

St. Xavier's College (Autonomous), Ahmedabad

Syllabus of Semester – VI of the following departments under Faculty of
Computer Science based on Undergraduate Curriculum Framework to be
implemented from the Academic Year 2025-26.

DEPARTMENT OF COMPUTER SCIENCE

BSC CS

Minor Course: Data Science Using R

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course Title & Code	Credit Distribution of The Course			Eligibility Criteria	Prerequisite(s) of the Course (if any)
	Lecture	Tutorial	Practical / Practice		
Data Science Using R CSMN661C	2		2		Nil

Course Objectives:

The course aims to provide students with a strong foundation in data science concepts and statistical methods using R programming. It enables students to understand the process of data collection, exploratory data analysis, and data visualization while developing practical skills in data manipulation and reporting using R. The course also introduces descriptive statistics, probability distributions, correlation analysis, simple linear regression, and basic time series analysis, empowering students to analyze real-world datasets and derive meaningful insights for informed decision-making.

Learning Outcomes:

After completing this course, students will be able to:

- Understand the fundamentals of data science, data visualization, and the role of statistics in data-driven decision making.
- Develop proficiency in R programming for data manipulation, analysis, visualization, and reporting.
- Apply descriptive statistical techniques and probability concepts to summarize and interpret datasets.
- Build and evaluate simple linear regression models and perform basic time series analysis using R.
- Develop analytical and problem-solving skills for extracting meaningful insights from real-world datasets.

UNIT	TOPICS / SUBTOPICS	HOURS
1	Introduction to Data Science and R programming	15
	Introduction to Data Science and Data Visualization • Meaning and Scope of data science, Role of statistics in data science • Type of data and data collection methods • Basic Charts for data visualization • Multi-dimensional visualization • Visualizing network and geographical data ◦ Exploratory Data Analysis (EDA) Practical Sessions: • Getting Started with R, Basic arithmetic in R, vectors, matrices, Tables, Data frames • Function, Data pipelines, Creating an R Markdown/knitr Document in RStudio, The Markdown Language, Running R Code in Markdown Documents • Data manipulation, Reading and Formatting Datasets • Visualizing Data: Basic graphics, ggplot2 Package	
2	Descriptive Statistics	15
	• Basic Statistical Summaries-Measures of central tendency, dispersion and shape, Measures of association-bivariate correlation: Pearson and Spearman Practical Sessions: • Descriptive Statistics for a single group, Descriptive statistics by groups, Dealing with missing values, Pearson and Spearman correlation	
3	Probability and Distributions	15
	• Theory of probability, addition and multiplication rules, conditional probability and Bayes' theorem, Random variables and probability distributions-discrete and continuous, expectation and variance, Specific probability distributions-Binomial, Poisson, Normal and their properties Practical Sessions: • Random sampling, Probability calculations and combinatorics, Discrete distributions, Continuous distributions, The built-in distributions in R	
4	Simple Linear Regression and Basic Time Series Analysis	15
	• Simple linear regression model, principle of least squares, goodness-of-fit of the model • Components of time series, Decomposition, Moving average and exponential smoothing Practical Sessions: • Simple linear regression, Time series analysis, decomposition, Moving and exponential smoothing Suggestive Readings: • Thomas Mailund (2017) Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist, Apress. • Peter Dalggaard (2008) Introductory Statistics with R, Springer • Cole, N.K., Storytelling with Data: A Visualization Guide for Business Professionals, Wiley. • Provost, F. and Fawcett, T., Data Science for Business, O'Reilly Media, Inc.	

- Bruce, P. and Bruce, A., Practical Statistics for Data Scientists: 50 Essential Concepts, O'Reilly Media, Inc.