Course Code	21CSS303T	Course Name	DATA SCIENCE	Course Category	S	ENGINEERING SCIENCES	L	T	P	C
							2	0	0	2

Pre-requisite Courses	*Nil*	Co- requisite Courses	*Nil*	Progressive Courses	*Nil*
Course Offering Department	*Data Science and Business Systems*	*Data Book / Codes / Standards*		*Nil*	

Course Learning Rationale (CLR):	***The purpose of learning this course is to:***
CLR-1:	*understand the basics of data*
CLR-2:	*learn the Pandas library to analyze data frames*
CLR-3:	*utilize different methods of data acquisition and data cleaning*
CLR-4:	*explore the visualization tools for different kinds of input data formats*
CLR-5:	*apply supervised and unsupervised learning to learn the hidden patterns from the data and predict the output*

Course Outcomes (CO):	***At the end of this course, learners will be able to:***	**Program Outcomes (PO)**												**Program Specific Outcomes**		
		1	2	3	4	5	6	7	8	9	10	11	12			
		Engineering Knowledge	Problem Analysis	Design/development of solutions	Conduct investigations of complex problems	Modern Tool Usage	The engineer and society	Environment & Sustainability	Ethics	Individual & Team Work	Communication	Project Mgt. & Finance	Life Long Learning	PSO-1	PSO-2	PSO-3
CO-1:	*understand the relationship between data*	-	-	-	-	*1*	-	-	-	-	-	-	-	-	-	-
CO-2:	*identify the different data structures to represent data*	-	-	-	-	*1*	-	-	-	-	-	-	-	-	-	-
CO-3:	*identify data manipulation and cleaning techniques using pandas*	-	-	-	-	*1*	-	-	-	-	-	-	-	-	-	-
CO-4:	*constructs the Graphs and plots to represent the data using python packages*	-	-	-	-	*1*	-	-	-	-	-	-	-	-	-	-
CO-5:	*apply the principles of the data science techniques to predict and forecast the outcome of real-world problem*	-	-	-	-	*1*	-	-	-	-	-	-	-	-	-	-

Unit-1 - Introduction to Data Science, Numpy and Pandas	***10 Hour***
Introduction to Data science: Facets of data, Data Science Process Introduction to Numpy: Numpy, creating array, attributes, Numpy Arrays objects: Creating Arrays, basic operations (Array Join, split, search, sort), Indexing, Slicing and iterating, copying arrays, Arrays shape manipulation, Identity array, eye function Pandas: Exploring Data using Series, Exploring Data using DataFrames, Index objects, Re index, Drop Entry, Selecting Entries, Data Alignment, Rank and Sort, Summary Statistics, Index Hierarchy Data Acquisition: Gather information from different sources, Web APIs, Open Data Sources, Web Scrapping.	
Unit-2 - Data Wrangling, Data Cleaning and Preparation	***10 Hour***
Data Handling: Problem faced when handling large data-General techniques for handling large volume of data- General programming tips for dealing large data sets Data Wrangling: Clean, Transform, Merge, Reshape: Combining and Merging Datasets, Merging on Index, Concatenate, Combining with overlap, Reshaping, Pivoting Data Cleaning and Preparation: Handling Missing Data, Data Transformation, String Manipulation, summarizing, Binning, classing and Standardization, outlier/Noise& Anomalies.	
Unit-3 - Visualization	***10 Hour***
Customizing Plots: Introduction to Matplotlib, Plots, making subplots, controlling axes, Ticks, Labels and legends, annotations and drawing on subplots, saving plots to files, matplotlib configuration using different plot styles, Seaborn library. Making sense of data through advanced visualization: Controlling line properties of chart, creating multiple plots, Scatter plot, Line plot, bar plot, Histogram, Box plot, Pair plot, playing with text, styling your plot, 3d plot of surface	

Learning Resources	1. *Grus, J. (2019). Data Science from Scratch, 2nd Edition. O'Reilly Media, Inc.* 2. *Jiawei Han, Micheline Kamber and Jian Pei (2012), Data Mining Concepts and Techniques, Third Edition, Elsevier.* 3. *Davy Cielen, Arno D. B. Meysman, and Mohamed Ali (2016), Introducing Data Science: Big data, machine learning, and more, using Python tools, Manning Publications.* 4. *McKinney, W. (2018). Python for data analysis: Data wrangling with pandas, NumPy, and IPython. O'Reilly Media, Inc.*	5. *Vanderplas, J. T. (2017). Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc.* 6. *Jeffrey S. Saltz and Jeffrey M. Stanton (2018), An Introduction to Data Science, Sage Publication.* 7. *Shai Vaingast (2014), "Beginning Python Visualization Crafting Visual Transformation Scripts", Second Edition, Apress.* 8. *Wes Mc Kinney (2012). "Python for Data Analysis", O'Reilly Media.*

Learning Assessment

	Bloom's Level of Thinking	*Continuous Learning Assessment (CLA)*				*Summative Final Examination (40% weightage)*	
		Formative CLA-1 Average of unit test (50%)		*Life-Long Learning CLA-2 (10%)*			
		Theory	*Practice*	*Theory*	*Practice*	*Theory*	*Practice*
Level 1	*Remember*	*40%*	-	*20%*	-	*40%*	-
Level 2	*Understand*	*40%*	-	*20%*	-	*40%*	-
Level 3	*Apply*	*10%*	-	*20%*	-	*10%*	-
Level 4	*Analyze*	*10%*	-	*20%*	-	*10%*	-
Level 5	*Evaluate*	-	-	*10%*	-	-	-
Level 6	*Create*	-	-	*10%*	-	-	-
	Total	*100 %*		*100 %*		*100 %*	

Course Designers

Experts from Industry	**Experts from Higher Technical Institutions**	**Internal Experts**
1. *Dr. Veeramanickam. M.R.M, Associate Professor Chitkara University Institute of Engineering and Technology*	1. *Mr. Snehith Allam Raju Senior Manager Advanced Analytics & Architecture Envista Holdings Corporation, Hyderabad.*	1. *Dr.V.Kalpana, SRMIST*
		2. *Dr.G.Vadivu, SRMIST*