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

COURSE CONTENT

DATA HANDLING AND VISUALIZATION								
III Semester: CSE (DS)								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACDD03	**Core**	3	-	-	3	40	60	100
Contact Classes: 48	**Tutorial Classes: Nil**	**Practical Classes: Nil**				**Total Classes: 48**		
Prerequisite: Nil								

I. COURSE OVERVIEW:
This course introduces the fundamentals of data and its types, collection of data, data preservation and organization techniques. The course also discusses data visualization techniques that are useful for preliminary analysis or exploration of the data. This course is an introductory data science course which lays foundation for further data processing that involves data modelling, analysis and mining in all scientific and engineering domains.

II. COURSE OBJECTIVES:
The students will try to learn:

I. The data fundamentals, data collection and data preservation techniques.
II. The data presentation for preliminary understanding of the data.
III. The data visualization techniques.

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

CO1	Tabulate the data from the CSV, XLS, TXT and JSON files and export the table data to files.
CO2	Make use of imputation techniques for wrangling the data for an authentic data analysis.
CO3	Manipulate the tabular data using pivoting and cross-tabling to generate contingency tables
CO4	Manipulate the tabular data by joining multiple data table.
CO5	Explore the data using the data visualization techniques.
CO6	Analyze the data for outliers to data trimming the data required for an authentic data analysis.

IV. COURSE CONTENT:

MODULE – I: DATA FUNDAMENTALS

Scientific research, Concepts, Measurement of Concepts, Variables, Types of Variables: Independent and dependent, Continuous and Discrete, Numerical and Nonnumerical, Numerical Variables: Interval and Ratio, Nonnumerical Variables: Categorical and Ordinal, Definition of data, Modalities of data, Types of data, Data collection methods, Popular online datasets.

MODULE – II: WORKING WITH DATA FILES USING PANDAS

Introduction to Python's Pandas Library, Pandas dataframes, Operations on dataframes, Importing Data from CSV and XLSX files to dataframes, exporting data from dataframes to CSV/XLSX files, Importing Data from and Exporting data to JSON/XML files.

MODULE – III: DATA CLEANING AND PRE-PROCESSING

Data Cleaning: Need for cleaning the Data, treating of missing values, treating duplicate values, treating the bad data, Data Imputation and methods, Fuzzy Matching and its applications.

Data Preprocessing: Standardizing the data, Normalizing the data, Outliers and their detection, treating the outliers.

MODULE – IV: WORKING WITH DATA TABLES

Contingency Tables, merging data tables, pivoting data tables, grouping data tables, summary tables.

MODULE – V: DATA VISUALIZATION

Data visualization: Charts for comparison- Bar graphs and types, Pie charts, line plots and multiline plots, scatter plots, Q-Q plots. Summary plots: Histograms and density plots, Box plots, Multiline plots. spanning trees, minimal spanning trees.

V.TEXT BOOKS:

1. Claus O. Wilke, "Fundamentals of Data Visualization", April 2019.
2. Robert Hoyt, Robert Muenchen, "Data Preparation and Exploration: Applied to Healthcare Data", PHI Learning Private Ltd., 5th edition, 2012.

VI. REFERENCE BOOKS

1. Editors: Chun-houh Chen, Wolfgang H¨ardle Antony Unwin "Handbook of Data Visualizations", Springer.

VII. ELECTRONIC RESOURCES

1. https://www.programiz.com/python-programming
2. https://towardsdatascience.com
3. https://www.vedantu.com