

NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	BTech CSE, Integrated BTech (CSE)-MBA
Course Code:	XXXX
Course Title:	Data Analysis and Visualisation
Course Type:	Department Elective-I
Year of Introduction:	2024-25

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

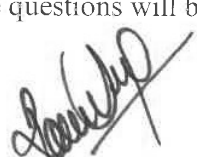
At the end of the course, students will be able to –

1. demonstrate data characteristics using visualisation tools (BL2)
2. identify common data types and corresponding analysis approaches (BL3)
3. analyse the data using various statistical tools (BL4)
4. build data visualisation systems for interdisciplinary problems. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	Introduction: Data Understanding types of data, information, and uncertainty, classes and attributes, interactions among attributes, relative distributions, summary statistics. Data Quality: inaccurate data, sparse data, missing data, insufficient data, imbalanced data	10
Unit-II	Definition, Purpose, Usage, Business Data Visualisation: Features of Business Data, Different Visualisation fields. Forms of Business Data Visualisation. Social Challenges: data ownership, data security, ethics, and privacy	10
Unit-III	The Data: Data Examination, Data Visualisation Patterns, the Categories of Data Visualisation. Data Visualisation: using different tools: refine data and create, edit, alter, and display their visualisations (x-y graph, bar chart, pie chart, cube, etc.)	10
Unit-IV	Data Reduction and Feature Enhancement: standardizing data, sampling data, using principal components to eliminate attributes, limitations, and pitfalls of principal component analysis (PCA), curse of dimensionality	10
Unit-V	Showing Complex Data: Organizational Models, Preattentive Variables, Sorting and Rearranging, Searching and Filtering, Datatips, Data Spotlight, Dynamic Queries, Data Brushing, Local Zooming, Sortable Table, Radial Table, Multi-Y Graphs, Treemap, Small Multiples	05

Self-Study:

The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study content



Suggested Readings/ References:

1. Jack G. Zheng, Data Visualization for Business Intelligence, Taylor and Francis
2. Jiahei Han & Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann
3. Jenifer Tidwell, Designing Interfaces, O'reilly Media, inc.
4. Edward Tufte, The Visual Display of Quantitative Information, Graphics Press LLC.
5. Ben Fry, Visualizing Data, O'reilly Media inc.
6. Noab Iliinsky, Julie Steele, Designing Data Visualization, O'reilly Media inc.
7. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson
8. Wes McKinney, Python for Data Analysis, O'Reilly
9. S. Nagabhushana, Data Warehousing, OLAP, and Data Mining, New Age Publishers

Laboratory Work:

Laboratory work will be based on the above syllabus with a minimum of 10 experiments to be incorporated. The students in a suitable group size will design and perform one experiment as a part of Laboratory work.

Sr. No.	List of Experiments/Exercises	Hours
1-2	(a) Aim: Data Domain selection and Identification of Characteristics of the selected Dataset of different formats. <ol style="list-style-type: none"> 1. What data domain have you selected? 2. What information does the dataset contain? 3. Identify the characteristics of various fields of the dataset. (The distribution, inference, etc.) 4. What insights (knowledge) can we generate for the selected dataset? 5. What patterns are available in the dataset? (b) For the selected dataset, generate a five-number summary using Python. Also generate mode and midrange, outlier detection using the concept of the Quartile method, and others. Compare the results.	06
2	Case Study for Data Visualisation using Tableau. Use the dataset selected in practical-1 and design an interactive Dashboard for analysing data for the selected KPI.	04
3	Data Preprocessing (Data Quality): Aim: Implement data smoothing and data normalization methods. Redundancy analysis using Pearson correlation and Chi-Square. Discretization by Intuitive Partitioning.	04
4	Data Reduction and Feature Enhancement: Aim: Implement Dimensionality reduction and Feature selection techniques with the selected dataset.	04
5-6	Data Analysis and Visualisation Aim: Apply Classification Techniques to selected datasets and visualize the results.	04
7-8	Data Analysis and Visualisation Aim: Apply Clustering Techniques for selected datasets and visualize the results.	04
10	Pattern Analysis and Visualisation Aim: Implementation of various classification and regression techniques and visualize the result with a selected visualization tool like Tableau.	04

