

	DPG Institute of Technology and Management Sector 34, Gurugram HR 122004
	Lesson Plan
	Course Name: DATA SCIENCE FUNDAMENTALS
	Faculty Name: RICHA NEHRA

No. of Lecture Hours/Week	3	Exam Hours	3
Total No. of Lecture Hours	35	Exam Marks	75
Course Code:	PCC-DS-301G	Semester	5th

Course Objectives:

1. To understand the Data science tasks, methods and process
2. To understand the Data Analytics lifecycle
3. To apply hypotheses and data into actionable predictions
4. To effectively document and communicate the findings using visualization techniques

Lecture No.	Topics to be covered	Teaching Methodology	Class Activity/ Event	Remark /CO
UNIT -1				
1	Introduction to Data Science, Definition, Data Scientist roles, Real-world examples	Chalk &Talk	PPT making	CO1
2	Data Science Project Lifecycle	Chalk &Talk		
3	Working with Data Files	Chalk &Talk		
4	Structured & Other Data Formats	PPT/Chalk &Talk		
5	Working with Databases, NoSQL Databases	Chalk &Talk		
6	Data Staging & Curation	Chalk &Talk		
7	Exploring & Cleaning Data	Chalk &Talk	Data Cleaning Mini Task	
8	Sampling & Data Splits	Chalk &Talk		
9	Revision of unit-1	Chalk &Talk		
UNIT -2				
10	Overview of Analytics Projects, Business Problem Framing	Chalk &Talk		CO2
11	Identifying Data Sources, External APIs, internal DBs, web scraping overview	Chalk &Talk		
12	Data Preparation & Understanding, Conditioning, normalization, encoding	Chalk &Talk		
13	Model Planning, Feature selection, exploratory analysis, correlation	Chalk &Talk	Assignment -1	

14	Model Selection & Building, Regression, classification, clustering overview	Chalk &Talk		
15	Communicating Results, Dashboarding, visualization, report structure	Chalk &Talk		
16	Operationalizing Models, Deployment workflow & analytics plan components	Chalk &Talk		
17	Revision of unit-2	Discussion		
UNIT -3				
18	Introduction to R Interface, Data Import/Export in R	Chalk &Talk		CO3
19	R Data Types & Structures	PPT		
20	Arrays, Matrices & Contingency Tables	PPT		
21	Descriptive Statistics in R, Mean, SD, variance, distribution visualization	SMART BOARD	Assignment -2	
22	Hypothesis Testing Basics,	SMART BOARD		
23	Statistical Tests in R	Chalk &Talk		CO3
24	Evaluating Models	Chalk &Talk		CO3
25	Revision of unit-1, unit-3 for sessional1	Discussion		CO1, CO3
UNIT -4				
26	Overfitting & Model Soundness, Bias-variance, validation strategies	Chalk &Talk		CO4
27	Linear Regression, Logistic Regression	SMART BOARD		
28	Unsupervised Learning, k-Means, hierarchical clustering in R	Chalk &Talk		
29	Model Documentation, Model summary report formatting	Chalk &Talk		
31	Model Deployment in R, knitr, RMarkdown, R HTTP services	PPT		
32	Visualization Techniques, Single & multiple variable visualization	Chalk &Talk		
33	Visualizing Complex Data, Boxplot, Scatter matrix, Dot/Bar/Hexbin plots in R	Chalk &Talk		
34	Revision of unit-3, unit-4	Discussion		CO3, CO4
35				

Assessment Methods: -

S.No.	Evaluation Component	Assessment Method	Marks
1	Internal Marks		25
		Attendance	5

2		Quiz/Presentation	5
3		Assignment	5
4		Avg of Sessional 1&2	10
5	External Marks	Final University Exam	75

Reference Books:

1. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big Data Analytics" , EMC Education Services,2015
2. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014
3. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2014
4. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc, 2012
5. W.N. Venables, D.M. Smith and the R Core Team, "An Introduction to R", 2013
6. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, "Practical Data Science Cookbook", Packt Publishing Ltd., 2014

Course Outcomes:

At the end of the course, the student will be able:

CO 1	Able to convert data into actionable insights
CO 2	Analyze and validate the models using appropriate performance metrics
CO 3	Build clustering and classification models using R environment
CO 4	Present the results using effective visualization techniques