INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal, Hyderabad -500 043

COMPUTER SCIENCE AND ENGINEERING

COURSE DESCRIPTOR

Course Title	**DATA PREPARATION AND ANALYSIS**				
Course Code	BCSB13				
Programme	M.Tech				
	II	CSE			
Course Type	Elective				
Regulation	IARE - R18				
Course Structure	**Theory**			**Practical**	
	Lectures	**Tutorials**	**Credits**	**Laboratory**	**Credits**
	3	0	3	0	0
Chief Coordinator	Dr. K Rajendra Prasad, Professor & HOD, CSE				
Course Faculty	Ms. G Sulakshana				

I. COURSE OVERVIEW:

The course covers the basics of data preparation and data cleaning is an inevitable step in statistical analysis. In business environments, it is frequently required to transfer data from databases and perform statistical analysis. Establish a linkage between data marts and statistical packages is an important task which occurs in professional organizations. This course introduces you to the concepts and the techniques to prepare data located in business intelligent data marts for statistical analysis and covers reading, cleaning, pre-analyzing data and visualization.

II. COURSE PRE-REQUISITES:

Level	Course Code	Semester	Prerequisites
PG	BCSB13	II	Fundamentals of data preparation with a statistical knowledge.

III. MARKS DISTRIBUTION:

Subject	SEE Examination	CIA Examination	Total Marks
Data Preparation and Analysis	70 Marks	30 Marks	100

IV. DELIVERY / INSTRUCTIONAL METHODOLOGIES:

✔	Chalk & Talk	✔	Quiz	✔	Assignments	✔	MOOCs
✔	LCD / PPT	✔	Seminars	✔	Mini Project	✔	Videos
✘	Open Ended Experiments						

IV. EVALUATION METHODOLOGY:

The course will be evaluated for a total of 100 marks, with 30 marks for Continuous Internal Assessment (CIA) and 70 marks for Semester End Examination (SEE). Out of 30 marks allotted for CIA during the semester, marks are awarded by taking average of two CIA examinations or the marks scored in the make-up examination.

Semester End Examination (SEE): The SEE is conducted for 70 marks of 3 hours duration. The syllabus for the theory courses is divided into fiveunits and each unit carries equal weightage in terms of marks distribution. The question paper pattern is as follows. Two full questions with "either" or "choice" will be drawn from each unit. Each question carries 14 marks. There could be a maximum of two sub divisions in a question.

The emphasis on the questions is broadly based on the following criteria:

50 %	To test the objectiveness of the concept.
50 %	To test the analytical skill of the concept OR to test the application skill of the concept.

Continuous Internal Assessment (CIA):
CIA is conducted for a total of 30 marks (Table 1), with 25 marks for Continuous Internal Examination (CIE), 05 marks for Quiz/ Alternative Assessment Tool (AAT).

Table 1: Assessment pattern for CIA

Component	**Theory**		**Total Marks**
Type of Assessment	CIE Exam	Quiz / AAT	
CIA Marks	25	05	30

Continuous Internal Examination (CIE):
Two CIE exams shall be conducted at the end of the 8th and 16th week of the semester respectively. The CIE exam is conducted for 25 marks of 2 hours duration consisting of two parts. Part–A shall have five compulsory questions of one mark each. In part–B, four out of five questions have to be answered where, each question carries 5 marks. Marks are awarded by taking average of marks scored in two CIE exams.

Quiz / Alternative Assessment Tool (AAT):
Two Quiz exams shall be online examination consisting of 25 multiple choice questions and are be answered by choosing the correct answer from a given set of choices (commonly four). Marks shall be awarded considering the average of two quizzes for every course. The AAT may include seminars, assignments, term paper, open ended experiments, five minutes video and MOOCs

V. HOW PROGRAM OUTCOMES ARE ASSESSED:

Program Outcomes (POs)		Strength	Proficiency assessed by
PO 1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.	3	Assignments
PO 2	**Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences	3	Assignments
PO 3	**Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.	2	Seminars, Viva
PO 4	**Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	1	5 minutes video
PO 5	**Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.	2	Mini Projects

3 = High; 2 = Medium; 1 = Low

VI. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED:

Program Specific Outcomes (PSOs)		Strength	Proficiency assessed by
PSO 1	**Professional Skills:** The ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, multimedia, web design, big data analytics, and networking for efficient design of computer-based systems of varying complexity.	2	Projects
PSO 2	**Problem-Solving Skills:** The ability to apply standard practices and strategies in software project development using open-ended programming environments to deliver a quality product for business success.	3	Lectures, Assignments
PSO 3	**Successful Career and Entrepreneurship:** The ability to employ modern computer languages, environments, and platforms in creating innovative career paths to be an entrepreneur, and a zest for higher studies.	1	5 minutes video

3 = High; 2 = Medium; 1 = Low

VII. COURSE OBJECTIVES (COs):

The course should enable the students to:	
I	Develop strategies for dealing with imperfect real world data
II	Read data from databases and clean the data for statistical analysis in SAS.
III	Prepare data marts for statistical analysis using SAS software.
IV	Convert the data for analysis and develop meaningful Data Visualizations.

VIII. COURSE LEARNING OUTCOMES (CLOs):

CLO Code	CLO's	At the end of the course, the student will have the ability to:	PO's Mapped	Strength of Mapping
BCSB13.01	CLO 1	Identify and understand the difference between data and information with formats.	PO 1, PO 4	2
BCSB13.02	CLO 2	Gain knowledge to identify the data parsing and transformations.	PO 2, PO 3	3
BCSB13.03	CLO 3	Describe the fundamentals of scalability with a desired real time issues.	PO 2, PO 4 PO 5	2
BCSB13.04	CLO 4	Explain the basic concept of data cleaning for valuable information with a minimum consistency checking.	PO 1, PO 3	3
BCSB13.05	CLO 5	Learn data transformations and segmentation to solve statistical problems.	PO 2,PO 3 PO 5	3
BCSB13.06	CLO 6	Understand statistical exploratory analysis with hypothesis generation.	PO 1, PO 4	2
BCSB13.07	CLO 7	Distinguish Clustering and association and apply them in solving statistical problems.	PO 1,PO 4	2
BCSB13.08	CLO 8	Designing visualizations for exploratory analysis.	PO 2, PO 3	3
BCSB13.09	CLO 9	Understand the concept of correlations and connections for geo located data.	PO 2	3
BCSB13.10	CLO 10	Visualize the basic hierarchies in a network for interactivity.	PO 1, PO 2	3

3 = High; 2 = Medium; 1 = Low

IX. MAPPING COURSE LEARNING OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES

CLOs	Program Outcomes POs												Program Specific Outcomes PSOs		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CLO 1	3			1										3	
CLO 2		3	2										2		1
CLO 3		3		1	2								2		
CLO 4	3			1									2		
CLO 5		3	2		2								2	3	
CLO 6	3			1											1
CLO 7	3			1										3	
CLO 8		3	2										2		
CLO 9		3												3	
CLO 10	3	3											2		1

X. ASSESSMENT METHODOLOGIES–DIRECT

CIE Exams	PO 1, PO2, PO 3, PO 4	SEE Exams	PO 1, PO2, PO 3, PO 5	Assignments	PO 2	Seminars	PO 3
Laboratory Practices	PO 1	Student Viva	PO 3	Mini Project	PO 5	Certification	-
Term Paper	-						

XI. ASSESSMENT METHODOLOGIES-INDIRECT

✔	Early Semester Feedback	✔	End Semester OBE Feedback
✔	Assessment of Mini Projects by Experts		

XII. SYLLABUS

II Semester: CSE

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
BCSB13	Elective	3	0	0	3	30	70	100
Contact Classes: 45	Total Tutorials: Nil	Total Practical Classes: Nil				Total Classes: 45		

OBJECTIVES:

The course should enable the students to:

I. Convert the data for analysis and develop meaningful Data Visualizations

UNIT-I	**DATA GATHERING AND PREPARATION**	**Classes: 09**
Data formats, parsing and transformation, Scalability and real-time issues		
UNIT-II	**DATA CLEANING**	**Classes: 09**
Consistency checking, Heterogeneous and missing data, Data Transformation and segmentation		
UNIT-III	**EXPLORATORY ANALYSIS**	**Classes: 09**
Descriptive and comparative statistics, Clustering and association, Hypothesis generation		
UNIT-IV	**VISUALIZATION -1**	**Classes: 09**
Designing visualizations, Time series, Geo located data, Correlations and connections		
UNIT-V	**VISUALIZATION -2**	**Classes: 09**
Hierarchies and networks, interactivity.		

Text Books:
1. Making sense of Data : A practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt
Web References:
1. http://www.sctie.iitkgp.ernet.in/ 2. http://www.rkala.in/softcomputingvideos.php 3. http://www.sharbani.org/home2/soft-computing-1 4. http://www.myreaders.info/html/soft_computing.html
E-Text Books:
1. https://www.books.google.co.in/books?id=bVbj9nhvHd4C 2. https://www.books.google.co.in/books?id=GrZHPgAACAAJ&dq=1.+J.S.R.Jang,+C.T.Sun+and+E. Mizutani,+Neuro,+Fuzzy+and+Soft+Computing,+PHI,+2004,Pearson+Education.

XIII COURSE PLAN:

The course plan is meant as a guideline. Probably there may be changes.

Lecture No	Topics to be covered	Course Learning Outcomes (CLOs)	Reference
1 – 2	Introduction to Data gathering and preparation from various sources.	CLO 1	T1:1.1-1.2
3 – 4	Difference between Data and Information with data formats.	CLO 1	T1:2.1-2.2
5 – 6	Explain briefly about parsing and transformation techniques.	CLO 2	T1:1.4-1.5
7 – 8	What are different types of scalability in real time issues.	CLO 3	T1:2.1-2.2
9 – 10	Explain the basic technique of data cleaning in mining of actual information.	CLO 3	T1: 2.3-2.6,7
11 – 12	Data cleaning concept with consistency checking.	CLO 6	T1:3.1-3.5
13 – 14	Explain the concept of heterogeneous and missing data	CLO 7	T1: 5.2-5.3
15 – 16	Review the concept of data transformation and segmentation under data cleaning process.	CLO 7	T1: 6.1-6.6
17 – 18	Exploratory statistics analysis – from various collection of data.	CLO 7	T1: 6.7
19 – 20	What are descriptive and comparative statistics with various examples.	CLO 7	T1:6.8
21-22	Explain about clustering and association in statistics analysis.	CLO 8	T1: 11.1-11.5
23-- 24	Hypothesis generation from exploratory statistics analysis.	CLO 8	T1: 4.1-4.5
25	What is visualization?	CLO 8	T1:7
26-27	Explain different visualizations in designing statistical analysis.	CLO 9	T1:1.0

28 – 29	What are various time series in visualization.	CLO 9	T1:10.3-10.5
30 – 31	What are geo located data.	CLO 9	T1:8
32 – 33	What are correlations and connections for data visualization.	CLO 10	T1: 12.3-12.4
34 – 35	Explain briefly about hierarchies and its networks.	CLO 10	T1:12
36 -- 38	How interactivity done for visualization.	CLO 10	T1:2.1-2.2
39 – 40	How data preparation and analysis done through a statistical analysis.	CLO 10	T1:12.4

XIV GAPS IN THE SYLLABUS - TO MEET INDUSTRY / PROFESSION REQUIREMENTS:

S No	Description	Proposed actions	Relevance with POs	Relevance with PSOs
1	Updating latest version and new features of the statistical analysis environment.	Laboratory Sessions	PO 5	PSO2
2	Familiarizing the role of converting data into valuable information.	Assignments / Industrial visits	PO1,PO2	PSO2
3	Familiarizing different areas where data analysis can be used.	Seminars	PO 5	PSO3
4	Solving various statistical analysis through sas software.	Extra Lab Sessions for analysis.	PO2	PSO3

Prepared by:
G. Sulakshana, Assistant Professor

HOD, CSE