

FOUNDATIONS OF DATA SCIENCE

I Year B. Tech. I Semester
(CSD)

Course Code:

L T P C

3 0 2 4

Prerequisites: Basic Mathematics

Course Outcomes:

CO1: Describe the significance of data science and understand the Data Science process. (L2)

CO2: Explain data concepts, sources, acquisition methods and common formats. (L2)

CO3: Apply basic data preprocessing techniques. (L3)

CO4: Apply basic visualization methods to perform exploratory data analysis. (L3)

CO5: Apply basic data analysis techniques and tools for data analysis. (L3)

UNIT I

10 Lectures

INTRODUCTION TO DATA SCIENCE:

Definition, Big Data, Datafication, Data Science Profile, Statistical Inference, Modeling, The Data Science Process, A Data Scientist's Role in Data Science Process, Multi-Disciplinary Science, Data Science Tasks, Data Science Objectives, Applications of Data Science. (Textbook1, Reference Book1)

Learning Outcomes: At the end of the unit, student will be able to

1. Understand the basics of data science(L1)
2. Summarize testable predictions for real-time data(L2)
3. Understand Data Scientist's Role in the analysis Process (L2)

UNIT II

10 Lectures

INTRODUCTION TO DATA AND DATA SOURCES:

Data and Information, Importance of Data in Decision Making, Types of Data – Qualitative and Quantitative, Structured and Unstructured Data, Data Sources - Primary and Secondary Data, Data Acquisition, Common Data Formats – CSV, Excel and JSON, Introduction to Datasets and Variables. (Textbook-2)

Learning Outcomes: At the end of the unit, student will be able to

1. Define basic concepts of data and information and identify different types of data. (L1)
2. Describe different data sources and distinguish between primary and secondary data. (L1)
3. Explain basic data acquisition methods and common data formats such as CSV, Excel and JSON. (L2)

UNIT III

10 Lectures

DATA PREPARATION AND CLEANING:

Data Quality, Characteristics of Good-quality Data, Missing Values, Duplicate Records, Inconsistent Data, Basic Data Cleaning Techniques, Data Transformation, Data Aggregation, Data Normalization, Data Integration. (Textbook-3)

Learning Outcomes: At the end of the unit, student will be able to

1. Identify common data quality issues and inconsistent data. (L2)
2. Apply basic techniques to handle missing, duplicate and inconsistent data. (L3)
3. Apply basic data transformation, sorting, filtering and aggregation techniques. (L3)

UNIT IV

10 Lectures

EXPLORATORY DATA ANALYSIS:

Introduction to Exploratory Data Analysis, Objectives and EDA Process, Data Summarization, Descriptive Statistics, Mean, Median and Mode, Range, Variance and Standard Deviation, Quartiles and Percentiles, Frequency Distribution, Univariate Analysis, Histograms, Bar Charts and Box Plots, Correlations. (Textbooks-2 and Reference Book 1)

Learning Outcomes: At the end of the unit, student will be able to

1. Explain the objectives and basic process of Exploratory Data Analysis (EDA). (L2)
2. Calculate and interpret basic descriptive statistics (L3)
3. Apply frequency distributions to summarize and explore data. (L3)

UNIT V

10 Lectures

DATA ANALYSIS AND VISUALIZATION Introduction to Excel for Data Analysis, Work Sheets and Excel Tables, Data Consistency, Building Datasheets, Sorting, Filtering, Data Cleaning Using Excel, Data Validation, Data Formatting, Charts.(Textbook-4)

Learning Outcomes: At the end of the unit, student will be able to

1. Illustrate relationships between variables using basic bivariate analysis techniques. (L3)
2. Apply cross-tabulation, correlation and scatter plots to examine relationships and patterns in data. (L3)
3. Apply Excel formulas, functions, sorting, filtering and conditional formatting for basic data analysis. (L3)

TEXTBOOKS:

1. Jugal K. Kalita, Dhruva K. Bhattacharyya, Swarup Roy “Fundamentals of Data Science Theory and Practice” Academic Press is an imprint of Elsevier, 2024.
2. S Sridhar, M Vijaya Lakshmi, “Machine Learning, Oxford University Press, 2021.
3. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to DataMining, 2nd edition, Pearson education, 2018.
4. Michael Olafusi " Microsoft Excel and Business Data Anaysis", Published by www.urbizedge.com Copyright © 2015

REFERENCE BOOKS:

1. Rachel Schutt&O’neil, “Doing Data Science”, Straight Talk from The Frontline O’REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013
2. S. Gaskin and A. Vargas, GO! Microsoft 365: Excel 2025. Boston, MA, USA: Pearson, 2025.
3. Sanjeev J. Wagh, Manisha S. Bhende, and Anuradha D. Thakare, “Fundamentals of Data Science”, CRC Press 2022
4. Steven Cooper “Data Science from Scratch” 2018.
5. K. Stewart and R. Nordell, “Microsoft Excel 365 Complete: In Practice” New York, NY, USA: McGraw Hill, 2026.
6. Steven S. Skiena, “The Data Science Design Manual”, Springer 2017.

WEB REFERENCES:

1. <https://support.microsoft.com/excel>
2. <https://support.microsoft.com/en-us/office/analyze-data-in-excel-322d75d6-06f2-4b67-a3c4-9c0f6f9b3f3e>

Course Outcomes:

CO1: Describe the significance of data science and understand the Data Science process. (L2)

CO2: Explain data concepts, sources, acquisition methods and common formats. (L2)

CO3: Apply basic data preprocessing techniques. (L3)

CO4: Apply basic visualization methods to perform exploratory data analysis. (L3)

CO5: Apply basic data analysis techniques and tools for data analysis. (L3)

LIST OF PROGRAMS:

(Any 12 programs must be completed)

1. Import CSV/Excel datasets and identify data types, structure, missing values, and basic data characteristics.
2. Clean a dataset by handling missing values, duplicate records, inconsistent values, and incorrect data types.
3. Apply sorting, filtering, aggregation, and basic transformations to prepare an analysis-ready dataset.
4. Calculate mean, median, mode, range, variance, standard deviation, quartiles, and percentiles using Excel.
5. Analyze a dataset using frequency tables, histograms, bar charts, and box plots.
6. Analyze relationships between variables using cross-tabulation, correlation, and scatter plots.
7. Analyze multiple variables to identify patterns, trends, and anomalies using Excel.
8. Apply Excel formulas and functions, including statistical, logical, and lookup functions, to analyze a real-world dataset.
9. Create PivotTables and Pivot Charts to summarize and analyze categorical and numerical data.
10. Create appropriate charts and visual summaries for a given dataset and interpret the results.
11. Write a R program to create and display vectors and data frames
12. Find missing values in data and perform R as Calculator and Explore Built in functions in R. Implement of matrix operations, array and factors in R.
13. Import CSV and Tabular Data Files and Construct data frames in R.
Consider any dataset and construct basic plots like histograms, scatter plots, box plots
14. Consider any dataset and present data in Tables, charts, Visualize Measures of Central Tendency, Variation and Shape for any dataset
15. Illustrate rnorm(), dnorm(), pnorm(), and qnorm() functions in R
16. Perform tests of hypotheses about the mean when the variance is known. Compute the p-value.

LABORATORY SOFTWARE / TOOLS

- Microsoft Excel / Microsoft 365 Excel
- AI-assisted spreadsheet analysis tools
- Power Query
- R Studio

SUGGESTED DATASETS

Student performance data, sales data, customer data, employee data, product data, financial data, healthcare data, transportation data, and other publicly available structured datasets.