INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING
DEFINITIONS AND TERMINOLOGIES

Course Title	**DATA MANAGEMENT AND REPRESENTATION**				
Course Code	ACDC03				
Program	B.Tech				
Semester	IV	CSE (DS)			
Course Type	Professional				
Regulation	UG-20				
Course Structure	Theory			Practical	
	Lecture	Tutorials	Credits	Laboratory	Credits
	3	1	4	-	-
Course Coordinator	Dr M V Krishna Rao, Professor				

COURSE OBJECTIVES:
The students will try to learn:

I	The data fundamentals, data collection, handling and preservation techniques.
II	The treatment of missed values in large data sets.
III	The data presentation and visual exploartion techniques needed before the data analysis

COURSE OUTCOMES:
After successful completion of the course, students should be able to:

CO 1	**Identify** the data importing methods from the data files of various formats. for data prsenetation and further exploration.	Remember
CO 2	**Make use of** imputation techniques for wrangling the data for subsequent data analysis.	Understand
CO 3	**Identify** the reasons of missing and bad data in various forms for applying cleaning techniques	Apply
CO 4	**Examine** different styles of tables and graphs for presenting and visualizing the data.	Analyze
CO 5	**Determine** the principles like clarity, precision and efficiency of data presentation and visualization.	Evaluate

CO 6	**Build** different data visualizations using tablular data or dataframes.	Apply

DEFINITION AND TERMINOLOGY:

S.No	DEFINITION	CO's
	MODULE I	
	PRINCIPLES OF DATA MANAGEMENT	
1	**What is Data?**	CO 1
	Data is raw, unorganized facts that need to be processed.	
2	**What are different forms of Data?**	CO 1
	The different forms of data are 1.Text 2. Numerical 3. Multimedia 4. Models 5. Software 6.Domain-specific 7.Instrument-specific	
3	**Define codebook.**	CO 1
	A codebook describes the contents, structure, and layout of a data collection. A well-documented codebook "contains information intended to be complete and self-explanatory for each variable in a data file.	
4	**What is a variable in data management?**	CO 1
	A variable usually contains data from observations such as a score from a questionnaire, age or sex. There are different types of variables and having their influence differently in a study viz. Independent and dependent variables, Active and attribute variables, Continuous, discrete and categorical variable, Extraneous variables and Demographic variables.	
5	**What are the 5 types of variables?**	CO 1
	There are different types of variables and having their influence differently in a study viz. Independent and dependent variables, Active and attribute variables, Continuous, discrete and categorical variable, Extraneous variables and Demographic variables.	

6	**What is Documentation?** Data documentation aims to describe the collected data to make it easier to use, retrieve and manage. Data documentation takes various forms and describes the data on multiple levels.	CO 1
7	**Define coding.** Coding of data refers to the process of transforming collected information or observations to a set of meaningful, cohesive categories. It is a process of summarizing and re-presenting data in order to provide a systematic account of the recorded or observed phenomenon.	CO 1
8	**What is data cleaning?** Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.	CO 1
9	**What is data screening?** Data screening means checking data for errors and fixing or removing these errors. ... Definition of data screening: The process of inspecting data for errors and correcting them prior to doing data analysis.	CO 1
10	**What are the principles of file management?** The 8 Principles are: Accountability, Transparency, Integrity, Protection, Compliance, Accessibility, Retention and Disposition.	CO 1
MODULE II		
SECONDARY, PRIMARY AND ADMINISTRATIVE DATA		
1	**What is Secondary data?** Secondary data is the data that have been already collected by and readily available from other sources.	CO 2
2	**Describe Conceptual model.** A Conceptual Data Model is a diagram identifying the business concepts (entities) and the relationships between these concepts in order to gain, reflect, and document understanding of the organization's business, from a data perspective. business meaning, along with their relationships.	CO 2
3	**What is a primary data?** It is a term for data collected at source. This type of information is obtained directly from first hand sources by means of surveys, observations and experimentation and not subjected to any processing or manipulation and also called primary data.	CO 2

4	**Define administrative data?** Administrative data is the data that organizations collect about their operations. It includes data for routine operations, and is frequently used to assess how well an organization is achieving its intended goals.	CO 2
5	**What is the data set?** A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set.	CO 2
6	**What is an Internal source data ?** Internal source data can easily be found within the organization such as market record, a sales record, transactions, customer data, accounting resources, etc. The cost and time consumption is less in obtaining internal sources.	CO 2
7	**What is an External source data?** This data which can't be found at internal organizations and can be gained through external third party resources is external source data. The cost and time consumption is more because this contains a huge amount of data. Examples of external sources are Government publications, news publications, Registrar General of India, planning commission, international labor bureau, syndicate services, and other non-governmental publications.	CO 2
8	**What is a data linking?** Data linking is used to bring together information from different sources in order to create a new, richer dataset. This involves identifying and combining information from corresponding records on each of the different source datasets.	CO 2
MODULE III		
MISSING DATA		
1	**What is missing data?** Missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest.	CO 3
2	**What are missing data patterns?** A missing data pattern is said to be univariate if there is only one variable with missing data.	CO 3
3	**What is a form?** A form is a database object that you can use to enter, edit, or display data from a table or a query. You can use forms to control access to data, such as which fields of data are displayed.	CO 3

4	**What is data analysis process?** Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.	CO 3
5	**What are the three patterns of missing data?** Missing data are typically grouped into three categories: 1.Missing completely at random (MCAR). 2.Missing at random (MAR). 3. Missing not at random (MNAR).	CO 3
6	**What is a missing data pattern ?** A missing data pattern is said to be univariate if there is only one variable with missing data. ... A missing data pattern is said to be monotone if the variables Yj can be ordered such that if Yj is missing then all variables Yk with k¿j are also missing. This occurs, for example, in longitudinal studies with drop-out.	CO 3
7	**What are the reasons of a missing data?** Many reasons for missing data are 1. People do not respond to survey (or specific questions in a survey) 2. Species are rare and cannot be found or sampled. 3.The individual dies or drops out before sampling. 4. Some things are easier to measure than others. 5. Data entry errors.	CO 3
MODULE IV		
DATA PRESENTATION		
1	**What is data visualization?** Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.	CO 4
2	**What is clarity data?** The Clarity database is a large subset of data that comes from the PennChart (Epic) application.	CO 5
3	**Define precision.** Precision is the degree to which future measurements or calculations yield the same or similar results — it is a measure of the spread of repeated measurement results and depends only on the distribution of random errors – it gives no indication of how close those results are to the true value.	CO 5
4	**What is efficiency?** Efficiency is defined as the ability to produce something with a minimum amount of effort. An example of efficiency is a reduction in the number of workers needed to make a car. The ratio of the effective or useful output to the total input in any system.	CO 5

5	**What are the types of presenting data?** The three main forms of presentation of data are: 1. Textual presentation. 2. Data tables. 3.Diagrammatic presentation.	CO 4
6	**How many methods there are in presenting data?** Data is initially collected from a given source, whether they are experiments, surveys, or observation, and is presented in one of four methods: Textual Method. The reader acquires information through reading the gathered data. Tabular Method.	CO 4
7	**What is the most effective method of presenting data?** While graphs are effective for presenting large amounts of data, they can be used in place of tables to present small sets of data. A graph format that best presents information must be chosen so that readers and reviewers can easily understand the information.	CO 4
8	**What are data presentation tools?** Data tools include standard charts and graphs, such as a bar chart, block histogram, bubble chart, scatterplot, pie chart, line graph, and so on. Users can also choose to display data as networks of related words and ideas, such as a word tree, tag cloud, or word cloud.	CO 4
MODULE V		
DESIGNING TABLES AND GRAPHICS FOR DATA PRESENTATIONS		
1	**What is a table?** A Table refers to any data which is presented in orderly rows across and/or down the page, often enclosed within borders.	CO 6
2	**What are graphics?** Graphical representation refers to the use of charts and graphs to visually display, analyze, clarify, and interpret numerical data, functions, and other qualitative structures.	CO 6
3	**What are five formats for the visual display of data?** 1.Column Chart. 2.Bar Graph. 3. Stacked Bar Graph. 4.Line Graph. 5.Dual-Axis Chart. 6.Mekko Chart. etc.	CO 6
4	**What are tables used for in a presentation?** A Table refers to any data which is presented in orderly rows across and/or down the page, often enclosed within borders. A Figure refers to any other form of presentation such as a bar or pie chart, a graph, a diagram, a map, a photograph, a line drawing or a sample of material.	CO 6

5	**What is a graphical table**	CO 6
	A graphical table is a summarizing visualization designed to provide a lot of information at one glance. It can be set up to show columns with dynamic items such as sparklines, calculated values, conditional icons, or bullet graphs. One value is shown for each row as specified on the Rows axis	
6	**What is Tabular Presentation of Data?**	CO 6
	It is a table that helps to represent even a large amount of data in an engaging, easy to read, and coordinated manner. The data is arranged in rows and columns. This is one of the most popularly used forms of presentation of data as data tables are simple to prepare and read.	
7	**What is graphic presentation of data?**	CO 6
	Graphical representation refers to the use of charts and graphs to visually display, analyze, clarify, and interpret numerical data, functions, and other qualitative structures.	
8	**hat are the different methods of graphical presentation of data?**	CO 6
	IGenerally, four methods are used to represent a frequency distribution graphically. These are Histogram, Smoothed frequency graph and Ogive or Cumulative frequency graph and pie diagram.	
9	**What is Tabular Presentation of Data?**	CO 6
	It is a table that helps to represent even a large amount of data in an engaging, easy to read, and coordinated manner. The data is arranged in rows and columns. This is one of the most popularly used forms of presentation of data as data tables are simple to prepare and read.	
10	**What is a Dichotomous variables?**	CO 6
	These variables are those that have only two categories, i.e., only two response options. Typical examples of this type of variable are sex (male and female) and presence of skin cancer (yes or no).	

Course Coordinator: **HOD,CSE(DS)**
Dr M V Krishna Rao, Professor