Anekant Education Society's

Tuljaram Chaturchand College

of Arts, Science and Commerce, Baramati

(Empowered Autonomous)

M.Sc. Degree Program in Data Science
(Faculty of Science & Technology)

CBCS Syllabus

M.Sc. (Data Science) Part – I Semester – I

For Department of Statistics

Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati

Choice Based Credit System Syllabus (2026 Pattern)

(As Per NEP 2020)

To be implemented from Academic Year 2026-27

Title of the Programme: M.Sc. (Data Science) Part – I

Preamble

AES's Tuljaram Chaturchand College has made the decision to change the syllabus of across various faculties from June, 2023 by incorporating the guidelines and provisions outlined in the National Education Policy (NEP), 2020. The NEP envisions making education more holistic and effective and to lay emphasis on the integration of general (academic) education, vocational education and experiential learning. The NEP introduces holistic and multidisciplinary education that would help to develop intellectual, scientific, social, physical, emotional, ethical and moral capacities of the students. The NEP 2020 envisages flexible curricular structures and learning based outcome approach for the development of the students. By establishing a nationally accepted and internationally comparable credit structure and courses framework, the NEP 2020 aims to promote educational excellence, facilitate seamless academic mobility, and enhance the global competitiveness of Indian students. It fosters a system where educational achievements can be recognized and valued not only within the country but also in the international arena, expanding opportunities and opening doors for students to pursue their aspirations on a global scale.

In response to the rapid advancements in science and technology and the evolving approaches in various domains of Statistics and related subjects, the Board of Studies in Statistics at Tuljaram Chaturchand College, Baramati - Pune, has developed the curriculum for the first semester of M.Sc. Data Science, which goes beyond traditional academic boundaries. The syllabus is aligned with the NEP 2020 guidelines to ensure that students receive an education that prepares them for the challenges and opportunities of the 21st century. This syllabus has been designed under the framework of the Choice Based Credit System (CBCS), taking into consideration the guidelines set forth by the National Education Policy (NEP) 2020, LOCF (UGC), NCrF, NHEQF, Prof. R.D. Kulkarni's Report, Government of Maharashtra's General Resolution dated 20th April and 16th May 2023, and the Circular issued by SPPU, Pune on 31st May 2023.

The preamble of an MSc Data Science course typically provides an overview and introduction to the program, outlining its objectives, structure, and key features. It sets the context and expectations for students pursuing a Master's degree in Data Science.

The Master of Science (MSc) in Data Science program is designed to equip students with the knowledge, skills, and expertise necessary to excel in the rapidly evolving field of data science. This interdisciplinary program combines principles and techniques from statistics, computer science, and domain-specific areas to enable students to extract actionable insights and make data-driven decisions.

The MSc Data Science program is structured to provide a balance between theoretical foundations, practical skills, and hands-on experience. The curriculum consists of a combination of core courses, elective courses, and a capstone project. The program also offers opportunities for specialization in areas such as machine learning, big data analytics, natural language processing, or business analytics.

Upon successful completion of the MSc Data Science program, graduates will be equipped with the knowledge and skills to take on roles such as data scientists, data analysts, machine learning engineers, or data consultants in diverse industries, including finance, healthcare, e-commerce, and technology.

Overall, revising the M.Sc. Data Science syllabus in accordance with the NEP 2020 ensures that students receive an education that is relevant, comprehensive, and prepares them to navigate the dynamic and interconnected world of today. It equips them with the knowledge, skills, and competencies needed to contribute meaningfully to society and pursue their academic and professional goals in a rapidly changing global landscape.

Programme Specific Outcomes (PSOs)

PSO1. Advanced Data Analysis: Apply advanced statistical and machine learning techniques to analyze complex datasets, identify patterns, and derive actionable insights.

PSO2. Data Visualization and Communication: Effectively visualize and communicate data insights through compelling visualizations, reports, and presentations.

PSO3. Statistical Computing and Programming: Utilize statistical software packages, such as R, Python, Pawer BI, SQL etc. to implement statistical analyses and simulations.

PSO4. Research and Problem-Solving: Identify research problems, formulate appropriate hypotheses, and design research studies.

PSO5. Ethical and Legal Considerations: Understand and navigate ethical and legal challenges related to data privacy, security, and governance in the field of data science.

PSO6. Deep Learning: Apply deep learning algorithms and neural networks to solve complex data analysis problems, such as image recognition and natural language processing.

PSO7. Predictive Modeling: Build predictive models using machine learning algorithms to make accurate predictions and forecasts.

PSO8.Data Mining and Knowledge Discovery: Utilize data mining techniques to extract valuable knowledge and patterns from large and complex datasets.

Program Outcomes for M.Sc.(Data Science)

PO1	**Advanced Disciplinary Knowledge & Originality:** Demonstrating comprehensive and advanced knowledge in the chosen field of science, extending beyond the undergraduate level, providing a specialized foundation for developing and applying original ideas, particularly within a research context.
PO2	**Research, Analysis, and Complexity:** Ability to formulate hypotheses and design experiments while demonstrating the capacity to integrate knowledge and handle complex information, even when it is incomplete or limited.
PO3	**Problem Solving in New Contexts:** Apply theoretical knowledge and problem-solving abilities to unfamiliar, real-world, or multidisciplinary environments, moving beyond standard classroom scenarios to innovative applications.
PO4	**Technical Mastery and Scientific Reasoning:** Utilize modern tools, specialized techniques, and instruments with high proficiency, underpinned by a deep rationale and scientific reasoning for the choice of methodology.
PO5	**Integrated Communication:** Clearly and unambiguously communicate complex scientific conclusions, and the knowledge/rationale supporting them, to both specialist peers and non-specialist stakeholders.
PO6	**Ethical, Social, and Professional Judgment:** Adhere to strict ethical standards in research while reflecting on the social and environmental responsibilities linked to the application of scientific knowledge and professional judgments.
PO7	**Autonomous and Lifelong Learning:** Exhibit the learning skills necessary to pursue further study or professional development in a largely self-directed and autonomous manner.
PO8	**Employability, Innovation, and Entrepreneurship:** Translate advanced technical skills and independent thinking into professional excellence within industry, academia, or entrepreneurial ventures.

Title of the Programme: M.Sc. – I (Data Science)

Anekant Education Society's

Tuljaram Chaturchand College

of Arts, Science and Commerce Baramati, Dist.-Pune, MS, India.

(Empowered Autonomous)

Board of Studies in Statistics

(Academic Year 2025-26 to 2027-28)

Sr. No.	Name of Members	Designation
1.	**Prof. Dr. Kakade Vikas Chintaman** Head & Professor, Department of Statistics, T. C. College, Baramati.	**Chairperson**
2.	**Prof. Dr. Jagtap Avinash Srirangrao** Principal, Department of Statistics, T. C. College, Baramati	Member
3.	**Dr. Dhane Neeta Kishor** Associate Professor, Department of Statistics, T. C. College, Baramati	Member
4.	**Dr. Patil Vaishali Vilas** Associate Professor, Department of Statistics, T. C. College, Baramati	Member
5.	**Dr. Swami Chandrashekhar Panchayya** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
6.	**Ms. Wadkar Sarita Dipak** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
7.	**Dr. Malusare Priti Sandeep** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
8.	**Dr. Jagtap Nilambari Arvind** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
9.	**Dr. Gaikwad Pooja Sujit** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
10.	**Ms. Kalange Tejshri Chetan** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
11.	**Dr. Arekar Trupti Shantanu** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
12.	**Miss. Rakate Priya Nanasaheb** Assistant Professor, Department of Statistics,	Member

	T. C. College, Baramati	
13.	**Ms. Choudhar Shital Balu** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
14.	**Miss. Dhokchaule Rutuja Babasaheb** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
15.	**Miss. Ghadge Kiran Tanaji** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
16.	**Miss. Ranmode Snehal Sanjay** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
17.	**Miss. Prabhune Utkarsha Shrinivas** Assistant Professor, Department of Statistics, T. C. College, Baramati	Member
18.	**Dr. Akanksha Kashikar**	Vice-Chancellor Nominee Subject Expert from SPPU, Pune
19.	**Dr. Koshti Rohan**	Subject Expert from Outside the Parent University
20.	**Prof. Gardi Chandrakant Gopal**	Subject Expert from Outside the Parent University
21.	**Mr. Kadam Saurabh**	Representative from industry/corporate sector/allied areas
22.	**Dr. Limbore Jaya Laxman**	Member of the College Alumni
23.	**Miss. Shirke Shatakshi Shrikant**	UG Student
24.	**Miss. Pathak Siddhi Rajendra**	PG Student

Course Structure for M.Sc. Part-I (Data Science) (2026 Pattern)

Level	Sem	Course Type	Course Code	Course Title	Theory/ Practical	No. of Credits
6.0	I	Major (Mandatory)	DSC-501-MRM	Probability and Statistics for Data Science	Theory	04
			DSC -502-MRM	Data Analytics Using R	Theory	04
			DSC -503-MRM	Data Base Management System	Theory	02
			DSC -504-MRM	Data Science Practical – I	Practical	02
			DSC -505-MRM	Practical Based on Python	Practical	02
		Major (Elective)	DSC -506-MJE(A)	Fundamentals of Data Science	Theory (Any one)	02
			DSC -506-MJE(B)	Stochastic Models and Applications		
			DSC -507-MJE(A)	Practical Based on Fundamentals of Data Science	Practical (Any one)	02
			DSC -507-MJE(B)	Practical Based on Stochastic Models and Applications		
		Research Methodology (RM)	DSC -508-RM	Research Methodology	Theory	04
		Total Credits Semester I				**22**
6.0	II	Major (Mandatory)	DSC -551-MRM	Machine Learning and Artificial intelligence	Theory	04
			DSC -552-MRM	Regression Analysis and Predictive Models	Theory	04
			DSC -553-MRM	Inferential Statistics for Data Science	Theory	02
			DSC -554-MRM	Data Science Practical – II	Practical	02
			DSC -555-MRM	Data Science Practical – III	Practical	02
		Major (Elective)	DSC -556-MJE(A)	Bayesian Inference	Theory (Any one)	02
			DSC -556-MJE(B)	Computational Statistics		
			DSC -557-MJE(A)	Data Visualization Using Tableau and Power BI	Practical (Any one)	02
			DSC -557-MJE(B)	Data Science Practical – IV		
		On Job Training (OJT)	DSC -581-OJT/FP	On Job Training	Training	04
		Total Credits Semester-II				**22**
		Cumulative Credits Semester I and II				**44**

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	**:** M.Sc. Data Science
Program Code	**:** PSDSC
Class	**:** M.Sc. Part – I
Semester	**:** I
Course Type	**:** Major Mandatory Theory
Course Name	**:** Probability and Statistics for Data Science
Course Code	**:** DSC-501-MRM
No. of Credits	**:** 4
No. of Teaching Hours	**:** 60

Course Objectives:

1. To develop fundamental understanding of statistical concepts used in data science and analytics.
2. To enable students to organize, summarize and visualize real world datasets using descriptive statistical techniques.
3. To develop understanding of correlation analysis for identifying relationships between variables.
4. To build strong foundation in probability theory for modeling uncertainty in real world data science problems.
5. To introduce random variables and probability models used in machine learning and data analytics.
6. To develop ability to apply probability laws and conditional probability concepts in real data situations.
7. To provide knowledge of discrete probability distributions used for modeling count and event-based data.
8. To provide knowledge of continuous probability distributions used for modeling real valued and natural phenomena data.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Apply descriptive statistical methods to summarize and analyze real world datasets.

CO2. Compute and interpret correlation measures to identify relationships between variables.

CO3. Apply probability laws and conditional probability concepts to solve uncertainty based data science problems.

CO4. Model random phenomena using discrete and continuous random variables.

CO5. Apply discrete probability distributions to solve real world event and count based data problems.

CO6. Apply continuous probability distributions to model real world continuous data scenarios.

CO7. Analyze probabilistic models to support data driven decision making.

CO8. Interpret statistical and probabilistic results in data science, business and technology applications.

Topics and Learning Points

Unit 1: **(20 L)**

Descriptive Statistics and Correlation:

Introduction to Statistics, role of statistics in data science, types of data (qualitative and quantitative), population and sample, methods of data collection, scrutiny and cleaning of data, classification and tabulation of data, diagrammatic presentation of data , graphical presentation of data , measures of central tendency, measures of dispersion , measures of shape including skewness and kurtosis with interpretation, bivariate data and scatter diagram, types of correlation (positive, negative, zero), Karl Pearson's coefficient of correlation (r), properties, interpretation of correlation, coefficient of determination (r^2), introduction to multiple correlation coefficient (concept, definition, interpretation), introduction to partial correlation coefficient (concept, definition, interpretation)

Unit 2: **(20 L)**

Probability and Random Variables: Introduction – Random Experiments, Empirical basis of probability, Algebra of events, laws of Probability; Conditional Probability, Independence, Bayes' law; Application of probability to business and economics. One-dimensional Random variable- Discrete and Continuous; Distribution functions and its properties; Bivariate Random Variables- Joint Probability functions, marginal distributions, conditional distribution functions, Notion of Independence of Random variables.

Unit 3: **(10 L)**

Discrete Distributions: Bernoulli, Binomial, Poisson, Geometric, Hypergeometric, Negative Binomial, Multinomial, distributions and Discrete Uniform distribution - definition, properties and applications with numerical problems.

Unit 4: (10L)

Continuous Distributions: Uniform, Normal, Exponential, Gamma, Beta distributions (First and Second kind), - definition, properties and applications with numerical problems.

References:

1. Parimal Mukhopadhyay; An Introduction to the Theory of Probability, World scientific, 2012.
2. Irwin Miller, Marylees Miller, John E. Freund's; Mathematical Statistics, Pearson, 2017.
3. Fetsje Bijma, Marianne Jonker and Aadvander Vaart; Introduction to Mathematical Statistics, Amsterdam University Press, 2018.
4. Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, 2006.
5. Shanmugam, R., Chattamvelli, R. Statistics for scientists and engineers, John Wiley, 2015.
6. Casella G. and Beregar R.L. (2002) Statistical Inference, 2nd Edition (Duxbury Advanced Series)
7. Dudewitz E.J. & Mishra S.N.(1988) Modern Mathematical Statistics (John Wiley)
8. Kale B.K. (1999) A First course on Parametric Inference (Narosa)
9. Lehman E.L (1988) Theory of point estimation (John Wiley)
10. Lehman E.L(1986) Testing of Statistical hypotheses (John Wiley)
11. Rohatagi V.K. (1976) Introduction to theory of probability & mathematical statistics (John Wiley & sons)

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	2	1	3	1	2	2
CO2	2	2	2	1	2	1	2	2
CO3	3	3	3	2	2	1	3	3
CO4	3	3	3	3	1	1	3	3
CO5	3	3	3	3	1	1	3	3
CO6	3	3	2	2	2	1	2	2
CO7	2	2	3	1	3	2	2	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- The course develops strong theoretical and applied knowledge of statistical inference, estimation theory, confidence intervals and data-driven statistical decision making used in data science and research. CO1, CO2, CO3 and CO5 strongly support disciplinary statistical knowledge, while CO4 moderately supports computational statistical implementation.

PO2: Research, Analysis, and Complexity

- The course strengthens analytical thinking through estimator comparison, interval estimation, and statistical inference using real datasets. CO3, CO4 and CO5 strongly support research and data analysis, while CO1 and CO2 moderately support theoretical statistical understanding.

PO3: Problem Solving in New Contexts

- The course enables students to solve real world data science problems using estimation techniques and statistical inference methods. CO2, CO3, CO4 and CO5 strongly support applied problem solving, while CO1 moderately supports theoretical problem understanding.

PO4: Technical Mastery and Scientific Reasoning

- The course develops computational and statistical reasoning skills using statistical software and programming tools. CO2, CO4 and CO5 strongly support technical

mastery, CO3 moderately supports applied statistical reasoning, while CO1 partially supports theoretical reasoning.

PO5: Integrated Communication

- The course enables students to interpret estimation results, confidence intervals and statistical outputs effectively. CO3 and CO5 strongly support interpretation and result communication, CO2 and CO4 moderately support technical explanation, while CO1 partially supports conceptual understanding.

PO6: Ethical, Social, and Professional Judgment

- The course supports responsible use of statistical inference in data-driven decision making and real-world data analysis. CO4 and CO5 moderately support responsible data interpretation, while CO1–CO3 partially support professional statistical responsibility.

PO7: Autonomous and Lifelong Learning

- The course builds a strong statistical foundation for advanced learning in data science, machine learning and applied statistics. CO2, CO4 and CO5 strongly support lifelong learning, while CO1 and CO3 moderately support continuous statistical skill development.

PO8: Employability, Innovation, and Entrepreneurship

- The course develops industry-relevant skills in statistical modeling, estimation, and data analysis for data science careers. CO2, CO3, CO4 and CO5 strongly support employability and industry readiness, while CO1 moderately supports theoretical industry foundation.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: I
Course Type	: Major Mandatory Theory
Course Name	: Data Analytics Using R
Course Code	: DSC-502-MRM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

Students successfully completing this course will be able to:

1. Provide an overview of R and R Studio, including installation, basic operations, and the use of R as a calculator for arithmetic and logical operations.
2. Develop a solid understanding of different data types (numeric, integer, character, logical, factor) and how to create, index, and operate on various data structures.
3. Enable students to generate random samples from different probability distributions and compute probabilities, cumulative probabilities, and quantiles.
4. Teach students to test the normality of data using the Shapiro-Wilk test and interpret the results.
5. Introduce the concepts of null and alternative hypotheses, type I and type II errors, and conduct various parametric (z test, t test, ANOVA) and non-parametric tests.
6. Provide a thorough understanding of control structures (if-else, for loops, while loops, repeat loops) and their use in R programming.
7. Develop skills in writing R programs, including debugging and error handling, to solve data analysis problems effectively.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Demonstrate the ability to install, navigate, and utilize R and RStudio for data analysis tasks.

CO2. Identify and manipulate different data types (numeric, integer, character, logical, factor) and data structures (vectors, matrices, lists, data frames, factors) in R.

CO3. Read, write, and manipulate data from various sources (CSV, Excel, text files)

and create diverse types of plots (pie charts, bar charts, scatter plots, histograms, boxplots) using basic R functions and ggplot2.

CO4. Generate random samples from various probability distributions and compute probabilities, cumulative probabilities, and quantiles.

CO5. Create graphs of probability mass functions (pmf) and probability density functions (pdf) and fit probability distributions to data.

CO6. Calculate and interpret descriptive statistics, including measures of central tendency, variability, and distribution shape (mean, mode, median, quartiles, variance, standard deviation, skewness, kurtosis).

CO7. Implement control structures (if-else, for loops, while loops, repeat loops) in R to automate and streamline data analysis tasks.

Topics and Learning Points

Unit 1: Introduction to R **(15 L)**

Overview of R and RStudio: features, applications in Data Science. Installations and configuration of R and RStudio, R-Script and R-markdown R as a calculator: arithmetic, relational, logical, and assignment operations, Data types: numeric, integer, character, logical, and factor. Data structures: vectors, matrices, lists, data frames, and factors. Creation, indexing, modification and operations on data structures. Built-in-functions and user-defined basic functions use of cat( ) and print( ) commands. Reading and writing data in R (CSV, Excel, text files, etc.), Creating various types of plots: pie chart, bar chart, group bar chart, stacked bar chart, scatter plots, line plots, histograms, boxplots and Introduction to ggplot2.

Unit2: Probability Distributions with R **(15 L)**

Generating random samples from discrete and continuous probability distributions, computations of probabilities, cumulative probabilities, and quantiles. Graphs of pmf/pdf by varying parameters of the distributions. Fitting probability distributions to real data, Testing normality of data by Q-Q plots and Shapiro Wilks test.

Unit3: Statistical Analysis with R **(18 L)**

Descriptive statistics: mean, mode, median, quartiles, minimum and maximum value, percentiles, variance, standard deviation, coefficient of variation, covariance and correlation,

moments, skewness and kurtosis. Hypothesis testing: Null and alternative hypotheses, type I and type II errors, Parametric test: z test, t test, proportion test, variance test, chi-square test, correlation test and ANOVA. Non-parametric tests: Bartlett's test for homoscedasticity, Kruscal-Wallis test, Kolmogorov-Smirnov test, Sign test, Sign test for paired data, Wilcoxon's signed rank test, Mann Whitney test.

Unit4: Programming in R (12 L)

Control structures and loops: if, if-else statements, for loops, while loops, repeat loops, use of break( ) and next() in loops, Functions in R: Defining and calling functions and returning values functions arguments and default parameters, Apply family of functions: apply(), lapply(), sapply(), tapply(), mapply().Writing efficient and modular programs in R. Debugging and error handling in R.

References:

1. Hadley Wickham & Garrett Grolemund. R for Data Science
2. Norman Matloff. The Art of R Programming
3. Hadley Wickham. Advanced R
4. David S. Moore, George P. McCabe, and Bruce A. Craig. Introduction to the Practice of Statistics"
5. Crawley, M. J. (2006). Statistics - An introduction using R. John Wiley, London
6. Purohit, S.G.; Gore, S.D. and Deshmukh, S.R. (2015). Statistics using R, second edition. Narosa Publishing House, New Delhi.
7. Online resources and R documentation

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)							
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	2	-		3	-	1	2	2
CO2	2	2	2	3	-	1	2	2
CO3	2	2	2	3	2	1	2	2
CO4	3	3	2	3	-	1	2	2
CO5	3	3	2	3	2	1	2	2
CO6	3	3	2	2	2	1	2	2
CO7	2	2	3	3	-	1	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

Justification

PO1: Advanced Disciplinary Knowledge & Originality

- CO4: Generate random samples from various probability distributions and compute probabilities, cumulative probabilities, and quantiles, thereby strengthening advanced statistical knowledge. (Weightage: 3)
- CO5: Create graphs of probability mass functions and probability density functions and fit probability distributions to data, contributing to deeper disciplinary understanding. (Weightage: 3)
- CO6: Calculate and interpret descriptive statistics, reinforcing conceptual clarity and analytical depth in statistics. (Weightage: 3)
- CO1, CO2, CO3, CO7: Support foundational disciplinary knowledge related to statistical computing and data handling. (Weightage: 2)

PO2: Research, Analysis, and Complexity

- CO4: Strongly contributes to research skills by enabling simulation, probability computation, and analytical experimentation. (Weightage: 3)
- CO5: Facilitates analysis of data through distribution fitting and graphical interpretation in research contexts. (Weightage: 3)
- CO6: Supports research analysis through interpretation of statistical measures and data summaries. (Weightage: 3)
- CO2, CO3, CO7: Moderately related by enabling data manipulation, visualization, and structured analytical workflows. (Weightage: 2)

PO3: Problem Solving in New Contexts

- CO7: Strongly related as it enables the use of control structures in R to solve complex and unfamiliar data analysis problems efficiently. (Weightage: 3)
- CO2, CO3, CO4, CO5, CO6: Moderately related by applying computational and statistical methods to real-world and interdisciplinary problems. (Weightage: 2)

PO4: Technical Mastery and Scientific Reasoning

- CO1: Demonstrates strong technical proficiency in installing and using R and RStudio for scientific data analysis. (Weightage: 3)
- CO2, CO3: Enable mastery of data structures, file handling, and visualization using modern statistical tools. (Weightage: 3)
- CO4, CO5, CO7: Strongly related through application of computational techniques, modeling, and algorithmic logic. (Weightage: 3)
- CO6: Moderately supports scientific reasoning through statistical interpretation. (Weightage: 2)

PO5: Integrated Communication

- CO3: Moderately contributes through graphical visualization and presentation of data using plots and charts. (Weightage: 2)
- CO5: Supports effective communication of statistical results through graphical representation of distributions. (Weightage: 2)
- CO6: Enhances communication by interpreting and summarizing statistical findings clearly. (Weightage: 2)

PO6: Ethical, Social, and Professional Judgment

- All Cos: Partially related as students are exposed to ethical data handling practices, responsible analysis, and professional use of statistical software. (Weightage: 1)

PO7: Autonomous and Lifelong Learning

- CO7: Strongly contributes by encouraging independent learning, logical thinking, and self-directed problem-solving using programming constructs. (Weightage: 3)
- CO1, CO2, CO3, CO4, CO5, CO6: Moderately related by fostering continuous learning of statistical and computational skills. (Weightage: 2)

PO8: Employability, Innovation, and Entrepreneurship

- CO7: Strongly related as automation and programming skills enhance employability and innovation in data-driven roles. (Weightage: 3)

- CO1, CO2, CO3, CO4, CO5, CO6: Moderately related by developing industry-relevant competencies in data analysis and statistical computing. (Weightage: 2)

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: I
Course Type	: Major Mandatory Theory
Course Name	: Database Management System
Course Code	: DSC-503-MRM
No. of Credits	: 2
No. of Teaching Hours	: 30

Course Objectives:

1. Students should gain a solid understanding of the basic concepts and principles of database management systems.
2. Students should learn how to design a relational database, including identifying entities, attributes, and relationships.
3. Students should become proficient in SQL, the standard language for interacting with relational databases.
4. Students should learn techniques for optimizing database queries to improve performance
5. Examine the logical, physical, and database modelling designs.
6. Students should be exposed to emerging trends and technologies in the field of database management systems.
7. Recognize how to create, modify, and query databases for data

Course Outcomes:

By the end of the course, students will be able to:

CO1. be familiar with the fundamentals of database concepts and database management systems.

CO2. understand the fundamental concepts and principles of database management systems, including data models, schemas, instances, and database architecture.

CO3. implement mechanisms for ensuring data integrity, such as primary keys, foreign keys, and constraints. utilize.

CO4. conceptual modelling techniques, like as the ER model and relational model, to model the data requirements for an application.

CO5. Write SQL commands to create tables, insert, update, delete and querying

data.

CO6. create and manage database objects, such as tables, views, indexes, and constraints, using SQL.

CO7. normalize a database schema to eliminate redundancy and ensure data integrity.

Topics and Learning Points

Unit 1: **(10L)**

Introduction to file organization &DBMS, Database-system Applications, Purpose of Database Systems, Types of file Organization, File system Vs. DBMS, Data models, Levels of abstraction, Data in dependence, Structure of DBMS, Users of DBMS, Database Architecture, Speciality Databases. Structure of Relational Databases, Database Schema, Keys, Relational Operations, Conceptual Design (E-R model), Overview of DB design, ER data model (entities, attributes, entity sets, relations, relationship sets), Additional constraints (Key constraints, Mapping constraints), Conceptual design using ER modelling. Relational data model, Conversion of ER to Relational model, Integrity constraints, Relational algebra, Preliminaries.

Unit 2: **(10L)**

Introduction to SQL, Basic structure, set operations, Aggregate functions, Null values, PL/PgSqL: Data types, Language structure, Operations with SQL, Nested Sub queries, Modifications to Database, DDL and DML commands with examples, SQL mechanisms for joining.

Unit 3 : **(10L)**

Intermediate and advanced SQL: Join Expressions- Join conditions, Outer joins, Join types and conditions, Views- View definition, using views in SQL queries, Materialized views, update a view 4.3 Create table extensions, Schemas, Catalogs and Environments, The relational Algebra, The tuple relational calculus.

References:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarashan, Database System Concepts, McGraw-Hill International Edition, Sixth Edition
2. Elmasri, Navathe, Fundamentals of Database Systems, Pearson Education, Third Education

3. Ramakrishnan, Gehrke, Database Management Systems, McGraw Hill International Edition, Third Edition
4. Peter Rob, Carlos Coronel, Database System Concepts, Cengage Learning, India Edition
5. S.K. Singh, "Database Systems Concepts, Design and Applications", First Edition, Pearson Education, 2006
6. Redmond, E. & Wilson, Seven Databases in Seven Weeks: A Guide to Modern Databases and the No SQL Movement Edition:1st Edition.
7. Shamkant B. Navathe, RamezElmasri,(2010), Database Systems, ISBN:9780132144988, PEARSON HIGHER EDUCATION
8. Richard Stones, Neil Matthew, (2005), Beginning Databases with PostgreSQL: From Novice to Professional, ISBN:9781590594780, Apress
9. Korry, Douglas, (2005), Postgre SQL, ISBN:9780672327568, Sams Publishing.
10. Joshua D. Drake, John C. Worsley, Practical Postgre SQL, (2002), ISBN:9788173663925 O'Reilly Media, Inc., ISBN: 9781565928466.

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	1	1	1	1	1	2	1
CO2	3	3	2	1	2	1	3	2
CO3	2	2	2	2	1	2	2	2
CO4	3	3	3	1	3	1	3	2
CO5	2	1	3	3	3	1	2	3
CO6	2	2	3	3	2	1	2	3
CO7	3	3	3	2	2	2	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- This course develops strong foundational and advanced knowledge of database systems required for data science and software applications. CO1, CO2, CO4 and

CO7 strongly support core disciplinary knowledge, while CO3, CO5 and CO6 moderately support applied database implementation knowledge.

PO2: Research, Analysis, and Complexity

- Database design, normalization and data integrity mechanisms support structured data analysis and handling complex data systems. CO2, CO4 and CO7 strongly support analytical database design thinking, CO3 and CO6 moderately support system implementation analysis, while CO1 and CO5 partially support fundamental and operational understanding.

PO3: Problem Solving in New Contexts

- Database skills help solve real world data storage, retrieval and application integration problems. CO4, CO5, CO6 and CO7 strongly support real world database problem solving, CO2 and CO3 moderately support system level problem solving, while CO1 partially supports conceptual foundation.

PO4: Technical Mastery and Scientific Reasoning

- The course builds technical proficiency in database creation, querying and management tools. CO5 and CO6 strongly support technical mastery in SQL and database objects, CO3 and CO7 moderately support technical reasoning, while CO1, CO2 and CO4 partially support conceptual technical understanding.

PO5: Integrated Communication

- Database schema design and querying help in structured data representation and communication. CO4 and CO5 strongly support structured data communication, CO2, CO6 and CO7 moderately support technical data presentation, while CO1 and CO3 partially support database concept communication.

PO6: Ethical, Social, and Professional Judgment

- Data integrity, constraints and secure data handling support responsible data usage. CO3 and CO7 moderately support ethical data management, while CO1, CO2, CO4, CO5 and CO6 partially support professional database practices.

PO7: Autonomous and Lifelong Learning

- Database fundamentals form a base for advanced learning in data engineering, big data and cloud databases. CO2, CO4 and CO7 strongly support future learning capability, while CO1, CO3, CO5 and CO6 moderately support continuous technical skill development.

PO8: Employability, Innovation, and Entrepreneurship

- Database skills are highly relevant in industry applications such as data engineering and software development. CO5, CO6 and CO7 strongly support employability and technical industry readiness, CO2, CO3 and CO4 moderately support applied database design skills, while CO1 partially supports fundamental employability knowledge.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: I
Course Type	: Major Mandatory Practical
Course Name	: Data Science Practical – I
Course Code	: DSC-504-MRM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. To develop fundamental understanding of matrix algebra and its applications in data science and scientific computing.
2. To provide knowledge of numerical methods for solving linear algebraic systems.
3. To enable students to understand eigen values, eigen vectors and their applications in data analysis.
4. To develop skills in matrix decomposition techniques used in data science and machine learning.
5. To provide practical exposure to probability distributions and statistical computations using R.
6. To develop ability to perform exploratory data analysis and correlation analysis using real datasets.
7. To introduce statistical estimation techniques and confidence interval construction using computational tools.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Perform matrix operations and apply generalized inverse concepts using R programming.

CO2. Compute eigen values and eigen vectors and apply them in matrix transformations and data analysis.

CO3. Apply numerical methods to solve systems of linear equations efficiently.

CO4. Apply matrix decomposition techniques and quadratic form transformations in

practical problems.

CO5. Analyze and visualize probability distributions using computational tools.

CO6. Perform exploratory data analysis and correlation analysis on real world datasets.

CO7. Implement statistical estimation methods and construct confidence intervals using real data.

Topics and Learning Points

Sr. No.	Title of Experiments
1	Matrix Operations in R
2	Computation of Generalized Inverse and MPG-Inverse
3	Computation of Eigen values and Eigenvectors of a Given Matrix
4	Performing Spectral Decomposition of Symmetric Matrices
5	Matrix Power Calculation via Eigen value and Eigenvector Methods
6	Solution of Linear Systems Using the Gauss Elimination, Gauss-Jordan Method, Gauss-Seidel Method
7	Verification and Application of the Cayley-Hamilton Theorem
8	Classification and Reduction of Quadratic Forms Using eigen values and Eigenvectors
9	Plotting of density function of univariate Probability distribution
10	Exploratory Data Analysis and Correlation Analysis Using Real World Dataset
11	Computation of probability of events related to discrete and continuous probability distribution
12	Applications of probability distributions

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	1	1	3	1	1	2	2
CO2	3	3	2	3	2	1	3	2
CO3	2	2	3	2	1	1	2	3
CO4	3	3	3	3	1	1	3	2
CO5	2	2	3	2	3	1	2	3
CO6	2	3	3	1	3	2	2	3
CO7	3	3	3	2	2	2	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- The practical course develops advanced mathematical and computational knowledge useful in data science and research applications. CO1, CO2, CO4 and CO7 strongly support advanced disciplinary knowledge, while CO3, CO5 and CO6 moderately support applied analytical and statistical understanding.

PO2: Research, Analysis, and Complexity

- The course builds analytical and computational skills required for handling complex mathematical and statistical data problems. CO2, CO4, CO6 and CO7 strongly support research level analysis and interpretation, CO3 and CO5 moderately support computational analysis, while CO1 partially supports analytical foundation.

PO3: Problem Solving in New Contexts

- Students learn to apply computational mathematics and statistics to real world and interdisciplinary problems. CO3, CO4, CO5, CO6 and CO7 strongly support practical problem solving, CO2 moderately supports transformation based problem solving, while CO1 partially supports fundamental computational application.

PO4: Technical Mastery and Scientific Reasoning

- The course provides hands-on technical skill development using R programming and numerical methods. CO1, CO2 and CO4 strongly support technical mastery, CO3, CO5 and CO7 moderately support technical reasoning, while CO6 partially supports technical interpretation.

PO5: Integrated Communication

- Data visualization, statistical interpretation and reporting improve technical communication skills. CO5 and CO6 strongly support data interpretation and

communication, CO2 and CO7 moderately support technical result explanation, while CO1, CO3 and CO4 partially support structured technical understanding.

PO6: Ethical, Social, and Professional Judgment

- Proper data analysis and interpretation supports ethical data usage and professional responsibility. CO6 and CO7 moderately support responsible data analysis, while CO1–CO5 partially support professional technical practices.

PO7: Autonomous and Lifelong Learning

- The course builds strong computational and analytical base for advanced learning in AI, ML and analytics. CO2, CO4 and CO7 strongly support lifelong learning, while CO1, CO3, CO5 and CO6 moderately support continuous technical skill development.

PO8: Employability, Innovation, and Entrepreneurship

- The course develops industry relevant computational, statistical and data analysis skills. CO3, CO5, CO6 and CO7 strongly support employability and industry readiness, CO1, CO2 and CO4 moderately support technical job skills.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: I
Course Type	: Major Mandatory Practical
Course Name	: Practical Based on Python
Course Code	: DSC-505-MRM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Develop proficiency in Python programming language and its syntax for data analytics.
2. Understand and apply core programming constructs such as loops, functions, and control structures in Python.
3. Gain hands-on experience with Python scientific libraries for numerical and statistical computations.
4. Apply data manipulation and preprocessing techniques using Pandas.
5. Perform exploratory data analysis and data visualization using Python libraries.
6. Implement basic statistical analysis and summary techniques using Python.
7. Apply Python programming to solve real-world data analysis problems.
8. Develop problem-solving ability and analytical thinking using Python-based mini tasks and datasets.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Demonstrate a strong command of Python programming language and its syntax.

CO2. Apply Python programming constructs such as loops, conditional statements, and regular expressions.

CO3. Use Python scientific libraries (NumPy, SciPy) for numerical and scientific computations.

CO4. Perform data manipulation, cleaning, and transformation using Pandas.

CO5. Conduct exploratory data analysis and data visualization using Matplotlib and Seaborn.

CO6. Analyze datasets using descriptive statistics and summary measures in Python.

CO7. Implement control structures and logical programming to automate data analysis tasks.

CO8. Demonstrate proficiency in using Python libraries and frameworks commonly used to solve real-world data analysis problems and interpret results effectively.

Topics and Learning Points

Sr. No.	Title of Experiments
1.	Basics of Python Language: When and why to use Python for Analytics, Introduction & Installation of Python, Python Syntax, Strings, Lists, tuples, and Dictionaries
2.	Operators in Python: Arithmetic, Relational, Assignment, logical, bitwise, Ternary, Membership, Identity Operators with suitable examples.
3.	Conditional statements: if, if else, if-elif-else, return multiple values at a time using return statement. Loops: while loop, for loop.
4.	Defining functions in python: simple python functions, lambda function
5.	Scientific Libraries in Python: NumPy, SciPy
6.	Introduction to Pandas: Selecting data from Pandas DataFrame, Slicing and dicing using Pandas
7.	Introduction to Pandas: GroupBY / Aggregate, Strings with Pandas, cleaning up messy data with Pandas, Dropping Entries, Selecting Entries
8.	Data Manipulation using Pandas – I: Data alignment, sorting and ranking, summary statistics, missing values, merging data
9.	Data Manipulation using Pandas – I: Concatenation, Combining DataFrames, Pivot, Duplicates, Binning
10.	Data visualization on using matplotlib and seaborn libraries: Scatter plot, Line plot, Bar plot, Histogram, Box plot, Pair plot
11.	Control structures using Toyota Corolla dataset: if-else family, for loop, for loop with if break, while loop
12.	Case Study

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)							
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	1	1	2	--	--	2	1
CO2	2	2	2	2	--	--	2	2
CO3	3	2	2	3	--	--	1	2
CO4	3	3	2	3	--	--	1	2
CO5	2	2	2	2	3	--	1	2
CO6	3	3	2	2	--	--	1	2
CO7	2	2	3	2	--	--	2	3
CO8	3	3	3	3	2	--	2	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

Justification:

PO1: Advanced Disciplinary Knowledge & Originality

- CO1: Demonstrate a strong command of Python programming language and its syntax. (Weightage: 3)
- CO3: Use Python scientific libraries (NumPy, SciPy) for numerical and scientific computations. (Weightage: 3)
- CO4: Perform data manipulation, cleaning, and transformation using Pandas. (Weightage: 3)
- CO6: Analyze datasets using descriptive statistics and summary measures in Python. (Weightage: 3)
- CO8: Demonstrate proficiency in using Python libraries and frameworks to solve real-world data analysis problems. (Weightage: 3)

PO2: Research, Analysis, and Complexity

- CO2: Apply Python programming constructs such as loops, conditional statements, and regular expressions. (Weightage: 2)
- CO3: Use Python scientific libraries for analytical and numerical problem solving. (Weightage: 2)
- CO4: Perform data manipulation and transformation for complex datasets. (Weightage: 3)
- CO6: Conduct statistical analysis and interpret summary measures. (Weightage: 3)
- CO8: Analyze real-world datasets and interpret results effectively. (Weightage: 3)

PO3: Problem Solving in New Contexts

- CO2: Apply Python constructs to solve computational problems. (Weightage: 2)
- CO3: Apply numerical and scientific methods in unfamiliar contexts. (Weightage: 2)
- CO5: Perform exploratory data analysis to extract insights. (Weightage: 2)
- CO7: Implement control structures to automate data analysis tasks. (Weightage: 3)
- CO8: Solve real-world and multidisciplinary data analysis problems. (Weightage: 3)

PO4: Technical Mastery and Scientific Reasoning

- CO1: Demonstrate proficiency in Python syntax and programming foundations. (Weightage: 2)
- CO3: Use advanced scientific libraries with appropriate methodology. (Weightage: 3)
- CO4: Apply professional data handling tools and techniques. (Weightage: 3)
- CO5: Visualize and interpret data using appropriate plotting tools. (Weightage: 2)
- CO8: Select and apply suitable Python frameworks for real-world problems. (Weightage: 3)

PO5: Integrated Communication

- CO5: Communicate analytical insights effectively using visualizations. (Weightage: 3)
- CO8: Interpret and communicate results clearly to technical and non-technical audiences. (Weightage: 2)

PO6: Ethical, Social, and Professional Judgment

- No direct CO mapping. Ethical considerations are implicitly addressed through responsible data usage and interpretation.

PO7: Autonomous and Lifelong Learning

- CO1: Build foundational programming skills supporting independent learning. (Weightage: 2)
- CO2: Develop logical thinking for self-directed problem solving. (Weightage: 2)
- CO7: Enhance automation skills enabling independent analytical workflows. (Weightage: 2)
- CO8: Demonstrate continuous learning through application of evolving Python tools. (Weightage: 2)

PO8: Employability, Innovation, and Entrepreneurship

- CO2: Apply programming skills relevant to industry problem solving. (Weightage: 2)
- CO3: Use scientific libraries applicable to professional analytics roles. (Weightage: 2)

- CO7: Automate analytical processes for efficient professional practice. (Weightage: 3)
- CO8: Translate Python expertise into real-world, industry-ready solutions. (Weightage: 3)

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: I
Course Type	: Major Elective Theory
Course Name	: Fundamentals of Data Science
Course Code	: DSC-506-MJE(A)
No. of Credits	: 2
No. of Teaching Hours	: 30

Course Objectives:

After completing this course, the student will be able to:

1. Understand the basic concepts, scope, and significance of Data Science and its applications in real-world problems.
2. Explain the characteristics of data using the 3 V's (Volume, Velocity, Variety) and identify different sources of data.
3. Distinguish between structured, semi-structured, and unstructured data and understand the challenges associated with unstructured data.
4. Apply appropriate similarity and dissimilarity measures for different types of data attributes.
5. Identify data quality issues and explain the need for data preprocessing in the data science process.
6. Perform data cleaning and data wrangling operations to handle missing values, noise, inconsistencies, and formatting problems.
7. Apply data transformation, normalization, discretization, and integration techniques on datasets from multiple sources.
8. Understand feature representation, feature selection, and dimensionality reduction techniques to handle high-dimensional data.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Explain the fundamentals of Data Science, its lifecycle, and its role in solving data-driven problems.

CO2. Identify and classify different types of data and data sources suitable for analytical tasks.

CO3. Compute similarity and dissimilarity measures for nominal, binary, ordinal, and numeric data.

CO4. Analyze data quality issues and apply appropriate preprocessing techniques to improve data reliability.

CO5. Clean and prepare datasets by handling missing values, noisy data, duplicates, and inconsistencies.

CO6. Perform data transformation, scaling, normalization, and integration for effective data analysis.

CO7. Represent data using appropriate features and address high-dimensionality using feature reduction techniques.

CO8. Students will be able to apply data transformation techniques such as normalization, standardization, discretization and attribute construction

Topics and Learning Points

Unit 1: Introduction to Data Science (7L)

Introduction to Data Science, The 3 V's of Data: Volume, Velocity, Variety, Why learn Data Science, Applications of Data Science, The Data Science Lifecycle, Data Scientist's Toolbox, Types of Data: Structured, Semi-structured and Unstructured Data, Problems with Unstructured Data, Data Sources: Open Data, Social Media Data, Multimodal Data, Standard Datasets, Measuring Data Similarity and Dissimilarity, Data Matrix versus Dissimilarity Matrix, Proximity Measures for Nominal Attributes, Proximity Measures for Binary Attributes, Dissimilarity of Numeric Data: Euclidean, Manhattan and Minkowski Distances, Proximity Measures for Ordinal Attributes

Unit 2: Data Pre-processing and Data Quality (8L)

Data Objects and Attribute Types, Concept of Attributes, Nominal, Binary and Ordinal Attributes, Numeric Attributes, Discrete and Continuous Attributes, Data Quality and Data Preparation, Need for Data Pre-processing, Data Munging and Data Wrangling Operations, Data Cleaning Techniques, Handling Missing Values, Noisy Data and Inconsistencies, Duplicate Entries and Multiple Records for a Single Entity, Missing and NULL Values, Outliers, Out-of-date Data, Artificial and Invalid Entries, Formatting and Structural Issues, Irregular Spacing and Extra Whitespace, Inconsistent Capitalization, Inconsistent Delimiters, Irregular NULL Formats, Invalid Characters, Incompatible Date-Time Formats

Unit 3: Data Transformation, Integration (8L)

Data Transformation, Normalization and Standardization, Scaling Techniques and Their Importance, Discretization and Binning Methods, Attribute Construction and Transformation, Data Integration from Multiple Sources, Entity Identification and Schema Matching, Redundancy and Correlation Analysis, Data Conflict and Inconsistency Resolution.

Unit 3: **Feature Representation (7L)**

Feature Representation for Different Data Types, High-dimensional Data Challenges, Curse of Dimensionality, Feature Selection and Feature Extraction, Introduction to Data Reduction, Data Reduction Strategies, Sampling Methods, Dimensionality Reduction.

References:

1. Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 3rd Edition.
2. Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V., *Introduction to Data Mining*, Pearson Education, 2nd Edition.
3. Provost, F., & Fawcett, T., *Data Science for Business*, O'Reilly Media.
4. James, G., Witten, D., Hastie, T., & Tibshirani, R., *An Introduction to Statistical Learning*, Springer.
5. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J., *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann.
6. Leskovec, J., Rajaraman, A., & Ullman, J. D., *Mining of Massive Datasets*, Cambridge University Press.
7. Grus, J., *Data Science from Scratch*, O'Reilly Media.
8. Aggarwal, C. C., *Data Mining: The Textbook*, Springer.

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	2	1	3	2	2	2
CO2	2	2	2	1	1	1	2	2
CO3	3	3	2	3	2	1	2	3
CO4	2	3	3	2	3	3	2	3
CO5	2	3	3	2	3	3	2	3
CO6	3	3	3	3	2	2	3	3

CO7	3	3	3	3	3	2	3	3
CO8	3	3	3	3	2	2	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- The course develops strong foundational and advanced knowledge in data science concepts such as lifecycle, data types, similarity measures, preprocessing and feature engineering. CO1, CO3, CO6, CO7 and CO8 strongly support disciplinary knowledge, while CO2, CO4 and CO5 moderately support applied data science knowledge.

PO2: Research, Analysis, and Complexity

- The course develops ability to analyze complex real-world datasets, data quality issues and high dimensional data problems. CO3, CO4, CO5, CO6, CO7 and CO8 strongly support analytical and research skills, while CO1 and CO2 moderately support conceptual and dataset evaluation skills.

PO3: Problem Solving in New Contexts

- Students learn to solve real world data problems using preprocessing, similarity computation and transformation techniques. CO4, CO5, CO6, CO7 and CO8 strongly support real world problem solving, while CO1, CO2 and CO3 moderately support analytical problem solving.

PO4: Technical Mastery and Scientific Reasoning

- The course builds technical expertise in preprocessing, transformation, feature representation and similarity computations. CO3, CO6, CO7 and CO8 strongly support technical mastery, CO4 and CO5 moderately support technical implementation, while CO1 and CO2 partially support technical conceptual understanding.

PO5: Integrated Communication

- The course helps students interpret and communicate data analysis results through structured data representation and preprocessing insights. CO1, CO4, CO5 and CO7 strongly support data interpretation and communication, CO3, CO6 and CO8 moderately support technical explanation, while CO2 partially supports conceptual communication.

PO6: Ethical, Social, and Professional Judgment

- The course supports responsible data handling through data cleaning, preprocessing and data quality assessment. CO4 and CO5 strongly support ethical data usage, CO1, CO6, CO7 and CO8 moderately support responsible data handling, while CO2 and CO3 partially support professional practices.

PO7: Autonomous and Lifelong Learning

- The course develops strong base for advanced learning in machine learning, AI and analytics. CO6, CO7 and CO8 strongly support lifelong learning, while CO1, CO2, CO3, CO4 and CO5 moderately support continuous skill development.

PO8: Employability, Innovation, and Entrepreneurship

- The course builds practical industry skills in data preprocessing, feature engineering and similarity analysis. CO3, CO4, CO5, CO6, CO7 and CO8 strongly support employability and industry readiness, while CO1 and CO2 moderately support professional knowledge foundation

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: I
Course Type	: Major Elective Theory
Course Name	: Stochastic Models and Applications
Course Code	: DSC-506-MJE(B)
No. of Credits	: 2
No. of Teaching Hours	: 30

Course Objectives:

1. Students should acquire a fundamental understanding of stochastic processes, including the definition, types, and basic properties.
2. To understand discrete and continuous Markov chains models to compute the probability of events.
3. Formulate and solve problems by computing the long-term probabilities of a Markov chain model.
4. Write Python/R code to simulate Markov chains, and compute probabilities of events that may be difficult to derive by hand.
5. Apply Poisson processes to model the occurrence of events in various applications.
6. Students understand the practical relevance and utility of stochastic processes in modeling and analyzing complex systems.
7. Students should learn about the properties of Markov chains, including the Markov property, transition probabilities, stationary distributions, and ergodicity.

Course Outcomes:

By the end of the course, students will be able to:

CO1. formulate transition probability matrix, n-step transition probabilities

CO2. classify of states of Markov Chain.

CO3. familiar with Poisson process and its properties.

CO4. understanding of stationary distributions in the context of stochastic processes and their key properties.

CO5. explore applications of stationary distributions in various fields, including queueing theory, reliability analysis, and population dynamics.

CO6. develop a deep understanding of the definition and fundamental properties of Poisson process.

CO7. develop skills in using stochastic processes for modeling and forecasting future events and outcomes. Explain the fundamental principles of probability theory and random variables as they pertain to stochastic processes.

Topics and Learning Points

UNIT 1: (10 L)

Definition of a Stochastic process and notations, state space, parameter space, types of stochastic processes, Introduction Markov Chains (MC) $\{Xn, n \geq 0\}$, finite MC, time homogeneous MC one step transition probabilities, and transition probability matrix (t.p.m.), stochastic matrix, Chapman Kolmogorov equation, n-step transition probability matrix, initial distribution, joint distribution function of $\{X0, X1, \ldots, Xn\}$, partial sum of independent and identically distributed random variables as Markov Chain, illustrations such as random walk, Gambler's ruin problem, Ehrenfest chain.

UNIT 2: (10 L)

Classification of states: Communicating states, first return probability, probability of ever return Classification of states, as persistent and transient states. Decomposition of state space, closed set of states, irreducible set of states, irreducible MC, periodicity of M.C. aperiodic M.C. ergodic M.C.

UNIT 3: (10 L)

Stationary distribution for an irreducible ergodic finite M.C., Long run behaviour of a MC. Poisson process: Postulates and properties of Poisson process, probability distribution of N(t), the number of occurrences of the event in (0, t], Poisson process and probability distribution of inter-arrival time, mean, variance and covariance functions. Definition of compound Poisson.

References:

1. Bhat B.R. (2000). Stochastic Models: Analysis and Applications, New Age International.
2. Medhi, J. (2010) Stochastic Processes, New Age Science Ltd.
3. Pinsky M. A. and Karlin, S. (2010). An Introduction to Stochastic Modeling, 4thEdn. Academic Press.
4. Ross, S. (2014). Introduction to Probability Models, 11th Edn. Academic Press.

5. Feller, W. (1972). An Introduction to Probability Theory and its Applications, Vol. 1, Wiley Eastern.
6. Hoel, P.G. Port, S.C. & Stone, C.J. (1972). Introduction to Stochastic Processes, Houghton Mifflin
7. Karlin, S & Taylor, H.M. (1975). A First Course in Stochastic Processes (Second. Edition), Academic Press.
8. Serfozo, R. (2009). Basics of Applied Stochastic Processes, Springer.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)							
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	2	2	1	1	2	2
CO2	3	3	2	2	1	1	2	2
CO3	3	2	2	2	1	1	2	2
CO4	3	3	2	2	2	1	2	2
CO5	2	3	3	2	2	2	2	3
CO6	3	2	2	2	1	1	2	2
CO7	3	3	3	3	2	2	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

Justification for CO-PO Mapping

PO1: Advanced Disciplinary Knowledge & Originality

- **Strongly related (3)** with CO1, CO2, CO3, CO4, CO6, CO7
- These COs involve deep understanding of Markov Chains, stationary distributions, and Poisson processes — core advanced stochastic process concepts beyond undergraduate level.
- CO5 is moderately related (2) because it focuses more on applications rather than core theory.

PO2: Research, Analysis, and Complexity

- **Strong (3)** with CO1, CO2, CO4, CO5, CO7
- Formulating transition matrices and stationary distributions requires hypothesis formulation and analytical reasoning.

- CO5 (applications in queueing/reliability/population) involves handling complex real-world stochastic systems.
- CO3 and CO6 are moderately related (2) as they emphasize conceptual understanding more than research design.

PO3: Problem Solving in New Contexts

- Strong (3) with CO5 and CO7
- Application of stationary distributions and stochastic modeling in real-world domains (queueing, reliability) directly aligns with problem solving in new contexts.
- CO1–CO4 & CO6 moderately related (2) as they provide foundational tools used in applied settings.
-

PO4: Technical Mastery and Scientific Reasoning

- Strong (3) with CO7
- Modeling and forecasting require methodological selection and technical reasoning.
- CO1–CO6 moderately related (2) since they build technical foundations of stochastic models.
-

PO5: Integrated Communication

- **Moderately related (2)** with CO4, CO5, CO7
- Interpretation of stationary distributions and modeling outputs requires communicating mathematical conclusions clearly.
- Other COs partially related (1) as communication is implicit but not central.
-

PO6: Ethical, Social, and Professional Judgment

- Moderately related (2) with CO5 and CO7
- Applications in reliability and population dynamics have societal implications.
- Other COs partially related (1) as they are mainly theoretical.

PO7: Autonomous and Lifelong Learning

- **Strong (3)** with CO7
- Independent modeling and forecasting encourage self-directed learning.
- Other COs moderately related (2) as advanced stochastic concepts build foundation for research growth.

PO8: Employability, Innovation, and Entrepreneurship

- Strong (3) with CO5 and CO7

- Applications in queueing theory, reliability analysis, and forecasting have strong industry relevance (analytics, operations research, actuarial science).
- Other COs moderately related (2) as theoretical foundation supports professional excellence.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: I
Course Type	: Major Elective Practical
Course Name	: Practical Based on Fundamentals of Data Science
Course Code	: DSC-507 MJE(A)
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. To introduce learners to different types of data including structured, semi-structured, and unstructured data using real-world datasets.
2. To develop the ability to evaluate datasets using the fundamental characteristics of Big Data such as Volume, Velocity, and Variety.
3. To provide hands-on experience in implementing the complete Data Science lifecycle for solving practical problems.
4. To familiarize students with data representation techniques such as data matrices and dissimilarity matrices.
5. To enable understanding and computation of various proximity and distance measures for different types of attributes.
6. To develop skills for identifying and handling data quality issues such as missing values, noise, outliers, and duplicates.
7. To expose learners to feature engineering techniques including feature selection, feature extraction, and dimensionality reduction.

Course Outcomes:

1. After successful completion of the course, the student will be able to:

CO.1 Students will be able to explore, classify, and analyze structured, semi-structured, and unstructured datasets obtained from real-world sources.

CO.2 Students will be able to assess datasets based on the 3 V's of Data: Volume, Velocity, and Variety.

CO.3 Students will be able to design and implement an end-to-end Data Science workflow for a given real-world problem.

CO.4 Students will be able to construct and analyze data matrices and dissimilarity matrices for multivariate data.

CO.5 Students will be able to compute and compare proximity measures for nominal, binary, ordinal, and mixed-type attributes.

CO.6 Students will be able to apply and compare different distance measures such as Euclidean, Manhattan, and Minkowski for numerical data analysis.

CO.7 Students will be able to perform data quality assessment, data cleaning, outlier detection, and data transformation techniques effectively.

Topics and Learning Points

Sr. No.	Title of Experiment	No. of Experiment
1	Exploratory analysis and classification of structured, semi-structured, and unstructured datasets using real-world data sources	2
2	Evaluation of datasets using the 3 V's of Data (Volume, Velocity, Variety)	1
3	End-to-end implementation of the Data Science lifecycle for a real-world problem	1
4	Construction and analysis of data matrices and dissimilarity matrices for multivariate datasets	1
5	Computation and comparison of proximity measures for nominal, binary, ordinal, and mixed-type attributes	1
6	Implementation and comparison of distance measures (Euclidean, Manhattan, Minkowski) for numeric data	1
7	Comprehensive data quality assessment and data cleaning including treatment of missing values, noise, and inconsistencies	1
8	Detection and treatment of outliers, duplicate records, and invalid observations	1
9	Application and comparison of data transformation techniques such as normalization, standardization, and scaling	1
10	Data integration from multiple sources and conflict resolution	2
11	Feature selection, feature extraction, and dimensionality reduction techniques	3

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	2	1	3	2	2	3
CO2	2	2	2	1	1	1	2	2
CO3	3	3	3	2	3	2	3	3
CO4	3	3	2	3	2	1	2	2
CO5	3	3	3	3	2	1	3	3
CO6	3	3	3	3	2	1	3	3
CO7	2	3	3	2	3	3	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- The course builds advanced knowledge in data science concepts such as data types, proximity measures, multivariate data structures and data preprocessing. CO1, CO3, CO4, CO5 and CO6 strongly support disciplinary knowledge, while CO2 and CO7 moderately support applied data science understanding.

PO2: Research, Analysis, and Complexity

- Handling real-world datasets, multivariate matrices and workflow design requires strong analytical and research capability. CO3, CO4, CO5, CO6 and CO7 strongly support complex data analysis and research thinking, while CO1 and CO2 moderately support dataset understanding and evaluation.

PO3: Problem Solving in New Contexts

- The course focuses on solving real-world data science problems using workflows, distance measures and preprocessing techniques. CO3, CO5, CO6 and CO7 strongly support real-world problem solving, while CO1, CO2 and CO4 moderately support applied analytical problem solving.

PO4: Technical Mastery and Scientific Reasoning

- Students develop technical skills in data preprocessing, similarity computation and multivariate data handling. CO4, CO5 and CO6 strongly support technical reasoning and methodology selection, CO3 and CO7 moderately support technical implementation, while CO1 and CO2 partially support technical understanding.

PO5: Integrated Communication

- Data exploration, data quality analysis and workflow design support clear interpretation and communication of analytical results. CO1, CO3 and CO7 strongly support data communication, CO4, CO5 and CO6 moderately support technical explanation, while CO2 partially supports conceptual communication.

PO6: Ethical, Social, and Professional Judgment

- Data cleaning, data quality assessment and real-world data handling support ethical and responsible data usage. CO7 strongly supports ethical data practices, CO1 and CO3 moderately support responsible data handling, while CO2, CO4, CO5 and CO6 partially support professional analytical practices.

PO7: Autonomous and Lifelong Learning

- The course develops foundational and advanced skills needed for continuous learning in data science and analytics. CO3, CO5, CO6 and CO7 strongly support lifelong learning, while CO1, CO2 and CO4 moderately support analytical skill development.

PO8: Employability, Innovation, and Entrepreneurship

- The course develops practical industry skills such as workflow design, data preprocessing and similarity analysis. CO1, CO3, CO5, CO6 and CO7 strongly support employability and industry readiness, while CO2 and CO4 moderately support applied data science skills.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme : M.Sc. Data Science
Program Code : PSDSC
Class : M.Sc. (Part – I)
Semester : I
Course Type : Major Elective Practical
Course Name : Practical Based on Stochastic Models and Applications
Course Code : DSC-507 MJE(B)
No. of Credits : 2
No. of Teaching Hours : 60

Course Objectives:

1. To understand the theoretical foundations of stochastic processes, especially finite Markov chains and Poisson processes.
2. To develop simulation skills using R or Python for modeling stochastic systems.
3. To analyze transition probability matrices and interpret system behavior over time.
4. To verify theoretical results such as Chapman–Kolmogorov equations through computational experiments.
5. To study classical probability problems like Gambler's Ruin using simulation techniques.
6. To visualize stochastic processes such as random walks and interpret their statistical properties.
7. To perform statistical analysis of simulated data including inter-arrival time distributions.
8. To apply stochastic modeling concepts to real-life case studies and practical scenarios

Course Outcomes:

By the end of the course, students will be able to:

CO1. Explain the structure and properties of finite Markov chains and Poisson processes.

CO2. Construct and simulate Markov chain models using R or Python.

CO3. Compute and interpret transition probability matrices and steady-state distributions.

CO4. Validate stochastic process properties using numerical and simulation-based approaches.

CO5. Implement simulation-based solutions for classical stochastic problems.

CO6. Generate and visualize random walks and interpret long-run stochastic behavior.

CO7. Analyze simulated event data and evaluate inter-arrival time characteristics statistically.

CO8. Develop stochastic simulation models for real-world applications and case studies

Topics and Learning Points

Sr. no.	Title of Experiments
1	**S**imulation and Analysis of Finite Markov Chains using R / Python
2	Study of Gambler's Ruin Problem using Simulation
3	Verification of Chapman–Kolmogorov Equations Using Transition Probability Matrices (2 Practical)
5	Analysis of Transition Probability Matrix and Long-Run Behavior of Markov Chains (2 Practical)
7	Random Walk Simulation and Visualization using Markov Chains
8	Simulation and Statistical Analysis of Poisson Process and Inter-Arrival Times (2 Practical)
9	Simulation Study of Markov Chains and Poisson Processes with Real-Life Applications (2 Practical)
11	Case study (3 Practical)

Programme Outcomes and Course Outcomes Mapping:

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	2	1	1	1	2	2
CO2	3	2	3	3	2	1	3	3
CO3	3	3	2	2	3	1	2	2
CO4	3	3	3	3	2	1	3	2
CO5	2	3	3	3	2	1	2	3
CO6	3	2	3	2	3	1	2	3
CO7	2	3	2	2	3	2	2	3
CO8	3	3	3	3	2	2	3	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

CO – PO Mapping with Justification

PO1: Advanced Disciplinary Knowledge & Originality

- The course develops advanced knowledge of stochastic processes, Markov chains, Poisson processes and simulation modeling used in data science and research. CO1, CO2, CO3, CO4, CO6 and CO8 strongly support disciplinary knowledge, while CO5 and CO7 moderately support applied stochastic analysis.

PO2: Research, Analysis, and Complexity

- The course builds strong analytical and research skills through stochastic modeling, simulation validation and statistical analysis of event data. CO3, CO4, CO5, CO7 and CO8 strongly support research and complex data analysis, while CO1, CO2 and CO6 moderately support stochastic system understanding.

PO3: Problem Solving in New Contexts

- The course enables students to solve real world problems using stochastic models and simulation techniques. CO2, CO4, CO5, CO6 and CO8 strongly support real world problem solving, while CO1, CO3 and CO7 moderately support analytical problem solving.

PO4: Technical Mastery and Scientific Reasoning

- The course develops computational and simulation skills using programming tools and stochastic modeling techniques. CO2, CO4, CO5 and CO8 strongly support technical mastery, CO3, CO6 and CO7 moderately support scientific reasoning, while CO1 partially supports theoretical technical understanding.

PO5: Integrated Communication

- The course enables students to interpret stochastic results, simulation outputs and probabilistic insights effectively. CO3, CO6 and CO7 strongly support interpretation and communication, CO2, CO4, CO5 and CO8 moderately support technical explanation, while CO1 partially supports conceptual communication.

PO6: Ethical, Social, and Professional Judgment

- The course supports responsible use of stochastic models and simulation-based decision making. CO7 and CO8 moderately support responsible data and model interpretation, while CO1–CO6 partially support professional analytical responsibility.

PO7: Autonomous and Lifelong Learning

- The course builds strong mathematical and computational base for advanced learning in stochastic modeling, AI and simulation science. CO2, CO4 and CO8 strongly support lifelong learning, while CO1, CO3, CO5, CO6 and CO7 moderately support continuous skill development.

PO8: Employability, Innovation, and Entrepreneurship

- The course develops industry relevant skills in simulation modeling, stochastic analysis and computational data science. CO2, CO5, CO6, CO7 and CO8 strongly support employability and industry readiness, while CO1, CO3 and CO4 moderately support applied modeling skills.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2026 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: I
Course Type	: Research Methodology
Course Name	: Research Methodology
Course Code	: DSC-508-RM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To introduce the concept, purpose, and significance of research in data science and allied disciplines.
2. To enable students to identify, formulate, and define appropriate research problems and hypotheses.
3. To familiarize students with various research approaches, research designs, and data science research methods.
4. To impart knowledge of primary and secondary data collection methods and appropriate sampling techniques.
5. To develop skills in data pre-processing, cleaning, and preparation for analysis.
6. To train students in organizing, reviewing, and synthesizing scholarly literature effectively.
7. To provide comprehensive understanding of research report writing, journal articles, theses, and dissertations.
8. To inculcate ethical research practices and effective use of ICT tools and software for research and data management.

Course Outcomes:

By the end of the course, students will be able to:

CO1: Understand the concept, characteristics, significance, and applications of research in data science.

CO2: Identify research problems, formulate research hypotheses, and design structured questionnaires for data collection.

CO3: Apply appropriate primary and secondary data collection methods and sampling techniques for research studies.

CO4: Determine suitable sample sizes and select probability and non-probability sampling methods based on research objectives.

CO5: Perform data pre-processing and cleaning, including handling missing values, outliers, and data transformations.

CO6: Conduct effective literature reviews and organize scholarly information systematically.

CO7: Prepare structured research reports, journal articles, abstracts, theses, and dissertations using standard formats and referencing styles.

CO8: Demonstrate ethical research practices, avoid plagiarism, and effectively use ICT tools and software for data analysis, reference management, and research documentation.

Topics and Learning Points

Unit – 1 **(18L)**

Introduction to Research: The concept of research, characteristics of good research, Application of Research, Meaning and sources of Research problem, characteristics of good Research problem, Research process, outcomes, Meaning and types of Research hypothesis, Importance of Review of Literature, Organizing the Review of Literature. Research approaches, significance of research, overview of research methods in data science, defining the research problem, designing a questionnaire.

Unit – 2 **(12L)**

Data collection methods: Primary, secondary data collection methods, introduction to sampling techniques, finite population techniques SRSWR, SRSWOR, stratified sampling, systematic sampling, clustering , probability proportional to size with replacement(PPSWR) methods, Non probability sampling techniques, sample size determination.

Unit – 3 **(10L)**

Data pre-processing and cleaning, data pre-processing technique (missing values, imputation, outlier detection and treatment, data transformation), cleaning and preparing data set for analysis

Unit – 4 **(10L)**

Research Report: Research report and its structure, journal articles. Components of journal article. Explanation of various components. Structure of an abstract and keywords. Thesis and dissertations. Components of thesis and dissertations. Referencing styles and bibliography

Unit – 5 **(10L)**

Ethics in Research: Plagiarism - Definition, different forms, consequences, unintentional plagiarism, copyright infringement, collaborative work.

ICT Tools for Research : Role of computers in research, maintenance of data using software such as Mendeley, Endnote, Tabulation and graphical presentation of research data and software tools(MS-Excel, Tableau).

Use of tools or techniques for research: methods to search required information effectively, reference management software like Zotero/ Mendeley, software for paper formatting like LaTex/ MS office, software for detection of plagiarism.

References:

1. Des Raj &Chandhok P. (1998), Sample survey theory. (Narosa)
2. Murthy M.N. (1977) Sampling theory and methods. (Statistical Publishing Society)
3. Parimal Mukhopadhyay, Theory and methods of survey sampling, Prentice Hall of India private limited, 2nd Edition, 2008.
4. W.G.Cochran, (1977) Sampling techniques. (John Wiley and sons)
5. Sukhatme P.V. Sukhatme B.V. and C. Ashok Sampling theory of surveyand applications.(Indian society for Agricultural statistics)
6. Research Methodology: Methods and Techniques, Kothari C.R., 1990. New Age International.
7. An introduction to Research Methodology; Garg B.L., Karadia, R., Agarwal, F. and Agarwal, U.K., 2002. RBSA Publishers.
8. Research Methodology; Sinha S.C. and Dhiman, A.K., 2002. Ess Publications. 2 volumes.
9. Research Methods: the concise knowledge base; Trochim W.M.K., 2005. Atomic Dog Publishing. 270p.
10. Research Methodology; Panneerselvam R., PHI, Learning Pvt. Ltd., New Delhi – 2009.

Programme Outcomes and Course Outcomes Mapping:

COs \ POs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	1	1	1	1	1	1
CO2	2	3	2	2	2	1	2	1
CO3	2	3	2	3	1	1	1	2
CO4	2	3	2	3	1	1	1	2
CO5	2	3	2	3	1	1	1	2
CO6	3	3	1	1	2	1	2	1
CO7	2	2	1	1	3	1	2	1
CO8	2	2	1	2	2	3	2	2

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1: Advanced Disciplinary Knowledge & Originality

- Strongly related to CO1 and CO6 (3): These COs develop conceptual understanding of research and scholarly literature, building advanced disciplinary knowledge. Moderately related to CO2–CO5, CO7, CO8 (2): Application of research methods and reporting strengthens originality and specialization.

PO2: Research, Analysis, and Complexity

- Strongly related to CO2, CO3, CO4, CO5, CO6 **(3)**: Hypothesis formulation, sampling, data collection, pre-processing, and literature review directly reflect research competency and handling of complex data.
- Moderately related to CO1, CO7, CO8 (2): Understanding research basics, report writing, and ethical ICT usage support research ability.

PO3: Problem Solving in New Contexts

- Moderately related to CO2–CO5 (2): Designing research, sampling, and data handling require solving real-world research problems.
- Partially related to remaining COs (1): Conceptual understanding and reporting contribute indirectly.

PO4: Technical Mastery and Scientific Reasoning

- Strongly related to CO3, CO4, CO5 (3): Sampling design, data collection and preprocessing demonstrate technical proficiency and reasoning.
- Moderately related to CO2 and CO8 (2): Questionnaire design and ICT tools usage involve methodological reasoning.
- Partially related to CO1, CO6, CO7 (1): Limited technical execution.

PO5: Integrated Communication

- Strongly related to CO7 (3): Report writing, journal articles, and thesis preparation require high-level scientific communication.
- Moderately related to CO2, CO6, CO8 (2): Questionnaire design, literature review, and referencing enhance communication skills.
- Partially related to CO1, CO3–CO5 (1): Indirect role in communication.

PO6: Ethical, Social, and Professional Judgment

- Strongly related to CO8 (3**):** Ethical research practices and plagiarism avoidance are core components.
- Partially related to all other COs (1): Ethical considerations exist but are not the primary focus.

PO7: Autonomous and Lifelong Learning

- Moderately related to CO2, CO6, CO7, CO8 (2): Independent research design, literature review, and documentation promote self-learning.
- Partially related to CO1, CO3–CO5 (1): Learning is guided rather than autonomous.

PO8: Employability, Innovation, and Entrepreneurship

- Moderately related to CO3, CO4, CO5, CO8 (2): Data handling, sampling, and ICT skills enhance employability and innovation.
- Partially related to remaining COs (1): Foundational research knowledge supports professional growth indirectly.