

13. INTRODUCTION TO DATA SCIENCE

Discipline	COMPUTER SCIENCE				
Course Code	UK1MDCCSC103				
Course Title	INTRODUCTION TO DATA SCIENCE				
Type of Course	MDC				
Semester	I				
Academic Level	1				
Course Details	Credit	Lecture per week	Tutorial per week	Practical per week	Total Hours/Week
	3	2 hours	-	2 hours	4 hours
Pre-requisites	NIL				
Course Summary	This course aims to introduce the main concepts of data science, understand the essential principles and to implement spreadsheet-based data analysis. Through a blend of theoretical understanding and hands-on practice, learners will develop a solid foundation in data preprocessing, data integration, data transformation, data reduction and skills to apply statistical analysis techniques using Spreadsheet.				

Detailed Syllabus:

Module	Unit	Content	Hrs (L+P)
I	Fundamentals of Data Science		12
	1	Introduction, Why Data Science, Types of Data analysis: Descriptive analysis, Diagnostic analysis, Predictive analysis and Prescriptive analysis.	
	2	Data Analytics life cycle: Data discovery, Data Preparation, Model planning, Model Building, Communicate Results, and Operationalization.	
	3	Data Science tools: Python programming, R programming, SAS, Spreadsheet, Tableau Public, RapidMiner, Knime, Apache Spark.	
	4	Fundamental areas of study in data science: Machine Learning, Deep Learning, NLP, Statistical data analysis, Knowledge discovery and data	

		mining, Text mining, Recommender systems, Data visualization, Computer Vision, and Spatial data management.	
	5	Role of SQL in data science, Pros and Cons of data science	
II	Data Pre-processing		12
	6	Introduction, data types and forms, possible data error types,	
	7	Various data pre -processing operations: Data Cleaning: Filling missing values, Smoothing noisy data, Detecting and removing outliers.	
	8	Data Integration: Virtual integration, physical data integration, Application based integration, Manual Integration, and middleware data integration.	
	9	Data Transformation: Rescaling data, Normalizing data, Binarizing data, Standardizing data.	
	10	Data Reduction: Dimensionality reduction, Data cube aggregation, Numerosity reduction. Data Discretization: Top-down discretization, Bottom-up discretization.	
III	Data Analysis with Worksheet		12
	11	Introduction to Worksheet: Creation and Formatting.	
	12	Ranges and Tables-Data Cleaning with Text Functions, Containing Date Values and Containing Time Values	
	13	Conditional Formatting, Sorting and Filtering	
	14	Subtotals with Ranges, Creating Macros, Pivot Table.	
IV	Data Plotting and Visualization		12
	15	Introduction, Visual encoding, Basic data visualization tools: Histograms, Bar Charts/Graphs, Scatter plots and Area plots. Data visualization types: Temporal data, Hierarchical data, Network data, Multi-dimensional data, Geospatial data and Multivariate data.	
	16	Lookup Functions: LOOKUP and VLOOKUP and HLOOKUP.	
	17	Data Visualization using Band Chart, Thermometer Chart, Gantt chart, Waterfall Chart and Pivot Charts. Types of jobs in data analytics: Data Analyst, Data scientist, Data engineer, Database administrator, Data architect, and Analytics manager.	

V		Flexi Module (Not Included for End Semester Examination)	12
	18	Advanced data visualization tools	
	19	Visualization of geospatial data	
	20	Statistical Data Analysis : Probability theory	

REFERENCES

Core

1. Gypsy Nandi and Rupam Kumar Sharma, Data Science fundamentals and practical approaches, First Edition, BPB Publication, 2020 .
2. Bernd Held, Excel Functions and Formulas, BPB Publications.

Additional

1. V K Jain, Data Science and Analytics, Khanna Publishing.
2. Joel Grus, Data Science From Scratch, Second Edition, Oreilly.

LAB EXERCISES

PART A

1. Create a workbook and perform the operations: Selecting range of columns, hiding /show rows and columns and rename the worksheet.
2. Create a workbook with student mark details. Include formulas to calculate total, percentage and grade.
3. Create worksheet with student mark details and perform the following operations
 - i. Find the number of students having a percentage more than 70.
 - ii. Find the number of students having percentage between 60 and 80.
 - iii. Find the number of students passed in a subject
 - iv. Find the student who got highest mark in a subject.
4. Create a worksheet with Employee salary details. Find mean, median, mode, standard deviation and variance.
5. Create a workbook with sales details and use the functions: TRIM and CLEAN.
6. Create a worksheet with student mark details. Use sorting and filtering functions.
7. Create a worksheet with employee details. Use date and time values. Calculate salary details and bonus using functions.
8. Create a worksheet with the student name as a column. Add three more columns: First name, Last name and e-mail. Find the values of First name, Last name and e-mail(Firstname_lastname@gmail.com). Use text functions.
9. Enter your date of birth and today's date in two cells. Find your age in days, months and years.
10. Prepare a worksheet with sales details. Make a pivot table having product and category in row label.

PART B

11. Create a worksheet for the flower shop with invoice_id, flower name, price, qty and total price. Enter 10 records. Make pivot table and pivot charts.
12. Create a worksheet with Fruits supply details. Apply LOOKUP, VLOOKUP and HLOOKUP functions.
13. Assign a macro to a command button to display “welcome” in a cell.
14. Assign a macro to a command button to display “welcome” in a message box.
15. Assign a macro to a command button to find the total number of sheets in a workbook.
16. Assign a macro to a command button to add a new worksheet.
17. Assign a macro to a command button to add a new workbook.
18. Prepare a worksheet with wildlife populations of different states in India. Display in Pie chart and Bar chart.
19. Prepare a worksheet with the total number of primary schools in each district of kerala. Include different charts.
20. Create a worksheet with employee salary details. Include charts.

Course Outcomes

No.	Upon completion of the course the graduate will be able to	Cognitive Level	PSO addressed
CO-1	Discuss about the fundamentals of Data Science	U	PSO -1
CO-2	Illustrate the usage of Data Pre-processing techniques	Ap	PSO-1, 3
CO-3	Use data science concepts in real world problems	An	PSO-1,2,3
CO-4	Build Data Analytics and management Skill	Ap	PSO-1,2,3

R-Remember, U-Understand, Ap-Apply, An-Analyse, E-Evaluate, C-Create

Note: 1 or 2 COs/module

Name of the Course: INTRODUCTION TO DATA SCIENCE

Credits: 2:0:1 (Lecture: Tutorial: Practical)

CO No.	CO	PO/PSO	Cognitive Level	Knowledge Category	Lecture (L)/Tutorial (T)	Practical (P)
CO-1	Discuss about the fundamentals of Data Science	PO- 6,7 PSO-1,2	U	F, C	L	-
CO-2	Illustrate the usage of Data Pre-processing	PO-6,7 PSO-1,2,3	Ap	F, C, P	L	P

	techniques					
CO-3	Use data science concepts in real world problems	PO-1, 2, 6,7 PSO-1,2,3	An	F, C, P	L	P
CO-4	Build Data Analytics and management Skill	PO-1, 2, 6,7 PSO-1,2,3	Ap	F, C, P	L	P

F-Factual, C- Conceptual, P-Procedural, M-Metacognitive

Mapping of COs with PSOs and POs :

	PO 1	PO2	PO 3	PO 4	PO 5	PO 6	PO 7	PO8	PSO 1	PSO 2	PSO3	PSO4
CO 1	-	-	-	-	-	1	1	-	1	-	-	-
CO 2	-	-	-	-	-	1	2	-	2	2	2	-
CO 3	1	2	-	-	-	1	2	-	2	2	2	-
CO 4	2	2	-	-	-	1	2	-	2	2	2	-

Correlation Levels:

Level	Correlation
-	Nil
1	Slightly / Low
2	Moderate / Medium
3	Substantial / High