ASSESSMENT RUBRICS

	Marks
End Semester Evaluation	**65**
Continuous Evaluation	**35**

KU2MDCELE117: R and Python for Data Analysis

Semester	Course Type	Course Level	Course Code	Credits (L+T+P)	Total Hours (L+T+P)
2	MDC	100	KU2MDCELE117	2+0+1	30+0+30

Learning Approach (Hours/ Week)			Marks Distribution			Duration of ESE (Hours)
Lecture	Practical	Tutorial	CE	ESE	Total	
2	2	0	25	50	75	1.5Hr

Course Description:

This course offers a comprehensive introduction to two of the most widely used programming languages in data analysis: Python and R. Participants will learn the fundamentals of both languages and will explore data visualization techniques. The course will also cover statistical analysis topics such as probability distributions and hypothesis testing using both Python and R. By the end of this course, participants will have the skills and confidence to analyze and visualize data effectively using Python and R, making them valuable assets in the rapidly growing field of data analytics.

Course Prerequisite: Basic knowledge of programming concepts is helpful but not mandatory.

Course Outcomes:

CO No.	Expected Outcome	Learning Domains
1	Explain the fundamentals of programming in Python	U
2	Explain the fundamentals of programming in R	U
3	Demonstrate data visualization tasks using Python and R	U
4	Apply probability distributions to draw insights from data using Python and R	A
5	Apply hypothesis testing to draw insights from data using Python and R	A

****Remember (R), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C)***

Mapping of Course Outcomes to PSOs

	PSO 1	PSO 2	PSO 3	PSO 4	PSO 5	PSO 6	PSO 7
CO 1	2						
CO 2	2						
CO 3	2	3			3	3	
CO 4	3	3			3	3	
CO 5	3	3			3	3	

COURSE CONTENTS

Contents for Classroom Transaction:

MODULE	**UNIT**	**DESCRIPTION**	**HOURS**
1	**Module I : Introduction to Python and Pandas**		**5**
	1.1	Python installation, variables and operators, reading data from keyboard, Data types - numbers, lists, dictionary	

	1.2	Decision making Loops - for, while; Loop control statements - break, continue, pass	
	1.3	Introduction to Pandas - Pandas Installation, data structures - series, data frame	
	1.4	Reading from csv files, loc(), iloc() function, writing back to csv files	
2	**Module II : Introduction to R Programming**		**10**
	2.1	R programming installation and running, Variables, Operators, Data Types	
	2.2	Decision making, Loops, Functions	
	2.3	Packages - installation, loading, Introduction to dplyr Library, dplyr functions	
	2.4	Reading from csv files, writing back to csv files,	

3	**Module III : Data Visualization**		**5**
	3.1	Bar charts, grouped bar charts, stacked bar charts, histogram, line charts using Python	
	3.2	Pie charts, box plots, scatter plots, density plots, strip charts, QQ plots using Python	
	3.3	Bar charts, grouped bar charts, stacked bar charts, histogram, line charts using R	
	3.4	Pie charts, box plots, scatter plots, density plots, strip charts, QQ plots using R	

4	**Module IV : Data Analysis**		**5**
	4.1	Binomial and exponential distribution using Python	
	4.2	One sample Z-Test and two sample Z-Test using Python	
	4.3	Binomial and exponential distribution using R	
	4.4	One sample Z-Test and two sample Z-Test using R	
5	**Teacher Specific Module**		**5**

	Directions: Teacher can design this module based on their expertise and demand from the students	
	Space to fill the selected area/ activity	

Essential Readings:

1. PYTHON FOR DATA ANALYSIS: Data Wrangling with pandas, NumPy, and Jupyter by Wes Mckinney : O′Reilly

2. Python Data Science Handbook: Essential Tools For Working With Data by Jake VanderPlas : O′Reilly

3. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data by Hadley Wickham and Garrett Grolemund : Shroff/O'Reilly, 2017

4. Hands on Programming With R: Write Your Own Functions and Simulations by Garrett Grolemund, Shroff/O'Reilly

Practicals:

ASSESSMENT RUBRICS

	Marks
End Semester Evaluation	**50**
Continuous Evaluation	**25**