

	<b>Program</b>	<b>B. Tech Computer Science and Engineering</b>					<b>Program Code</b>	<b>24BTHCSE</b>
	<b>Course</b>	<b>Fundamentals of Data Science</b>					<b>Course Code</b>	<b>24BTELY107</b>
	<b>Semester</b>	<b>I</b>	<b>Credits</b>	<b>3</b>	<b>Theory</b> 3	<b>Practical</b> 0	<b>Total Hours</b>	<b>40</b>
<b>COURSE OBJECTIVES:</b> <ol style="list-style-type: none"> <li>1) Describe the significance of data science and understand the Data Science process.</li> <li>2) Understanding how data is collected, managed and stored in data warehouse.</li> <li>3) To learn how to analyse data sets for use with a variety of statistical methods and models.</li> <li>4) To Analyse data using various Visualization techniques.</li> <li>5) Learn the concepts of machine learning models, map reduce technique and social network graphs.</li> </ol>								
<b>General Instructions for Teaching-Learning:</b> <ol style="list-style-type: none"> <li>1) Appropriate Models PowerPoint presentation, Charts and Videos shall be used to enhance the learning process.</li> <li>2) Flipped Learning can be used.</li> <li>3) Simple case studies can be selected by the teacher to enhance teaching process.</li> <li>4) Use some tutorials and suggest to learn various visualization tools.</li> <li>5) Refer tutorials and concepts from web sources.</li> </ol>								
<b>Module</b>	<b>Topics</b>							<b>Hours</b>
I	<b>Introduction To Data Science</b>							8
	Definition—Big Data and Data Science Hype—Datafication—Data Science Profile—Meta Data—Definition—Data Scientist—Statistical Inference—Populations and Samples—Populations and Samples of Big Data—Modelling—Data Warehouse—Philosophy of Exploratory Data Analysis—The Data Science Process—A Data Scientist's Role in this Process Case Study: Real Direct—Housing Market Analysis							
II	<b>Mathematical Preliminaries</b>							8
	Probability—Descriptive Statistics—Correlation Analysis and Regression. <b>Data Munging:</b> Properties of Data—Collecting Data—Cleaning Data—Crowdsourcing.							
III	<b>Scores—Ranking and Statistical Analysis</b>							8
	Scores and Rankings: Developing Scoring Systems—Z-scores and Normalization Statistical Analysis: Sampling from Distributions—Statistical Distributions—Statistical Significance—Permutation Tests and P-values							
IV	<b>Data Visualization and Mathematical Models</b>							8
	Data Visualization: Basic principles—ideas and tools for data visualization Visualizing Data: Exploratory Data Analysis—Developing a Visualization Aesthetic—Chart Types Mathematical Models: Philosophies of Modelling—A Taxonomy of Models—Baseline Models in ML—Evaluating Models							
V	<b>Social Network Graphs</b>							8
	Data Engineering—Map reduce—Word Frequency Problem—Map Reduce Solution with Example Social Network Graphs: Social networks as graphs—Clustering of graphs—Partitioning of graphs							
<b>COURSE OUTCOMES:</b>								
<b>At the end of Course Students will be,</b>								
<b>CO1:</b> Able to understand Data scientist role, domains, big data and Data Science Process.								

**CO2:** Able to understand the data cleaning and data mining.  
**CO3:** Able to understand statistical analysis.  
**CO4:** Able to understand the exploratory data analysis and Data visualization technique.  
**CO5:** Able to understand map reduce technique and Social Network graphs.

#### **Text Books:**

1. Steven S. Skiena, "The Data Science Design Manual", Springer 2017.
2. Rachel Schutt & O'neil, "Doing Data Science", Straight Talk from The Frontline O'REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013.
3. Peter Bruce, Andrew Bruce, and Peter Gedeck, "Practical Statistics for Data Scientists", 2<sup>nd</sup> edition, 2020

#### **Reference Books:**

1. Joel Grus," Data Science from Scratch" First Edition, April 2015
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani , "An Introduction to Statistical Learning-with Applications in R", 2013
3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press, 2<sup>nd</sup> edition, 2014
4. R Programming for Data Science, Roger D. Peng, LeanPub, 2015.

#### **E-Learning Sources:**

1. Introduction to Data Science <https://www.geeksforgeeks.org/introduction-to-data-science/>
2. "Data science for beginners", edureka - <https://youtu.be/-ETQ97mXXF0?si=WvQ3GsTg7xFkE8Gd>
3. "Data science for beginners" full course-2023 by Simplilearn - <https://www.youtube.com/watch?v=SJuR41tIE9k>