# DATA MANAGEMENT AND REPRESENTATION

**IV Semester: CSE (DS)**

| Course Code | Category | Hours / Week | | | Credits | Maximum Marks | | |
|---|---|---|---|---|---|---|---|---|
| | | **L** | **T** | **P** | **C** | **CIA** | **SEE** | **Total** |
| **ACDC03** | **Core** | 3 | 1 | 0 | 4 | 30 | 70 | 100 |
| **Contact Classes: 45** | **Tutorial Classes: 15** | **Practical Classes: Nil** | | | | **Total Classes: 60** | | |

**Prerequisites: No prerequisites required**

## I.   COURSE OVERVIEW:

The course emphasis on the data management and representation. It is the fundamental course and is interdisciplinary in nature for all engineering applications. The students will understand principles of data management, different kinds of data, missing data, data representation in the form of tables and graphs. This course provides adequate knowledge to mange different kinds of data and visualizes data in different forms.

## II. COURSE OBJECTIVES:

**The students will try to learn:**

| | |
|---|---|
| I | Summarize the fundamental knowledge on data management for data science. |
| II | To provide knowledge on forms of data, Datasets, Missing data and datapresentation. |
| III | Able to document and transfer the results and effectively communicate thefindings using visualization techniques. |
| IV | To interpret and communicate data and results. |

## III. COURSE OUTCOMES:

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

| | | |
|---|---|---|
| CO 1 | **Recall** the basic concepts of data and principles of data management. | Remember |
| CO 2 | **Illustrate** secondary, primary and administrative data sources. | Understand |
| CO 3 | **Identify** the reasons of missing data and various forms of data set. | Apply |
| CO 4 | **Examine** different styles of tables and graphs for presenting the data. | Analyze |
| CO 5 | **Determine** the principles like clarity, precision and efficiency of datapresentation. | Evaluate |
| CO 6 | **Build** different data visualizations using tables and graphics. | Apply |

## IV. SYLLABUS:

**MODULE – I: Principles of Data Management (09)**
Data: Data, Sources of Data, From Concepts to Variables, Forms of Data
Data Management: Codebooks, Documentation, Coding, Data Cleaning and Screening, Naming Conventions, Principles of File Management

**MODULE – II: Secondary, Primary and Administrative Data (09)**
Secondary Data: Types of Secondary Data, Use of Secondary Data, Sources of Secondary Data, Examples of Searching for Downloading and Importing Data, A Simple Test of the Conceptual Model.
Primary and Administrative Data: Principles for Primary Data, Administrative Data and Linking Datasets.

**MODULE – III: Missing Data (09)**
Working with Missing Data: Why Are Missing Data a Problem?, Reasons for Missing Data, Types of Missing Data.
Forms and Patterns of Missing Data, Addressing Missing Data in the Analysis Stage

**MODULE - IV Data Presentation (09)**
Presenting Data, Visual Images, First Principles: Clarity, Precision, and Efficiency, Why Words Are Not Enough, Types of Tables and Graphics, Principles of Data Presentation.

**MODULE - V : Designing Tables and Graphics for Data Presentations (09)**
Tables, Examples of Tables, Graphics, Examples of Graphics.

**IV. TEXT BOOKS:**

1. Dr. John P. Hoffmann, "Principles of Data Management and Presentation", First Edition.

**V. REFERENCE BOOKS:**

1. Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt.

**VI. WEB REFERENCES:**

1. http://www.sctie.iitkgp.ernet.in/
2. https://towardsdatascience.com/data-management-strategy-part-3-f1ba544be2dc
3. https://towardsdatascience.com/how-to-handle-missing-data-8646b18db0d4
4. https://www.oclc.org/research/areas/data-science/data-science-presentations.html