



GUJARAT TECHNOLOGICAL UNIVERSITY

Master of Engineering

Subject Code: 3722319

Semester – II

Subject Name: Data Preparation and Analysis

Type of course: Regular

Prerequisite: Basics of statistics

Rationale: This course will introduce key concepts for the analysis and development of a meaningful Data Visualizations. In summary, this course will provide a broad view for the data gathering, cleaning and visualization techniques

Teaching and Examination Scheme:

Teaching Scheme			Credits C	Examination Marks				Total Marks
L	T	P		Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Total Hrs
1	Data Gathering and Preparation: Data formats, parsing and transformation, Scalability and real-time issues	09
2	Data Cleaning: Consistency checking, Heterogeneous and missing data, Data Transformation and segmentation	11
3	Exploratory Analysis: Descriptive and comparative statistics, Clustering and association, Hypothesis generation	13
4	Visualization: Designing visualizations, Time series, Geolocated data, Correlations and connections, Hierarchies and networks, interactivity	15

Reference Books:

1. Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, Glenn J. Myatt and Wayne P Johnson, Wiley
2. Python: Data Analysis and Visualization, Phuong Vo.T.H et al., Packt Publication
3. Data Points, Visualization that means something, Nathan Ayu, Wiley



GUJARAT TECHNOLOGICAL UNIVERSITY

Master of Engineering

Subject Code: 3722319

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Extract and transform the data into standard format	15%
CO-2	Clean the data before analysis	25%
CO-3	Describe and compare the descriptive statistics of the data	25%
CO-4	Analyze and design the visualization of the data analysis	35%

List of Experiments:

- Minimum 10 experiments based on the above contents.
- Mini Project in a group of max. 3 students
- Writing a research paper on selected topic from content with latest research issues in that topic

Major Equipments:

- Latest PCs with related software

List of Open Source Software/learning website:

- <https://www.kdnuggets.com/tag/data-preparation>
- <https://orange.biolab.si/>