FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: **Bachelor of Technology (Information Technology)**

Semester: **VII**

Course Code: **202046707**

Course Title: **Data Science and Visualization**

Course Group: **Professional Core Course**

Course Objectives: Purpose of this course is to introduce data science techniques to enhance the decision-making strategies through analysis, interpretation, and management of data. It also aims to provide overview of different data visualization techniques for efficient and effective data representation.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
				Theory		J/V/P*		Total
Lecture	Tutorial	Practical		Internal	External	Internal	External	
3	0	2	4	50/18	50/17	25/9	25/9	150/53

* **J**: Jury; **V**: Viva; **P**: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	**Introduction to Business Analytics:** Why Analytics, Business Analytics: The Science of Data-Driven Decision Making, Descriptive Analytics, Predictive Analytics, Prescriptive Analytics, Big Data Analytics, Web and Social Media Analytics.	02
2	**Descriptive Analytics:** Introduction to Descriptive Analytics, Data Types, and Scales, Types of Data Measurement Scales, Percentile, Decile and Quartile, Measures of Variation, Measures of Shape – Skewness and Kurtosis.	04

3	**Introduction to Probability:** Introduction to Probability Theory, Probability Theory – Terminology, Fundamental Concepts in Probability – Axioms of Probability, Application of Simple Probability, Bayes' Theorem, Random Variables, Probability Mass Function (PMF) and Cumulative Distribution Function (CDF) of a Discrete Random Variable, Binomial Distribution, Poisson Distribution, Geometric Distribution, Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of a Continuous Random Variable, Uniform Distribution, Exponential Distribution, Chi-Square Distribution, Student's t-Distribution, F-Distribution.	07
4	**Sampling and Estimation:** Introduction to Sampling, Population Parameters and Sample Statistic, Sampling, Probabilistic Sampling, Non-Probabilistic Sampling, Sampling Distribution, Central Limit Theorem (CLT), Sample Size Estimation for Mean of the Population, Estimation of Population Parameters, Estimation of Parameters Using Method of Moments, Estimation of Parameters Using Maximum Likelihood Estimation.	07
5	**Introduction to Data Visualization:** Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization (Control of Presentation, Faster and Better JavaScript processing, Rise of HTML5, Lowering the implementation Bar) Exploring the Visual Data Spectrum: charting Primitives (Data Points, Line Charts, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts, Infographics). Making use of HTML5 CANVAS, Integrating SVG	08
6	**Tabular Data Visualization:** Tables: Reading Data from Standard text files (.txt, .csv, XML), Outputting Basic Table Data (Building a table, Using Semantic Table, Configuring the columns), Assuring Maximum readability (Styling your table, Increasing readability, Adding dynamic Highlighting), Including computations, Using data tables library.	07
7	**Visualizing data Programmatically:** Creating HTML5 CANVAS Charts (HTML5 Canvas basics, A Simple Column Chart, Animations), Starting with Google charts (Google Charts API Basics, A Basic bar chart, A basic Pie chart).	05
	Total	40

List of Practicals / Tutorials:

1	Consider dataset with student name, gender, Enrollmentno, 4-semester result with marks of each subject, his mobile number, city. Implement the following in Python (For Practical 1,2) Perform descriptive analysis and identify the data type and implement a method to find out variation in data. For example, the difference between the highest and lowest marks in each subject semester-wise. Plot the graph showing the results of students in each semester.
2	Prepare Software Requirement Gathering for selected project. (Technical, Non-Technical and Functional requirements).

3	Plot the graph showing the geographical location of students, also plot the graph showing number of male and female students and implement a method to treat missing values for gender and missing value for marks.
4	Estimate project cost and prepare project schedule using Gantt chart.
5	Select the appropriate approach to track and monitor your project.
6	Introduction to data visualization setup tools.
7	Introduction to data visualization advanced setup tools.
8	Develop the different basic Graphical Shapes using HTML5 CANVAS .
9	Develop the different Advanced Graphical Shapes using HTML5 CANVAS.
10	Develop the different basic Graphical Shapes using SVG TAG.
11	Develop the different Advanced Graphical Shapes using SVG TAG.
12	Develop the simple bar chart using HTML5 CANVAS.

Reference Books:

1	U.Dinesh Kumar, Business Analytics, Wiley, India
2	V.K. Jain, Data Science & Analytics, Khanna Book Publishing, New Delhi
3	Data Science For Dummies by Lillian Pierson, Jake Porway
4	Jon Raasch, Graham Murray, Vadim Ogievetsky, Joseph Lowery, "JavaScript and jQuery for Data Analysis and Visualization", WROX
5	Ritchie S. King, Visual storytelling with D3" Pearson.
6	Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2008.

Supplementary learning material:

1	www.analyticsvidhya.com/
2	www.kaggle.com/
3	HTML5 (Canvas and SVG tags)

Pedagogy:
- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks in %						R: Remembering; U: Understanding;
R	**U**	**A**	**N**	**E**	**C**	**A**: Applying;
15%	25%	25%	15%	20%	---	**N**: Analyzing; **E**: Evaluating; **C**: Creating

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Identify the data types, the relationship between data, and the visualization techniques for data.	**20**
CO-2	To be able to apply the concept of probability, sampling, distribution, and estimation in solving real time problems.	**30**
CO-3	To be able to implement various data visualization techniques and draw basic and advanced graphics shapes using SVG and Canvas.	**25**
CO-4	Apply appropriate data visualization techniques on multiple data files and draw different charts.	**25**
CO-5	Identify the data types, the relationship between data, and the visualization techniques for data.	**20**

Curriculum Revision:	
Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025