DEPARTMENT OF CSE - DATA SCIENCE

Programme: CSE-DS					Semester: III
Course Code	Course Name	L	T	P	C
20CS3T03	**FUNDAMENTALS OF DATA SCIENCE**	3	0	0	3
Subject Category :	PCC				

Course Objectives:

- To provide a comprehensive knowledge of data science using Python.
- To learn the essential concepts of data analytics and data visualization

Course Outcomes:

Upon successful completion of the course, the student will be able to:

CO1 : Apply principles of NumPy and Pandas to the analysis of data.
CO2 : Make use of various file formats in loading and storage of data.
CO3 : Identify and apply the need and importance of pre-processing techniques.
CO4 : Show the results and present them in a pictorial format.

UNIT I

Data science: definition, Datafication, Exploratory Data Analysis, The Data science process, A data scientist role in this process.

NumPy Basics: The NumPy ndarray: A Multidimensional Array Object, Creating ndarrays ,Data Types for ndarrays, Operations between Arrays and Scalars, Basic Indexing and Slicing, Boolean Indexing, Fancy Indexing, Data Processing Using Arrays, Expressing Conditional Logic as Array Operations, Methods for Boolean Arrays , Sorting , Unique.

UNIT II

Getting Started with pandas: Introduction to pandas, Library Architecture, Features, Applications, Data Structures, Series, DataFrame, Index Objects, Essential Functionality Reindexing, Dropping entries from an axis, Indexing, selection, and filtering),Sorting and ranking, Summarizing and Computing Descriptive Statistics, Unique Values, Value Counts, Handling Missing Data, filtering out missing data.

UNIT III

Data Loading, Storage, and File Formats : Reading and Writing Data in Text Format, Reading Text Files in Pieces, Writing Data Out to Text Format, Manually Working with Delimited Formats, JSON Data, XML and HTML: Web Scraping, Binary Data Formats,Using HDF5 Format, Reading Microsoft Excel Files, Interacting with Databases, Storing and Loading Data in MongoDB

UNIT IV

Data Wrangling: Combining and Merging Data Sets, Database style DataFrame Merges, Merging on Index, Concatenating Along an Axis, Combining Data with Overlap , Reshaping and Pivoting, Reshaping with Hierarchical Indexing, Data Transformation, Removing Duplicates, Replacing Values.

UNIT V

Plotting and Visualization: A Brief matplotlib API Primer, Figures and Subplots, Colors, Markers, and Line Styles, Ticks, Labels, and Legends, Annotations and Drawing on a Subplot, Saving Plots to File,Plotting Functions in pandas, Line Plots, Bar Plots, Histograms and Density Plots, Scatter Plots.

Text Books :

1. Wes McKinney, "Python for Data Analysis",O'REILLY, ISBN:978-1-449-31979-3, 1st edition, October 2012.
2. Rachel Schutt & O'neil, "Doing Data Science", O'REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013

Reference Books:

1. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media, 2015
Matt Harrison, "Learning the Pandas Library: Python Tools for Data Munging, Analysis, and Visualization , O'Reilly, 2016.