UOE032: Introduction to Data Science							
University Open Elective-III(CSE & AIML)							
Lect.	**Tut.**	**Pract.**	**Credits**	**Evaluation Scheme**			
				Component	**Exam**	**Weightage %**	**Pass %**
2	-	-	2	Theory 100 Marks	FA	50	40
					SA	50	40

Pre-Requisite Courses: Basics of Database

Course Learning Outcomes: After completion of course, students would be able:

CO1 Understand[2] data collection, management in data science;
CO2 Explain[2] key concepts in data science applicable in real-world applications
CO3 Use[3] tools/techniques of data science for real world problem solving

I. Introduction

Introduction to Data Science, Different Sectors using Data science, Purpose and Components of Python in Data Science **7**

II. Data preprocessing **7**

Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization

III. Data analysis

Data Analytics Process, Knowledge Check, Exploratory Data Analysis (EDA), EDA- Quantitative technique, EDA- Graphical Technique, Data Analytics Conclusion and Predictions. **7**

IV. Data and features

Feature Generation and Feature Selection (Extracting Meaning from Data)- Motivating application: user (customer) retention- Feature Generation (brainstorming, role of domain expertise, and place for imagination)- Feature Selection algorithms. **7**

V. Data visualization

Data Visualization- Basic principles, ideas and tools for data visualization, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset. **7**

VI. Data privacy **7**

Applications of Data Science, Data Science and Ethical Issues- Discussions on privacy, security, ethics- A look back at Data Science- Next-generation data scientists

Text Books

1.Data Sciences & Analytics, V.K. Jain, Khanna Publishing House.
Business Analytics: The Science of Data - Driven Decision Making, U Dinesh Kumar, John Wiley & Sons.
2.Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, John Wiley & Sons.

References:

Joel Grus, Data Science from Scratch, Shroff Publisher/O'Reilly Publisher Media

1. Annalyn Ng, Kenneth Soo, Numsense! Data Science for the Layman, Shroff Publisher Publisher
2. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher.
3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.
4. Jake VanderPlas, Python Data Science Handbook, Shroff Publisher/O'Reilly Publisher Media.
5. Philipp Janert, Data Analysis with Open Source Tools, Shroff Publisher/O'Reilly Publisher Media.