 = 20 M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
1	I	1	
2	I	1	
3	II	1	
4	II	1	
5	III	1	
6	III	1	
7	IV	1	**No. of questions to be set**
8	IV	1	RBT1 – 8
9	V	1	RBT2 – 8
10	V	1	RBT3 – 2
11	I	1	RBT4 – 2
12	I	1	
13	II	1	
14	II	1	
15	III	1	
16	III	1	
17	IV	1	
18	IV	1	
19	V	1	
20	V	1	

SECTION – B

Answer any **FOUR** questions   **4 x 8 = 32 M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
21	I	8	**No. of questions to be set**
22	II	8	RBT1 – 2
23	III	8	RBT2 – 2
24	IV	8	RBT3 – 1
25	V	8	RBT4 – 1
26	I / II / III / IV / V	8	

SECTION – C

Answer any **FOUR** questions   **4 x 12 = 48 M**

Q. No.	UNIT	Marks Weightage	RBT LEVEL
27	I	12	**No. of questions to be set**
28	II	12	RBT1 – 2
29	III	12	RBT2 – 2
30	IV	12	RBT3 – 1
31	V	12	RBT4 – 1
32	I / II / III / IV / V	12	

MARIS STELLA COLLEGE, VIJAYAWADA-8
(An autonomous college affiliated to Krishna University)
Model Question Paper

Subject: **Computer Science** Semester: **V**

Course Title: **Python for Data Science** Course Code: **20CSSEC32PD3**

Time: **3 Hrs.** Max. Marks: **100**

SECTION – A

Answer **ALL** questions **20 x 1 = 20 M**

1. Is Python code compiled or interpreted?
   A. Python code is both compiled and interpreted
   B. Python code is neither compiled nor interpreted
   C. Python code is only compiled
   D. Python code is only interpreted

2. What will be the output of the following Python code?
```
i = 0
while i < 5:
   print(i)
i += 1
   if i == 3:
      break
else:
   print(0)
```
   A. 0 1 2 0
   B. 0 1 2
   C. error
   D. none of them

3. Which of these definitions correctly describes a module?
   A. Denoted by triple quotes for providing the specification of certain program elements
   B. Design and implementation of specific functionality to be incorporated into a program
   C. Defines the specification of how it is to be used
   D. Any program that reuses code

4. What will be the output of the following Python code?
```
def printMax(a, b):
   if a > b:
      print(a, 'is maximum')
   elif a == b:
      print(a, 'is equal to', b)
   else:
      print(b, 'is maximum')
printMax(3, 4)
```

A. 3

B. 4

C. 4 is maximum

D. None of the mentioned

5. Suppose B is a subclass of A, to invoke the __init__ method in A from B, what is the line of code you should write?

A. A.__init__(self)

B. B.__init__(self)

C. A.__init__(B)

D. B.__init__(A)

6. Which of these is not a fundamental features of OOP?

A. Encapsulation

B. Inheritance

C. Instantiation

D. Polymorphism

7. Which of the following is the most suitable definition for encapsulation?

A. Ability of a class to derive members of another class as a part of its own definition

B. Means of bundling instance variables and methods in order to restrict access to certain class members

C. Focuses on variables and passing of variables to functions

D. Allows for implementation of elegant software that is well designed and easily modified

8. Pandas is an open-source _____ Library?

A. Ruby

B. Javascript

C. Java

D. Python

9. Which of the following makes use of pandas and returns data in a series or dataFrame?

A. pandaSDMX

B. freedapi

C. OutPy

D. Inpy

10. Which of the following takes a dict of dicts or a dict of array-like sequences and returns a DataFrame?

A. DataFrame.from_items

B. DataFrame.from_records

C. DataFrame.from_dict

D. All of the above

11. NumPY stands for _____.

12. The most important object defined in NumPy is an N-dimensional array type called _____.

13. If a dimension is given as ____ in a reshaping operation, the other dimensions are automatically calculated.
14. What will be syntax for pandas dataframe? _____
15. _____ is also known as axis array.
16. What will be output for the following code?
    import pandas as pd
    s = pd.Series([1,2,3,4,5],
    index = ['a','b','c','d','e']) print s['a']
17. _____developed Python Programming Language.
18. _____ is the extension of the Python file.
19. _____ keyword is used for function.
20. _____ makes use of pandas and returns data in a series or dataframe.

SECTION – B

Answer any **FOUR** questions   **4 x 8 = 32 M**

21. What are the essential features of python?
22. Explain in detail about various operators.
23. Demonstrate exception handling.
24. Explain in detail about various functions in NumPy.
25. Explain about Indexing, Selection, and Filtering.
26. How is the data transformation done in pandas?

SECTION – C

Answer any **FOUR** questions   **4 x 12 = 48 M**

27. Explain in detail about looping statements.
28. Explain in detail about decision making statements.
29. Describe various OOPS concepts.
30. Explain about types of NumPy arrays.
31. How to read and write data in text format?
32. Explain about plotting with pandas.