PROGRAM	**Master of Business Administration**
SEMESTER	**3**
COURSE TITLE	**Data Science Using R**
COURSE CODE	**04MB0341**
COURSE CREDITS	**03**
COURSE DURATION	**42 Hrs (42 sessions of 60 minutes each)**

COURSE OUTCOMES:

- Apply R code to use the API (Application Programming Interface) to manage databases.
- Select the right functions and control structure of R Programming language.
- Combine various tools and packages of R programming language for business analytics.
- Analyze data graphically by creating various plots using visualization tools in R.
- Analyze business data using simple linear regression and multiple linear regression for prediction and decision making.

COURSE CONTENTS:

Unit No	Unit / Sub Unit	Sessions
I	**R Nuts and Bolts, Getting Data In and Out of R & Using Textual and Binary Formats for Storing Data** **R Nuts and Bolts:** Entering Input, Evaluation, R Objects, Numbers, Attributes, Creating Vectors, Mixing Objects, Explicit Coercion, Matrices, Lists, Factors, Missing Values, Data Frames, Names **Getting Data In and Out of R:** Reading and Writing Data. Reading Data Files with, read .table (), Reading in Larger Datasets with read. table, Calculating Memory Requirements for R Objects	06
II	**Sub setting R Objects, Managing Data Frames with the dplyr package & Control Structures** **Sub setting R Objects:** Sub setting a Vector, Sub setting a Matrix, Sub setting Lists, Sub setting Nested Elements of a List, Extracting Multiple Elements of a List, Partial Matching, Removing NA Values **Managing Data Frames with the dplyr package:** Data Frames, The dplyr Package, dplyr Grammar, Installing the dplyr package, select() , filter() , arrange() , rename() ,mutate(), group_by(), %>%	08
III	**Control Structures:** if-else for Loops, Nested for loops , while Loops , repeat Loops , next, break **Functions, Scoping Rules of R & Loop Functions:** **Functions:** Functions in R, Your First Function, Argument Matching, Lazy Evaluation, The ... Argument, Arguments Coming After the ... Argument **Loop Functions :** Looping on the Command Line, lapply(), sapply(), split(), Splitting a Data Frame, tapply, apply(), Col/Row Sums and Means, Other Ways to Apply, mapply() , Vectorizing a Function	08
IV	**Descriptive Statistics**: Basic Arithmetic Operations, **Standard Functions** like abs(), sqrt(), round(), sum(), product(), log(), log10(), **Statistical Functions** like min(), max(), range(), mean(), quantile (), summary(), var(), sd(), scale(), boxplot(), cov(), cor()	08

	Frequency Measures and Graphical Presentation frequency distribution and cumulative frequency distribution tables, Bar Chart, Pie Chart, Histogram, Box-Whisker Plot, Scatterplots, Matrix of Plots **Simulation:** Generating Random Numbers, Setting the random number seed, Simulating Random Sampling, R function for solution of Binomial, Poisson, Normal and Exponential distribution problems **Hypothesis Testing**: Testing Means (Single mean and Two Means)	
V	**Predictive Analytics:** Types of Analytics, Analytics in Decision Making, Introduction to Predictive Analytics. **Simple Linear Regression (SLR):** Simple Linear Regression: Overview, Model Development, Assumptions, Model Validation, Model fitness and R^2, Example of SLR. **Multiple Linear Regression and Logistic Regression:** MLR: Introduction, Estimation of Regression Parameters, Explanatory vs. Predictive Modeling, Assumptions and Model Diagnostics, MLR with categorical predictors (dummy variable), Derived & Interaction Variables, Multi-collinearity, Adjusted R^2, Model Deployment, Example of MLR.	12

Evaluation:

The students will be evaluated on a continuous basis and broadly follow the scheme given below:

		Weight age
A	Continuous Evaluation Component (Assignments / Quizzes /Class Participation etc.)	20% (C.E.C.)
B	Internal Assessment (Lab based Practical Examination using R-software)	30% (I.A.)
C	End-Semester Practical Examination	50% (Practical/VIVA)

SUGGESTED READINGS:

Text Books:

Sr. No	Author/s	Name of the Book	Publisher	Edition & Year
T-01	Roger D. Peng	R Programming for Data Science	Lean Publishing	1st edition, 2015
T-02	Nicholas J Horton	Using R and RStudio for Data Management, Statistical Analysis and Graphs	CRC Press – T&F Group	2015
T-03	Christian Heumann, Michael Schomaker, Shalabh Sinha	Introduction to Statistics and Data Analytics: With Exercise, Solutions and Applications in R	Springer	2016

Reference Books:

Sr. No	Author/s	Name of the Book	Publisher	Edition & Year
R-01	Roger D. Peng	Exploratory Data Analysis with R	Lean Publishing	1st Edition, 2015
R-02	Alain F Zuur, Elena Leno	A Beginner's Guide to R	Springer (Use R!)	1st Edition 2009
R-03	A. Ohri	R for Business Analytics	Springer	1st Edition, 2012
R-04	**Seema Acharya**	**Data Analytics Using R**	**McGraw Hill**	**1st edition, 2018**