

<b>Semester: 1</b>				
<b>Programme: Data Science</b>				
<b>Course: Introduction to Data Science</b>				
<b>Paper code: C1DS250112T/C1DS250112P</b>				<b>Credits: 4</b>
<b>Hours/week: 5 (3Th + 2Pr)</b>				
<b>Category: Core</b>				
<b>Theory / Practical / Composite: Composite</b>				
<b>No of Module: 1</b>				
<b>Course Outcome:</b>				
1. <b>Remember</b> the fundamental concepts of data, including the definitions of population and sample, scales of measurement, the distinction between structured and unstructured data, and the notion of Big Data.				
2. <b>Understand</b> the conceptual framework of exploratory data analysis (EDA) and the functional importance of simple correlation and regression in addressing real-world scenarios.				
3. <b>Apply</b> descriptive measures—such as central tendency, dispersion, skewness, and kurtosis—along with basic EDA tools like summary tables and graphs to explore datasets.				
4. <b>Analyze</b> the features and patterns of data through scatter plots and contingency tables while identifying the impact of missing values and outliers on data quality.				
5. <b>Evaluate</b> the effectiveness and appropriateness of using specific correlation, regression, and association measures for solving complex real-world problems.				
6. <b>Create</b> structured data science problem formulations and comprehensive data stories by integrating techniques for data wrangling, abstraction, and visualization.				
<b>Prerequisites: Basic knowledge about any prior course</b>				
<b>SYLLABUS</b>				
<b>UNIT</b>	<b>CONTENT</b>	<b>HOURS or NUMBER OF CLASSES</b>	<b>CO Mapping</b>	<b>COGNITIVE LEVEL</b>
<b>1.</b>	<b>Basics of Data:</b> Concept of population and sample. Classification of data, scales of measurement, structured and unstructured data, data sources. Data wrangling and data abstraction, storytelling through data, converting a real-world problem to a data science problem. Notion of big data	10	CO1, CO6	K1, K6
<b>2.</b>	<b>Summary Statistics:</b> Moments and quantiles. Concept and measures of Central Tendency, Dispersion, Skewness and Kurtosis. Scatter plot. Simple correlation and Regression and their importance with respect to real world problems. Contingency tables and measures of association for nominal and ordinal data. Idea of missing values and outliers.	20	CO2, CO3, CO4, CO6	K2, K3, K4, K6

<b>3.</b>	<b>Exploratory data analysis (EDA):</b> Philosophy of EDA, basic tools for EDA, Summary tables and Graphs. Case studies.	9	CO2, CO3, CO6	K2, K3, K6
<b>Text Books</b>				
1. Jeffrey M. Stanton (2013). Introduction to Data Science				
2. Hogg, R.V. and Tanis, E.A. (2009): A Brief Course in Mathematical Statistics. Pearson Education.				
3. Peng, R.D. and Matsui, E., (2015): The Art of Data Science: A guide for anyone who works with Data. Skybrude Consulting, LLC.				
4. Moore, D.S., McCabe, G.P. and Craig, B.A., (2014): Introduction to the Practice of Statistics. Macmillan Higher Education.				
<b>Evaluation</b>	Theory CIA: 15 Sem Exam: 45	Practical CA: 40 Sem Exam: NA		
<b>Paper Structure for Theory Semester Exam Module:</b>	Short questions (5 marks each)	Long questions (15 marks each)		
	3 out of 5	2 out of 3		

### Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
<b>CO1</b>	<b>Remember</b> the fundamental concepts of data, including the definitions of population and sample, scales of measurement, the distinction between structured and unstructured data, and the notion of Big Data.	K1
<b>CO2</b>	<b>Understand</b> the conceptual framework of exploratory data analysis (EDA) and the functional importance of simple correlation and regression in addressing real-world scenarios.	K2
<b>CO3</b>	<b>Apply</b> descriptive measures—such as central tendency, dispersion, skewness, and kurtosis—along with basic EDA tools like summary tables and graphs to explore datasets.	K3
<b>CO4</b>	<b>Analyze</b> the features and patterns of data through scatter plots and contingency tables while identifying the impact of missing values and outliers on data quality.	K4
<b>CO5</b>	<b>Evaluate</b> the effectiveness and appropriateness of using specific correlation, regression, and association measures for solving complex real-world problems.	K5
<b>CO6</b>	<b>Create</b> structured data science problem formulations and comprehensive data stories by integrating techniques for data wrangling, abstraction, and visualization.	K6