

Savitribai Phule Pune University

(Formerly University of Pune)

Bachelors Degree in Data Science

(Faculty of Science and Technology)



Syllabi for B.Sc. (Data Science)

(For Colleges Affiliated to Savitribai Phule Pune University)

Choice Based Credit System (CBCS) Syllabus
Under National Education Policy (NEP)

To be implemented from Academic Year 2024-2025

Preamble

Welcome to the B.Sc.(Data Science) programme! This programme is designed to empower students with knowledge and skills required to thrive in an era of data science and technology. By choosing B.Sc. (Data Science) Programme, students enter into the dynamic field of data science and data analytics. Students will engage and build strong foundation in mathematics, statistics, computer science and ethical data practices. This programme not only equips students with technical expertise but also fosters a mindset of continuous learning, adaptability and ethical leadership.

As you navigate this syllabus, consider it a roadmap to your future in data science. Welcome to the world where data becomes insight and insight drives innovation.

Eligibility

- (a) Higher Secondary School Certificate (HSC) (10+2) with any stream or its equivalent examination having Mathematics/ Statistics/ Information Technology(IT) at 10+2 level.

Note: If candidates having Information Technology (IT) at 10+2 level and not offered Mathematics/Statistics as one of the subject then such candidates should have to complete the Bridge Course (Specified in the syllabus) before completion of semester-I.

OR

- (b) Three Years Diploma Course after S.S.C. (10th standard) of Board of Technical Education conducted by Government of Maharashtra or its equivalent.

Programme Outcomes:

- PO 1: The programme seeks to develop strong foundation in Mathematics, Statistics and Computer Science that demonstrate proficiency in basic programming languages and tools.
- PO 2: The programme aims to understand the principles of data storage and retrieval by acquiring knowledge of data type structures and basic data manipulation techniques.
- PO 3: The programme helps to learn database management techniques with design and management of databases as well as executing SQL queries for data retrieval and manipulation.
- PO 4: By applying advanced statistical methods and machine learning techniques, the students can analyze complex datasets, interpret and communicate findings effectively.
- PO 5: The programme also aims to understand and work with big data technologies and apply these technologies to process and analyze large-scale datasets.
- PO 6: The students can create clear and effective data visualizations using various tools and communicate complex findings through visual representations.

- PO 7: The programme also seeks to develop comprehensive projects by applying data science techniques to solve real-world problems that will improve the ability of learner to integrate knowledge and skills acquired throughout the programme.
- PO 8: Through hands-on projects, practical assignments, and exposure to state-of-the-art tools and technologies, programme aim to develop the technical proficiency and problem-solving skills necessary for success in the professional world.
- PO 9: Depending on the chosen track, students can develop expertise in data analytics with areas such as Business, Social Media, HR, Financial, Healthcare, Supply Chain & Logistics and Big Data etc.
- PO 10: The program include On Job Training, internships and research work that provides learners with practical experience, applying their knowledge to real-world challenges.
- PO 11: Graduates will be adept at presenting complex technical concepts clearly and effectively, both in written and oral forms, to various audiences.
- PO 12: The programme places a strong emphasis on ethical considerations, responsible use of technology, and awareness of the societal impact of data science and computing solutions.
- PO 13: The programme aim to produce graduates who approach their work with integrity and a sense of social responsibility.
- PO 14: Acknowledging the dynamic nature of computer science, the programme aim to inspire students for continuous learning and professional development, empowering them to adapt and thrive in the face of technological advancements; prepared them to adapt to new technologies and methodologies throughout their careers.
- PO 15: The students will be encouraged to think creatively and innovatively, exploring new ideas and approaches to solve data science related problems and advance the state of the art in the field.

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

Bridge Course : Foundation of Mathematics and Statistics for Data Science

(This course is to be offered only by students having Information Technology (IT) at 10+2 level and not offered Mathematics/Statistics as one of the subject. Note that this Bridge Course must be completed by students before completion of semester-I)

No. of Credits: 2		Teaching Scheme Theory:2 Hrs /Week	Examination Scheme Continuous Internal Evaluation:50 Marks	
Objective : To learn and understand the fundamental concepts of Mathematics and Statistics required for Data Science				
Course Outcomes On Completion of this course, student will be able to – CO1: Understand concepts of determinants and matrices CO2: Learn fundamental measures of dispersion and probability CO3: Learn concept of permutations and combinations CO4: Understand working of sets and relations CO5: Construct equations of line, vector and plane				
Unit No.	Name of Unit	Teaching Hours	CO Targeted	
1	Determinants and Matrices	8	CO1	
1.1 Find value of determinants 1.2 Determinants of order 3 and expansion of determinants 1.3 Minors and cofactors of elements of determinants 1.4 Introduction to matrices, Types of matrices 1.5 Algebra of matrices – equality, multiplication by scalar, addition and multiplication of two matrices, transpose of matrix.				
2	Measures of Dispersion and Probability	5	CO2	
2.1 Measures of Dispersion – Range, Variance, Standard Deviation 2.2 Probability – Concept, Conditional probability, Baye’s Theorem				
3	Permutations and Combinations	5	CO3	
3.1 Factorial Notation 3.2 Permutation 3.3 Combination				
4	Sets and Relations	6	CO4	
4.1 Introduction to sets, Representation of sets 4.2 Types of sets, operations on sets 4.3 Introduction to relations				
5	Line, Vector and Plane	6	CO5	

5.1 Locus, Straight line, Equation of line in standard form
5.2 Representation of vectors, Types of vectors, Algebra of vectors
5.3 Introduction to plane, Equation of plane
Reference Books
1. Mathematics and Statistics (Arts and Science), Standard XI, Part-I and II 2. Mathematics and Statistics (Arts and Science), Standard XII, Part-I and II By Maharashtra State Board of Text Book Production and Curriculum Research, Pune
Evaluation Scheme
• The evaluation of this bridge course will be at college level for 50 marks. • Evaluation records are to be maintained by the college itself.

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

FY (Level 4.5) SEMESTER I

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Subject-1	DS-101-T	Problem Solving and Python Programming	2	-	2	-	15	35	50
	DS-102-P	Lab Course on DS-101-T (Python Programming)	-	2	-	4	15	35	50
Subject-2	DS-103-T	Descriptive Statistics	2	-	2	-	15	35	50
	DS-104-P	Lab Course on DS-103-T (Descriptive Statistics)	-	2	-	4	15	35	50
Subject-3	DS-105-T	Computational Mathematics	2	-	2	-	15	35	50
	DS-106-P	Lab Course on DS-105-T (Computational Mathematics)	-	2	-	4	15	35	50
GE/OE *	OE-101-DS-T* OE-102-DS-T* OE-103-DS-T* OE-104-DS-T*	• Office Automation I/ • Introduction to Computers and Basics of Internet / • Introduction to Google Apps I/ • Fundamentals of Computers I	2	-	2	-	15	35	50
SEC	SEC-101-DS-T	Computer Organization	2	-	2	-	15	35	50
IKS	IKS-100-T	Generic IKS	2	-	2	-	15	35	50
AEC	AEC-101-ENG	English	2	-	2	-	15	35	50
VEC	VEC-101- ENV	EVS-I	2	-	2	-	15	35	50
Total			16	6	16	12			550

** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.*

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

FY (Level 4.5) SEMESTER II

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Subject-1	DS-151-T	Advanced Python Programming	2	-	2	-	15	35	50
	DS-152-P	Lab Course on DS-151-T (Advanced Python Programming)	-	2	-	4	15	35	50
Subject-2	DS-153-T	Discrete Probability and Probability Distributions	2	-	2	-	15	35	50
	DS-154-P	Lab Course on DS-153-T (Discrete Probability and Probability Distributions)	-	2	-	4	15	35	50
Subject-3	DS-155-T	Graph Theory	2	-	2	-	15	35	50
	DS-156-P	Lab Course on DS-155-T (Graph Theory)	-	2	-	4	15	35	50
GE/OE *	OE-151-DS-T* OE-152-DS-T* OE-153-DS-T* OE-154-DS-T* OE-155-DS-T* OE-156-DS-T*	• Office Automation II/ • Computer Fundamentals / • Introduction to Google Apps II/ • Fundamentals of Computers II/ • Introduction to Data Science/ • AI Tools for Business	2	-	2	-	15	35	50
SEC	SEC-151-DS-P	Lab Course on Excel and Advanced Excel	-	2	-	4	15	35	50
AEC	AEC-151-ENG	English	2	-	2	-	15	35	50
VEC	VEC-151- ENV	EVS-II	2	-	2	-	15	35	50
CC	CC-151-T	From University Basket	2	-	2	-	15	35	50
Total			14	8	14	16			550

*** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.**

Exit option: Award of UG Certification in B.Sc. (Data Science) with 44 credits and an additional 04 credits as per University guidelines or else continue with Major and Minor.

Continue option: In second year Data Science will be the Major and Statistics as a Minor subject.

Savitribai Phule Pune University
Syllabus Structure as per NEP Guidelines
B.Sc. (Data Science) from 2024-25
SY (Level 5.0) SEMESTER III

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	C E	E E	Total
Major Core	DS-201-MJ-T	Database Management System	2	-	2	-	15	35	50
	DS-202-MJ-T	Data Structure-I	2	-	2	-	15	35	50
	DS-203-MJ-P	Lab Course on DS-201-MJ-T and DS-202-MJ-T	-	2	-	4	15	35	50
VSC	DS-221-VSC-T	Foundations of Data Science	2	-	2	-	15	35	50
FP/ OJT/ CEP	DS-231-FP	Mini Project	-	2	-	4	15	35	50
Minor	DS-241-MN-T	Probability Distribution and Modelling	2	-	2	-	15	35	50
	DS-242-MN-P	Lab Course on DS-241-MN-T	-	2	-	4	15	35	50
GE/OE *	OE-201-DS-T* OE-202-DS-T* OE-203-DS-T* OE-204-DS-T*	• E commerce-I / • Web Design-I/ • Digital Marketing-I/ • AI for Everyone-I	2	-	2	-	15	35	50
IKS	DS-200-IKS	Indian Knowledge System in Computing	2	-	2	-	15	35	50
AEC	AEC-201-T	From University Basket	2	-	2	-	15	35	50
CC	CC-201-T	From University Basket	2	-	2	-	15	35	50
Total			16	6	16	12			550

** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.*

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

SY (Level 5.0) SEMESTER IV

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	C E	E E	Total
Major Core	DS-251-MJ-T	Relational Database Management System	2	-	2	-	15	35	50
	DS-252-MJ-T	Data Structure-II	2	-	2	-	15	35	50
	DS-253-MJ-P	Lab Course on DS-251-MJ-T and DS-252-MJ-T	-	2	-	4	15	35	50
VSC	DS-271-VSC-P	Data Analytics	-	2	-	4	15	35	50
FP/OJT/CEP	DS-281-FP	Mini Project	-	2	-	4	15	35	50
Minor	DS-291-MN-T	Testing of Hypothesis and Sampling Distributions	2	-	2	-	15	35	50
	DS-292-MN-P	Lab Course on DS-291-MN-T	-	2	-	4	15	35	50
GE/OE *	OE-251-DS-T* OE-252-DS-T* OE-253-DS-T* OE-254-DS-T*	• E commerce-II / • Web Design-II / • Digital Marketing-II/ • AI for Everyone-II	2	-	2	-	15	35	50
SEC	SEC-251-DS-T	Software Engineering	2	-	2	-	15	35	50
AEC	AEC-251-T	From University Basket	2	-	2	-	15	35	50
CC	CC-251-T	From University Basket	2	-	2	-	15	35	50
Total			14	8	14	16			550

*** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.**

Exit option: Award of UG Diploma in B.Sc. (Data Science) with 88 credits and an additional 04 credits as per University guidelines or else continue with Major and Minor.

Continue option: Third year will be continued with Data Science as Major and Statistics as a Minor subject.

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

TY (Level 5.5) SEMESTER V

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	C E	E E	Total
Major Core	DS-301-MJ-T	NoSQL databases	4	-	4	-	30	70	100
	DS-302-MJ-T	R Programming	2	-	2	-	15	35	50
	DS-303-MJ-T	Foundations of Artificial Intelligence	2	-	2	-	15	35	50
	DS-304-MJ-P	Lab Course on DS-301-MJ-T (NoSQL databases)	-	2	-	4	15	35	50
	DS-305-MJ-P	Lab Course on DS-302-MJ-T (R Programming)	-	2	-	4	15	35	50
Major Elective	DS-310-MJ-T	Business Analytics	2	-	2	-	15	35	50
	DS-311-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR					-			
	DS-312-MJ-T	Social Media Analytics	2	-	2	-	15	35	50
	DS-313-MJ-P	Lab Course	-	2	-	4	15	35	50
VSC	DS-321-VSC-P	Lab Course on MATLAB	-	2	-	4	15	35	50
FP/CEP	DS-331-FP	Project	-	2	-	4	15	35	50
Minor	DS-341-MN-T	Categorical and Multivariate Data Analysis	2	-	2	-	15	35	50
Total			12	10	12	20			550

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

TY (Level 5.5) SEMESTER VI

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	PR	TH	PR	CE	E E	Total
Major Core	DS-351-MJ-T	Data Visualization and Modelling	4	-	4	-	30	70	100
	DS-352-MJ-T	Artificial Intelligence in Data Science	2	-	2	-	15	35	50
	DS-353-MJ-T	Data Security and Privacy	2	-	2	-	15	35	50
	DS-354-MJ-P	Lab on DS-351-MJ-T	-	2	-	4	15	35	50
	DS-355-MJ-P	Lab Course on DS-352-MJ-T	-	2	-	4	15	35	50
Major Elective	DS-360-MJ-T	HR Analytics	2	-	2	-	15	35	50
	DS-361-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR								
	DS-362-MJ-T	Financial Analytics	2	-	2	-	15	35	50
	DS-363-MJ-P	Lab Course	-	2	-	4	15	35	50
VSC	DS-371-VSC-P	Advance Data Science Tools	-	2	-	4	15	35	50
OJT	DS-381-OJT	On Job Training	-	4	-	8	30	70	100
Total			10	12	10	24			550

Exit option: Award of Bachelor of Data Science (B.Sc. Data Science) with Statistics as a minor with 132 credits or else continue.

Continue option: Fourth year will be continued for B.Sc.(Data Science) Honors with Research.

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25

(Level 6.0) SEMESTER VII (Honors with Research Degree)

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	C E	E E	Total
Major Core	DS-401-MJ-T	Machine Learning	4	-	4	-	30	70	100
	DS-402-MJ-T	Basics of Cloud Computing	2	-	2	-	15	35	50
	DS-403-MJ-P	Lab Course on DS-401-MJ-T	-	2	-	4	15	35	50
	DS-404-MJ-P	Lab Course on DS-402-MJ-T	-	2	-	4	15	35	50
Major Elective	DS-410-MJ-T	Supply Chain & Logistics Analytics	2	-	2	-	15	35	50
	DS-411-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR								
	DS-412-MJ-T	Healthcare Analytics	2	-	2	-	15	35	50
	DS-413-MJ-P	Lab Course	-	2	-	4	15	35	50
RP	DS-431-RP	Research Project	-	4	-	8	30	70	100
RM	DS-441-RM	Research Methodology	4	-	4	-	30	70	100
Total			12	10	12	20			550

OR

(Level 6.0) SEMESTER VII (Honors Degree)

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	C E	E E	Total
Major Core	DS-401-MJ-T	Machine Learning	4	-	4	-	30	70	100
	DS-402-MJ-T	Basics of Cloud Computing	2	-	2	-	15	35	50
	DS-403-MJ-P	Lab Course on DS-401-MJ-T	-	2	-	4	15	35	50
	DS-404-MJ-P	Lab Course on DS-402-MJ-T	-	2	-	4	15	35	50
	DS-405-MJ-T	Big Data Analytics	4	-	4	-	30	70	100
Major Elective	DS-410-MJ-T	Supply Chain & Logistics Analytics	2	-	2	-	15	35	50
	DS-411-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR								
	DS-412-MJ-P	Healthcare Analytics	2	-	2	-	15	35	50
	DS-413-MJ-P	Lab Course	-	2	-	4	15	35	50
RM	DS-441-RM	Research Methodology	4	-	4	-	30	70	100
Total			16	6	16	12			550

Savitribai Phule Pune University
Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2024-25
(Level 6.0) SEMESTER VIII (Honors with Research Degree)

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	C E	EE	Total
Major Core	DS-451-MJ-T	Data Mining and Warehousing	4	-	4	-	30	70	100
	DS-452-MJ-T	Deep Learning	2	-	2	-	15	35	50
	DS-453-MJ-P	Lab Course on DS-451-MJ-T	-	2	-	4	15	35	50
	DS-454-MJ-P	Lab Course on DS-452-MJ-T	-	2	-	4	15	35	50
Major Elective	DS-460-MJ-T	Geospatial Technology	2	-	2	-	15	35	50
	DS-461-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR								
	DS-462-MJ-T	E-Commerce	2	-	2	-	15	35	50
	DS-463-MJ-P	Lab Course	-	2	-	4	15	35	50
RP	DS-481-RP	Research Project	-	8	-	16	60	140	200
Total			8	14	8	28			550

Award of Bachelor of Data Science (B.Sc.Data Science) Honors with Research degree with 176 credits

OR

(Level 6.0) SEMESTER VIII (Honors Degree)

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	C E	E E	Total
Major Core	DS-451-MJ-T	Data Mining and Warehousing	4	-	4	-	30	70	100
	DS-452-MJ-T	Deep Learning	2	-	2	-	15	35	50
	DS-453-MJ-P	Lab Course on DS-451-MJ-T	-	2	-	4	15	35	50
	DS-454-MJ-P	Lab Course on DS-452-MJ-T	-	2	-	4	15	35	50
	DS-455-MJ-T	Natural Language Processing	4	-	4	-	30	70	100
Major Elective	DS-456-MJ-T	Geospatial Technology	2	-	2	-	15	35	50
	DS-457-MJ-P	Lab Course	-	2	-	4	15	35	50
	OR								
	DS-458-MJ-T	E-Commerce	2	-	2	-	15	35	50
	DS-459-MJ-P	Lab Course	-	2	-	4	15	35	50
OJT	DS-481-OJT	On Job Training	-	4	-	8	30	70	100
Total			12	10					550

Award of Bachelor of Data Science (B.Sc. Data Science) Honors degree with 176 credits

Detail Syllabus

B.Sc. (Data Science)

Semester-I

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-101-T : Problem Solving and Python Programming

No. of Credits: 2	Teaching Scheme Theory:2 Hrs /Week	Examination Scheme Continuous Evaluation:15 Marks End Semester:35 Marks	
Prerequisites • Basic knowledge of mathematics, logic. • Puzzle solving Aptitude • Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage			
Objectives • To teach students systematic and efficient problem-solving methods, including problem analysis, algorithm design, and solution implementation. • To provide a solid understanding of the Python programming language, including its syntax, data types, control structures, and functions. • To instill good programming habits, including code readability, commenting, and documentation. • To nurture the ability to think algorithmically and express solutions as step-by-step processes using Python programs. • To learn and understand Object Oriented Programming • To improve debugging techniques and error identification and correction in Python programs.			
Course Outcomes On Completion of this course, student will be able to – CO1: Create clear and efficient algorithms for solving a variety of problems. CO2: Write Python programs to implement algorithms and solve problems. CO3: Identify and correct errors in Python programs using systematic debugging techniques. CO4: Understand Object Oriented Concepts in Python CO5: Learn and understand modules and packages in Python CO6: Define and demonstrate the use of built-in data structures “lists” and “dictionary”.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Problem Solving	5	CO 1
1.1 What is problem solving?			
1.2 Problem solving steps.			
1.3 Algorithms - definition, characteristics, examples, advantages and limitations.			
1.4 Flowcharts - definition, notations, examples, advantages and limitations, Comparison with algorithms.			

1.5 Pseudo codes - notations, examples, advantages and limitations. 1.6 Introduction to Programming 1.7 Programming Languages as tools, programming paradigms, types of languages 1.8 Converting pseudo-code to programs.			
2	Introduction to Python	10	CO1, CO2, CO3, CO6
2.1 History, feature of Python, setting up path, working with python Interpreter, basic syntax, variable and data types, operators 2.2 Conditional statements-If, If-Else, nested if-else, Examples. 2.3 Looping-For, While, Nested loops, Examples 2.4 Control Statements-Break, Continue, Pass. 2.5 String Manipulation-Accessing String, Basic Operations, String Slices, Function and Methods, Examples. 2.6 Lists-Introduction, accessing list, operations, working with lists, function & methods. 2.7 Tuple-Introduction, accessing tuples, operations working, function & methods, Examples. 2.8 Dictionaries-Introduction, Accessing values in dictionaries, working with dictionaries, properties, function, Examples. 2.9 Functions-Defining a function, Calling a function, types of function, function arguments, anonymous function, global & local variable, Examples			
3	Classes, Objects and Inheritance	8	CO4
3.1 Classes and Objects 3.1.1 Classes as User Defined Data Type 3.1.2 Objects as Instances of Classes 3.1.3 Creating Class and Objects 3.1.4 Creating Objects By Passing Values 3.1.5 Variables & Methods in a Class 3.2 Inheritance 3.2.1 Single Inheritance 3.2.2 Multilevel 3.2.3 Multiple Inheritance 3.2.4 Hybrid Inheritance 3.2.5 Hierarchical Inheritance 3.2.6 IS-A Relationship and HAS-A Relationship			
4	Modules and Packages	7	CO5
4.1 Built in Modules 4.1.1 Importing modules in python program 4.1.2 Working with Random Modules. 4.1.3 E.g. - built-ins, time, date time, calendar, sys, etc 4.2 User Defined functions 4.2.1 Structure of Python Modules 4.3 Packages 4.3.1 Predefined Packages 4.3.2 User defined Packages			
Reference Books			
1. How to solve it by Computer, R.G. Dromey, Pearson Education.			

2. Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010
3. Dive into Python, Mike
4. Learning Python, 4th Edition by Mark Lutz
5. Programming Python, 4th Edition by Mark Lutz
6. Python Programming: An introduction to computer, John Zelle, 3rd Edition.
7. Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016
8. Introduction to Python Programming. Gowrishankar S., Veena A. CRC Press, Taylor & Francis Group, 2019

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-102-P : Lab Course on DS-101-T (Python Programming)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks	
Prerequisites - Basic knowledge of logic and Python programming concepts - Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage			
Objectives - Learn Programming fundamentals using Python - Understand the concepts and usage data types, variables and other basic elements - Learn about using operators and control statements in Python - Learn about using arrays and strings in Python. - Learn Object Oriented concepts in Python. - Learn how to use modules in packages in Python Programming			
Course Outcomes On Completion of this course, students will be able to - CO1: Implement the use of built-in data structures “lists” and “dictionary” “Tuples” and “Sets”. CO2; Implement programs on Arrays and Strings CO3: Implement programs on Object Oriented concepts in Python. CO4: Implement programs by importing modules and packages in Python.			
Practical Implementation: Implementation of Practical Assignments for all courses in B.Sc.(Data Science) will be Platform Independent.			
Unit No.	Name of Unit	Hours	CO Targeted
1	Introduction to Python Language	8	CO1
a. Write a Python program to explore various data types including numeric types, Boolean types and compound types. b. Write a Python program to perform Input and Output Operations. c. Write a Python program to demonstrate looping in python and use of break statement and continue statement.			
2	Functions	8	CO1
a. Write a Python program to define and use functions.			

b. Write a Python program to demonstrate the use of Built-in Functions. c. Write a Python Program to implement Lambda Functions.			
3	List, Tuples, Dictionaries and Sets	12	CO1
a. Write a Python Program to create list, apply various functions to it. b. Write a Python Program to demonstrate concept of aliasing and cloning. c. Write a Python Program to implement tuples for storing data. Verify the immutability property on tuples. d. Write a Python Program to implement Dictionary and operations on dictionaries. e. Write a Python Program to create sets and various operations on it.			
4	Arrays and String	12	CO2
a. Write a Python Program to implement arrays for storing homogeneous data items. b. Apply indexing and slicing operations to access elements of array. c. Write a Python Program to demonstrate operations and properties of string data types. d. Write a Python Program implement and demonstrate the use of Membership operators and Identity operators			
5	Object Oriented Programming	12	CO3
a. Write a Python program to define classes and create objects. b. Program to implement the inheritance. c. Program to implement the polymorphism.			
6	Modules and Packages	8	CO4
a. Write Python program to import built in and user defined modules. b. Write Python program to import built in and user defined packages.			
Reference Books			
1. How to solve it by Computer, R.G. Dromey, Pearson Education. 2. Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010 3. Dive into Python, Mike 4. Learning Python, 4th Edition by Mark Lutz 5. Programming Python, 4th Edition by Mark Lutz 6. Python Programming: An introduction to computer, John Zelle, 3rd Edition. 7. Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016 8. Introduction to Python Programming. Gowrishankar S., Veena A. CRC Press, Taylor & Francis Group, 2019			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-103-T : Descriptive Statistics

No. of Credits: 02	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Mathematical operations			
Objectives - To acquaint students with some basic concepts in Statistics - To introduce to some elementary statistical methods of analysis of data - To identify the nature and type of data - To apply statistical tools to numerical and categorical data			
Course Outcomes On Completion of this course, student will be able to – CO1: Identify the different types of variables and data. CO2:Compute various measures of central tendency, dispersion, CO3: Compute various measures of skewness and kurtosis. CO4: Find correlation coefficient between numerical variables. CO5: Fit linear regression lines. CO6: Fit non-linear regression lines.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Statistics	02	CO1
Meaning of Statistics and its importance in data science. Concept of population and sample. Types of characteristics (variables and attributes), Types of data (primary and secondary). Raw data and its classification. Ungrouped frequency distribution, grouped frequency distribution and cumulative frequency distribution.			
2	Measures of Central Tendency and Dispersion	10	CO2
Measures of central tendency: Concept of central tendency of statistical data. Statistical averages: Arithmetic mean (Definition, effect of change of origin and scale), Geometric Mean and Harmonic Mean, median and mode, partition values (Definitions and examples for ungrouped and grouped data). Situations where one kind of average is preferable to other.			
Measures of dispersion: Concept of dispersion. Range, Semi-interquartile range (Quartile deviation): Definition. Mean deviation: Definition, minimality property (without proof), Variance and standard deviation: Definition, effect of change of origin and scale. Mean squared deviation: Definition, minimality property of mean squared deviation (without proof), Measures of dispersion for comparison: coefficient of range, coefficient of quartile deviation and coefficient of mean deviation, coefficient of variation(C.V.).			

3	Moments, Skewness and Kurtosis	04	CO3
Moments: Raw moments and Central Moments (Definition for for ungrouped and grouped data). Relation between Raw moments and Central Moments (upto 4th order without proof). Skewness: Concept of skewness of frequency distribution, positive skewness, negative skewness, symmetric frequency distribution. Bowley's coefficient of skewness(Definition and Examples and Bowley's coefficient of skewness lies between -1 to 1 (without proof)). Karl Pearson's coefficient of skewness (Definition and Examples). Measures of skewness based on moments (Definition and Examples). Kurtosis: Concept of kurtosis, leptokurtic, mesokurtic and platykurtic frequency distributions. Measures of kurtosis based on moments (Definition and Examples).			
4	Correlation and Regression	10	CO4, CO5
Correlation: Bivariate data, Scatter diagram and its interpretation. Concept of Covariance and its properties. Correlation between two variables and its types. Karl Pearson's coefficient of correlation (r) and its computation for ungrouped data. Properties of correlation. Spearman's rank correlation coefficient and its computation. Regression: Concept of dependent (response) and independent (predictor or regerssor) variables. Meaning of regression, connection between correlation and regression. Fitting of line $Y = \beta_0 + \beta_1 X$, β_0 and β_1 are regression coefficients which are estimated using least-square method. Properties of regression coefficients. Concept of explained and unexplained variation, coefficient of determination, standard error of an estimate of line of regression. Concept of reverse regression.			
5	Non-linear Regression	04	CO6
Necessity and importance of fitting of non-linear regression. Fitting of second degree curve($Y = a + bX + cX^2$), Fitting of exponential curves of the type $Y = ab^x$ and $Y = ax^b$. Fitting of logistic curve.			
Reference Books			
1. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta. 2. Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi. 3. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson. 4. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, NewDelhi. 5. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentice Hall of India, NewDelhi. 6. W. and Cochran W. G.(1989). Statistical Methods, Eighth Ed. EastWest Press. 7. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill BookCompany.			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-104-P : Lab Course on DS-103-T (Descriptive Statistics)

No. of Credits: 02	Teaching Scheme Practical: 4 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks
Prerequisites Mathematical operations		
Objectives - To acquaint students with some basic concepts in Statistics - To introduce to some elementary statistical methods of analysis of data - To identify the nature and type of data - To apply statistical tools to numerical and categorical data		
Course Outcomes On Completion of this course, student will be able to – CO1: Identify the different types of variables and data. CO2: Compute various measures of central tendency, dispersion, CO3: Compute various measures of skewness and kurtosis. CO4: Find correlation coefficient between numerical variables. CO5: Fit linear regression lines. CO6: Fit non-linear regression lines.		
Sr.No.	List of Practical Assignments	Hours
1	Diagrammatic representation and interpretation of statistical data.	4
2	Graphical representation and interpretation of statistical data	4
3	Tabulation	4
4	Computation of measures of central tendency for grouped and ungrouped data	4
5	Computation of partition values for grouped and ungrouped data.	4
6	Computation of measures of dispersion for grouped and ungrouped data	4
7	Identification the nature of probability distribution based on measure of skewness and kurtosis.	4
8	Plotting of Scatter diagram and computation of correlation coefficient (ungrouped data).	8
9	Computation of Spearman's Rank correlation coefficient.	4
10	Fitting of simple linear regression model (for both cases Y on X and X on Y).	8
11	Fitting of second degree curve.	4
12	Fitting of exponential curve of type $Y = ab^x$, $Y = ax^b$.	8
Reference Books		

1. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta.
2. Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi.
3. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
4. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
5. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentice Hall of India, New Delhi.
6. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. EastWest Press.
7. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-105-T : Computational Mathematics

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks	
Prerequisites Basic Mathematics Skills			
Objectives - To understand the basic arithmetic operations on vectors and matrices, including determinants, using technology where appropriate. - To solving systems of linear equations, using technology to facilitate row reduction. - To understand the basic terminology of linear algebra in Euclidean spaces, including linear independence, spanning, basis, rank, nullity, subspace, and linear transformation. - To abstract notions of vector space and inner product space. - To understand find the eigen values and eigenvectors of a matrix or a linear transformation, and using them to diagonalize a matrix. - Enables to find projections and orthogonality among Euclidean vectors, including the Gram-Schmidt ortho normalization process and orthogonal matrices.			
Course Outcomes On Completion of this course, student will be able to - CO1: Solve systems of linear equations using methods by Gaussian elimination. CO2: Demonstrate understanding of the concepts of vector space, linear independence and basis. CO3: Determine eigen values and eigenvectors and solve eigenvalue problems. CO4: Demonstrate understanding the use of truth tables and laws of identity, distributive, commutative, and domination. CO5: Simplify and prove Boolean expressions, Compute sum of products and product of sum expansions.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	System of Linear Equation	6	CO1
Matrices, Determinants, Cramer’s Rule, Echelon form, Row reduction, Gaussian elimination method.			
2	Vector Spaces	8	CO1, CO2
Introduction to vector spaces, Some properties of vector spaces, Linear combination, Linear independence, Linear dependence, Basis and Dimension of a vector space, Row space, Column space.			
3	Eigen values and Eigen vectors	8	CO3
Eigen values and Eigen vectors, The characteristic equation, Diagonalization.			
4	Boolean function	8	CO4,CO5

Relations, Types of Relations, Equivalence relations, Digraphs of relations, Matrix representation and Composition of Relations, Transitive closure and Warshall's Algorithm, Poset, Hasse diagram, Boolean Functions : Introduction, Boolean variable, Boolean Function of degree n, Boolean identities, Definition of Boolean Algebra, Representation of Boolean Functions : Minterm, Maxterm Disjunctive normal form, Conjunctive normal Form.

Reference Books

1. Howard Anton, Chris Rorres, Elementary Linear Algebra, Application Version, Ninth Edition, Wiley, 11th edition.
2. K. Hoffman and R. Kunze, Linear Algebra, 2nd edition(2014), Prentice Hall of India, New Delhi.
3. Steven J. Leon, Linear Algebra with Applications, 4th edition(1994), Prentice Hall of India. New Delhi.
4. Discrete Mathematical Structures, by Kolman, Busby, Ross, Rehman, Prentice Hall

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

DS-106-P: Lab Course on DS-105-T (Computational Mathematics)

No. of Credits: 2	Teaching Scheme Practical : 4 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks
Prerequisites Basic Mathematic Skills		
Objectives - To understand the basic arithmetic operations on vectors and matrices, including determinants, using technology where appropriate. - To solve systems of linear equations, using software to facilitate row reduction. - To understand the basic terminology of linear algebra in Euclidean spaces, including linear independence, spanning, basis. - To abstract notions of vector space and inner product space. - To understand find the eigenvalues and eigenvectors of a matrix and using them to diagonalize a matrix. - Enables to Simplify and prove Boolean expressions. Compute sum of products and product of sum expansions. - To know how to use maxima software.		
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the systems of linear equations using methods by Gaussian elimination. CO2: Demonstrate understanding of the concepts of vector space, linear independence and basis. CO3: Compute eigenvalues and eigenvectors problems. CO4: Demonstrate the use of truth tables and laws of identity, distributive, commutative, and domination. CO5: Simplify and prove Boolean expressions, Compute sum of products and product of sum expansions. CO6: Students can solve the problem based on theory by using maxima software.		
Sr.No.	List of Practical Assignments	Hours
1	Problem Solving on Unit 1: System of Linear Equation (Written)	4
2	Problem Solving on Unit 2: Vector Spaces (Written)	8
3	Problem Solving on Unit 3: Eigen values and Eigen vectors (Written)	8
4	Problem Solving on Unit 4: Boolean function (Written)	8
5	Problem Solving on Unit 1: System of Linear Equation (Using Maxima Software)	8
6	Problem Solving on Unit 2: Vector Spaces (Using Maxima Software)	8
7	Problem Solving on Unit 3: Eigen values and Eigen vectors (Using Maxima Software)	8
8	Problem Solving on Unit 4: Boolean function (Using Maxima Software)	8

Reference Books
1. Howard Anton, Chris Rorres, Elementary Linear Algebra, Application Version, Ninth Edition, Wiley, 11th edition.
2. K. Hoffman and R. Kunze, Linear Algebra, 2nd edition(2014), Prentice Hall of India, New Delhi.
3. Steven J. Leon, Linear Algebra with Applications, 4th edition(1994), Prentice Hall of India. New Delhi.
4. Discrete Mathematical Structures, by Kolman, Busby, Ross, Rehman, Prentice Hall

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I

SEC-101-DS: Computer Organization

No. of Credits: 2	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester :35 Marks	
Prerequisites Number systems and basics of digital electronics.			
Objectives - To revise about different number systems, codes, logic gates with truth tables. - To understand combinational and sequential circuits of digital electronics. - To conceptualize the basics of organizational and architectural issues of a digital computer and learn about various data transfer techniques in digital computer and the I/O interfaces. - To know how I/O devices are accessed and its principles and to provide the knowledge on Instruction Level Parallelism. - To study architecture			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand number systems related to computer and their inter-conversion. CO2: Familiar with digital circuits, their types, and applications. CO3: Understand CPU and Memory organizations for the fundamentals of computer. CO4: Study interfacing of peripherals with CPU in serial and parallel manner with data convertors. CO5: Study basics of microprocessor architecture and concept of pipelining			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Digital Circuits	12	CO1, CO2
Number Systems: Binary, Hexadecimal, BCD and their inter-conversion. Gray code and ASCII code. Logic Gates: Basic gates, derived gates, positive and negative logic, Simplification of logic circuits, De-Morgan's theorem. Concept of K map and simplification of single expressions (upto 4 variables). Combinational circuits: Half adder, full adder, half Subtractor, Multiplexer (2:1 and 4:1), Demultiplexer (1:2 and 1:4) using basic gates, Encoder - Decimal to BCD, Decoder - 3 to 8 decoder. Sequential circuits: Concept of triggering, Flip-Flops: SR, JK, D and T.Counters: Synchronous and Asynchronous (3-bit), Shift registers: types and applications.			
2	CPU, Memory and I/O Organizations	12	CO3, CO4
CPU Organization: Functions of CPU, General registers used in CPU: PC, SP, instruction pointer, instruction register, instruction decoder, flag, general purpose registers, memory address register, memory byte register, General register organization of CPU, Concept of stack. Memory organization: Memory hierarchy, cache memory and its address mapping, Associative memory, Virtual memory, Memory management through segmentation and paging. I/O Organization: Block diagram of parallel interface and function of blocks, Concept of			

interrupt, IVT, Types of I/O transfer, CPU initiated, interrupt initiated, DMA (only concept), Data converters: R-2RDAC, ADC (flash, successive approximation), Serial communication and types.			
3	Architecture of Microprocessor and Parallel Processing	6	CO5
Architecture of Microprocessor: Block diagram of 8086 and function of blocks, 8086 Registers, Numeric co-processor - concept, block diagram and functions of blocks. Parallel Processing: Concept of parallelism, Parallel computer structures. Concept of pipelining, Pipelined computers, Instruction pipeline, Arithmetic pipeline, Concept of RISC and CISC. RISC pipelining.			
Reference Books			
1. Modern Digital Electronics, 4th edition, R P Jain, Tata McGraw Hill publication. 2. Digital Logic & Computer Design, Morris Mano, Pearson. 3. Computer Systems Architecture – Morris Mano, 3rd Edition, Pearson 4. Computer Systems Organization & Architecture- John D. Carinelli Pearson publication.			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-101-DS-T : Office Automation I

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites • Previous knowledge of Computer concepts is assumed. • Knowledge of Computer as operational tool is required.			
Objectives • To introduce the foundations of office automation especially word processing. • To develop the ability to prepare the well formatted word documents. • To prepare the documents using word processing tools such as tables, figures, shapes etc. • To prepare the word documents using advanced automated features.			
Course Outcomes On Completion of this course, student will be able to - CO1: Prepare the professional word documents CO2: Explore various tools in the word processing software CO3: Develop documents using word processing advanced tools			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Working with Documents	2	CO1
1.1. Opening & Saving files, 1.2. Editing text documents, Inserting, Deleting, 1.3. Cut, Copy, Paste, Undo, Redo, 1.4. Find, Search, Replace, 1.5. Formatting page &setting Margins, 1.6. Converting files to different formats, 1.7. Importing & Exporting documents, Sending files to others, 1.8. Using Tool bars, Ruler, Using Icons, using help			
2	Formatting Documents	2	CO1
2.1 Setting Font styles 2.2 Font selection- style, size, colour, etc. 2.3 Type face - Bold, Italic, Underline, 2.4 Case settings, Highlighting, Special symbols. 2.5 Setting Paragraph style. 2.6 Alignments, Indents, Line Space, Margins, 2.7 , Bullets & Numbering			

3	Setting Page Style	4	CO1
3.1 Formatting Page 3.2 Page tab : Margins, Layout settings, Paper tray 3.3 Border & Shading 3.4 Columns 3.5 Header & Footer 3.6 Setting Footnotes & End notes 3.7 Shortcut Keys; Inserting manual page break, Column break and line break 3.8 Creating sections & frames 3.9 Anchoring & Wrapping 3.10 Printing Documents			
4	Setting Document Styles	2	CO1
4.1 Table of Contents 4.2 Index 4.3 Page Numbering 4.4 date & Time, Author, etc. 4.5 Creating Master Documents 4.6 Web page			
5	Creating Tables	7	CO2
5.1 Table settings and Drawing - Inserting ClipArts, Pictures/Files etc., 5.2 Borders, Alignments, 5.3 Insertion, deletion, 5.4 , Merging, Splitting, 5.5 Sorting, 5.6 Formula			
6	Special Features	6	CO2, CO3
6.1 Inserting Formula, equation, symbols 6.2 Inserting Cliparts, pictures, objects, word art 6.3 Drawing: shapes, smart art, etc 6.4 Charts 6.5 Hyperlinks, bookmarks, cross-references, Digital Signature			
7	Tools	7	CO2, CO3
7.1 Word Completion, Spell Checks, 7.2 Mail merge 7.3 Templates, 7.4 Creating contents for books, Creating Letter/Faxes, Creating Web pages 7.5 Hyperlinks, bookmarks, cross-references 7.5 Using Wizards 7.6 Tracking Changes, Security.			
Reference Books			
1. Illustrated Microsoft Office 365 & Word 2019 Comprehensive by Jennifer Duffy 2. Microsoft Word 365 2019 by Joan Lambert 3. Microsoft Word 2013 Bible by Lisa A Bucki			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)

Semester-I

Open Elective(OE)/General Elective(GE)

(To be offered to faculty other than Science & Technology)

OE-102-DS-T : Introduction to Computers and Basics of Internet

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites ● Basic knowledge of Computer concepts is assumed. ● Knowledge of Computer as operational tool is required.			
Objectives ● To introduce the fundamental concepts of computers ● To introduce the basic concepts of Internet ● To develop the ability to analyses and use the computer peripherals effectively ● To develop the ability to analyses and use the internet effectively			
Course Outcomes On Completion of this course, student will be able to – Use the computer peripherals effectively CO1: Use the internet for the day to day life CO2: Explore various applications available over the internet.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Fundamentals of Computers	8	CO1
1.1 Overview of a Computer-Definition, functionalities of Computer 1.2 Generations and Classification of Computers 1.3 Functional Components of a Computer 1.4 Applications Of Computers 1.5 Software and Hardware-Definition, types of software 1.6 Introduction to various Operating systems-Windows, Linux, Android, IOS			
2	Introduction to various Computer applications	6	CO1
2.1 Various Explorers 2.2 Editors such as Notepad, wordpad 2.3 Calculator, calendar, etc 2.4 Paint. 2.5 Various browsers 2.6 Internet settings			
3	Basics of Internet	6	CO2
3.1 Definition and History of Internet 3.2 Uses and Applications of Internet			

3.3 Definition of Web 3.4 Website Address and URL 3.5 Different types of Internet Connections: ● Dial up Connection ● Broad Band (ISDN, DSL, Cable) ● Wireless (Wi-Fi, WiMax, Satellite, Mobile) naming convention 3.6 Modes of Connecting Internet (Hotspot, Wi-Fi, USB Tethering)			
4	Browsers and Email	10	CO2
4.1 Search Engines 4.2 Web Browsers ● Popular Web Browsers (Microsoft Edge, Google Chrome, Mozilla Firefox,Safari, etc.) ● Popular Search Engines.(Google, Bing, Startpage ,DuckDuckGo etc..) 4.3 Portals 4.4 Social Networking sites, blogs 4.5 Using Browsers : ● Viewing webpage ● Downloading and uploading the website 4.6 E-mail: ● Configuring an E-mail Account ● Composing and Sending Mail ● Receiving, Replying to and Forwarding Mail ● Attachments to email			
Reference Books			
1. Computer Fundamentals by P.K. Sinha &Priti Sinha, 3rd edition, BPB pub. 2.Fundamental of Computers – By V. Rajaraman B.P.B. Publications 3. The Internet Book by Douglas E Comer			
E-Books and Online Learning Material			
1. https://www.geeksforgeeks.org/computer-fundamentals-tutorial/ 2. https://www.javatpoint.com/computer-fundamentals			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-103-DS-T : Introduction to Google Apps I

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites ● Basic knowledge of Computer concepts is assumed. ● Knowledge of Computer as operational tool is required. ● Knowledge of Internet is required			
Objectives To introduce the foundations of various Google tools. ● To develop the ability to analyses and use the tools effectively			
Course Outcomes On Completion of this course, student will be able to - CO1: Use the Google tools for the day to day life CO2: Explore various applications available in the google tools. CO3: Develop the skills to implement the skills available in the google tools.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Gmail	2	CO1
1.1 Configuring an E-mail Account 1.2 Composing and Sending Mail 1.3 Receiving, Replying to and Forwarding Mail 1.4 Attachments to email			
2	Google Drive	3	CO1, CO3
2.1 Opening the Drive 2.2 Creating folders, Google docs, Google sheets, Google slides 2.3 Managing Files and folders 2.4 Sharing files and folders and managing permissions 2.5 Downloading the files and folders 2.6 Uploading files and folders 2.7 Printing files			
3	Google Docs, Sheets and Slides	8	CO1, CO2,CO3
3.1 Creating Google docs, sheets and slides 3.2 Formatting the documents 3.3 Managing the document permissions 3.4 Uploading/downloading the documents			

3.5 Special features in the docs, sheets and slides			
4	Google Forms	7	CO1, CO2.CO3
4.1 Creating a Google form 4.2 Adding various styles of the questions 4.3 settings of the Google form 4.4 Creating the links of the Google form and sharing the link 4.5 Creating and managing the permissions 4.6 Managing the data collected through Google form			
5	Other Google tools	10	CO2,CO3
5.1 Google Calendar 5.2 Google Meet 5.3 Google Chat 5.4 Google Contacts 5.5 Google Photos 5.6 Google Maps			
Reference Books			
1. Complete Beginners guide to Google Apps Script by Daniel Lawrie. 2.Google Apps made easy by James Bernstein 3. My Google Apps by Sherry Kinkoph Gunter			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-I
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-104-DS-T : Fundamentals of Computers I

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basic knowledge of Computer concepts is assumed.			
Objectives To converse with basic terminology of computer - To understand basics of Computer and working with Operating System - To develop working skills with productivity enhancing tools - To perform documentation and accounting operations			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the concept of input and output devices of Computers CO2: Learn the functional units and classify types of computers CO3: Understand concept of software and working of operating system CO4: Learn basic Word processing, Spreadsheet and Presentation Graphics Software skills CO5: Study to use the Information Technology safely, legally, and responsibly CO6: Describe various uses of offices automation tools in accounting Operations			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Computers	15	CO1,CO2,CO3
1.1 Basics of Computers - Definition, Block Diagram, Computer Hierarchy, (Classification), Characteristics of Computer, Computer Memory Input and Output Devices. 1.2 Introduction to Software - Software Types - System Software, Application Software, Types of Operating Systems, Functions of Operating Systems. 1.3 Working with Windows Operating System:- Structure of Windows, Windows Explorer, File and Folder Operations, The Search, The Recycle Bin, Adding or Removing New Programs using, Control Panel, Applications in windows (Paint, Notepad, WordPad, and Calculator) Data Processing: Files and Records, File Organization (Sequential, Direct/Random, Index)			
2	Office Automation Tools	15	CO4,CO5,CO6
2.1 Definition of Information Technology (IT) Benefits of Information Technology (IT) Applications of Information Technology (IT) 2.2 Office Automation Tools: 2.2.1 MS-Word: Introduction, Starting MS-Word, MS-Word Screen and its Components, Elementary Working with MS-Word			

2.2.2 MS-Excel: Introduction, Starting MS-Excel, Basics of Spread sheet, MS-Excel Screen and its Components, Elementary Working with MS-Excel MS-PowerPoint: Introduction, Starting MS-PowerPoint, Basics of PowerPoint, MS-PowerPoint Screen and Its Components, Elementary Working with MS PowerPoint.
Reference Books
1. Computer Fundamentals by: Anita Goel, Pearson Education India ISBN: 9788131742136 2. Connecting with Computer Science, by Greg Anderson, David Ferro, Robert Hilton, Course Technology, Cengage Learning, ISBN:9781439080351 3. Fundamentals of Computer : For undergraduate courses in commerce and management, ITL Education Solutions Limited, Pearson Education, ISBN:9788131733349 4. Introduction to Computer Science, 2/e, ITL Education Solutions Limited, Pearson Education, ISBN:9788131760307 5. Frontiers of Electronic Commerce, Ravi Kalakota, Andrew B. Whinston, Pearson Education, ISBN:9788177583922 6. Internet: The Complete Reference, Margaret Levine Young, Tata McGraw Hill Education Private Limited, ISBN: 9780070486997 7. On the Way to the Web: The Secret History of the Internet and Its Founders, A. Banks, Apress Publication, ISBN: 9781430208693 8. Computers and Commerce: A Study of Technology and Management at Eckert-Mauchly Computer Company, Engineering Research Associates, and Remington, Arthur L. Norberg, MIT Press (MA), ISBN:9780262140904 9. Essential of E-commerce technology by V. Rajaraman, Prentice Hall India Learning Private Limited ISBN 9788120339378 10. Fundamentals of Computers by E. Balagurusamy, McGraw Hill Computer Fundamentals by Priti Sinha, Pradeep K. Sinha, BPB Publications
Continuous Internal Evaluation – Max. Marks 15 Marks (Min. Passing Marks: 06)(Min. Passing Percentage: 40% of Max. Marks)
The colleges need to adopt any Two Methods out of the following Methods for Continuous Internal Evaluation: 1) Offline Written Examination 2) Power Point Presentations 3) Assignments / Tutorials 4) Oral Examination 5) Open Book Test 6) Offline MCQ Test 7) Group Discussion 8) Analysis of Case Studies
Semester End Examination: Max. Marks 35 and Duration of Examination is 2 Hours (Min. Passing Marks: 14) (Min. Passing Percentage: 40% of Max. Marks)
Instructions: 1. Attempt all questions Q. 1. Fill in the Blanks on all Units (05 Marks) Q. 2. Theory Question on Unit-1 OR Unit-2 (08 Marks) Q. 3. Numerical Problem on Unit-1 OR Unit-2 (14 Marks) Q. 4. Write Short Notes on all Units (Any 2 out of 4) (08 Marks)

Detail Syllabus

B.Sc. (Data Science)

Semester-II

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

DS-151-T : Advanced Python Programming

No. of Credits: 2		Teaching Scheme Theory:2 Hrs /Week	Examination Scheme Continuous Evaluation:15 Marks End Semester:35 Marks	
Prerequisites - Fundamentals of Python Programming Language. - Prior knowledge of computational mathematics.				
Objectives - To learn reading, writing and manipulating files - To implement libraries like Pandas, NumPy,SciPy, Matplotlib, Scikit-learn etc. in Python. - To implement the concepts of GUI controls and designing GUI applications. - To learn and know the concepts of file handling, exception handling.				
Course Outcomes On Completion of this course, student will be able to - CO1: Learn reading and writing into files using Python CO2: Design and implement a program to solve a computational problem. CO3: Understand implementation of libraries like Pandas, NumPy, SciPy, Matplotlib, Scikit-learn etc. in Python. CO4: How to handle exceptions and files. CO5: Design and implement GUI application				
Unit No.	Name of Unit		Teaching Hours	CO Targeted
1	File Handling		3	CO1
1.1 Introduction to Files 1.2 Types of Files 1.3 Opening and Closing a Text File 1.4 Writing to a Text File 1.5 Reading from a Text File 1.6 Setting Offsets in a File 1.7 Creating and Traversing a Text File				
2	Python Libraries		17	CO2 & CO3
2.1 Introduction to Python Libraries 2.1.1 Statistical Analysis- NumPy, SciPy, Pandas, StatsModels 2.1.2 Data Visualization- Matplotlib, Seaborn, Plotly 2.1.3 Data Modelling and Machine Learning- Scikit-learn, XGBoost, Eli5				

2.1.4 Deep Learning- TensorFlow, Pytorch, Keras 2.1.5 Natural Language Processing (NLP)- NLTK, SpaCy, Gensim 2.2 Working with Tabular Numeric Data(Numpy with Python) 2.2.1 NumPy Arrays Creation Using array() Function 2.2.2 Array Attributes, NumPy Arrays Creation with Initial Placeholder Content 2.2.3 Integer Indexing, Array Indexing, Boolean ArrayIndexing, Slicing and Iterating in Arrays Basic Arithmetic Operations on NumPy Arrays 2.2.4 Mathematical Functions in NumPy 2.2.5 Changing the Shape of an Array, Stacking and Splitting of Arrays, Broadcasting in Arrays. 2.3 Working with Data Series and Frames 2.3.1 Pandas Data Structures, Reshaping Data, Handling Missing Data 2.3.2 Combining Data, Ordering and Describing Data, Transforming Data, Taming Pandas File I/O 2.4 Plotting Basic Plotting with PyPlot, Matplotlib, Getting to Know Other Plot Types, Plotting with Pandas			
3	Exception Handling	5	CO4
3.1 Python Exception 3.2 Common Exception 3.3 Exception handling in Python (try-except-else) 3.4 The except statement with no exception 3.5 Multiple Exception 3.6 The try-finally clause 3.7 Custom Exception and assert statement			
4	GUI Programming	5	CO5
4.1 Introduction 4.2 Tkinter programming 4.3 Tkinter widgets 4.5 Frame 4.6 Button 4.7 Label 4.8 Entry			
Reference Books			
1.Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010 2.Dive into Python, Mike 3. Learning Python, 4th Edition by Mark Lutz 4. Programming Python, 4th Edition by Mark Lutz 5.Python Programming:An introduction to computer,John Zelle,3rd Edition. 6. Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016 7. Introduction to Python Programming. Gowrishankar S., Veena A. CRC Press, Taylor & Francis Group, 2019			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

DS-152-P : Lab Course on DS-151-T (Advanced Python Programming)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks	
Prerequisites • Practical Knowledge of Python Programming. • Prior knowledge of computational mathematics.			
Objectives • To learn reading, writing and manipulating files • To implement libraries like Pandas, NumPy, SciPy, Matplotlib, Scikit-learn etc. in Python. • To implement the concepts of GUI controls and designing GUI applications. • To learn and know the concepts of file handling, exception handling.			
Course Outcomes On Completion of this course, student will be able to - CO1: Learn reading and writing into files using Python CO2: Design and implement a program to solve a computational problem. CO3: Understand implementation of libraries like Pandas, NumPy, SciPy, Matplotlib, Scikit-learn etc. in Python. CO4: How to handle exceptions and files. CO5: Design and implement GUI application			
Unit No.	Name of Unit	Hours	CO Targeted
1	File Handling	8	CO1
a. Assignments on reading and writing files b. Assignments on accessing and manipulating files			
2	Python Numpy	12	CO2
a. Assignments on Numpy basics. b. Assignments on Numpy Arrays. c. Assignments on Numpy Linear Algebra d. Assignments on Numpy Statistics e. Assignments on Numpy Strings			
3	Python Pandas	8	CO3
a. Assignments on reading and writing text, CSV files using Pandas			
4	Plotting in Python	12	CO3
a. Assignments on plotting graphs in python using Matplotlib and Pyplot			
5	Assignments on Exception Handling	8	CO4
6	Assignments on GUI	12	CO5
Reference Books 1.Mark Lutz, Programming Python, O`Reilly, 4th Edition, 2010			

- 2.Dive into Python, Mike
3. Learning Python, 4th Edition by Mark Lutz
4. Programming Python, 4th Edition by Mark Lutz
- 5.Python Programming:An introduction to computer,John Zelle,3rd Edition.
6. Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016
7. Introduction to Python Programming. Gowrishankar S., Veena A. CRC Press, Taylor & Francis Group, 2019

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

DS-153-T : Discrete Probability and Probability Distributions

No. of Credits: 02	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basics of Set theory, Mathematical operations			
Objectives - To revise the basic concepts of probability, axiomatic theory of probability. - To understand the concept of random variable - To study probability distribution (univariate and bivariate) discrete random variables, expectation and moments of probability distribution - To find marginal distribution and conditional distribution of bivariate frequency distribution - To find conditional mean of bivariate frequency distribution - To find variance, covariance and correlation of bivariate frequency distribution			
Course Outcomes On Completion of this course, student will be able to – CO1: Find the probabilities of events and its expectation, mean, variance, etc. CO2:Distinguish between random and non-random experiments CO3:Identify the nature of distribution CO4: Find marginal distribution and conditional distribution CO5: Find mean of marginal distribution and conditional mean of bivariate frequency distribution CO6: Find correlation of bivariate frequency distribution			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Probability	07	CO1, CO2
Basics of Probability: Experiments/Models, Ideas of deterministic and non-deterministic models. Random Experiment.			
Definitions: Sample space, Discrete sample space: finite and countably infinite, Event, Elementary event, Complement of an event, Certain event, Impossible event.			
Occurrence of events: Concept of occurrence of an event. Algebra of events and its representation in set theory notation. Occurrence of following events (i) at least one of the given events, (ii) none of the given events, (iii) all of the given events, (iv) mutually exclusive events, (v) mutually exhaustive events, (vi) exactly one event out of the given events.			
Classical definition of probability and its limitations. Probability model, probability of an event, equiprobable and non-equiprobable sample space.			
Axiomatic definition of probability. Theorems and results on probability with proofs based on axiomatic definition such as $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Generalization $P(A \cup B \cup C), 0 \leq P(A) \leq 1, P(A) + P(A') = 1, P(\Phi) = 0$ and when $A \subseteq B$ then $P(A) \leq$			

$P(B)$.			
2	Conditional Probability	05	CO2
Definition of conditional probability of an event. Results on conditional probability. Definition of independence of two events $P(A \cap B) = P(A)P(B)$. Pairwise independence and mutual independence for three events. Multiplication theorem $P(A \cap B) = P(A)P(B A)$. Generalization to $P(A \cap B \cap C)$. Partition of the samplespace, prior and posterior probabilities. Proof of Bayes' theorem. Applications of Bayes' theorem in real life.			
3	Univariate Probability Distributions and its Mathematical Expectation	9	CO3
Univariate Probability Distributions defined on Discrete Sample Space: Concept and definition of a discrete random variable. Probability Mass Function (<i>pmf</i>) and cumulative Distribution Function (<i>cdf</i>), $F(\cdot)$ of discrete random variable, properties of <i>cdf</i>. Mode and median of a univariate discrete probability distribution. Mathematical Expectation: Definition of expectation (mean) of a random variable, expectation of a function of a random variable, Moment Generating Function (<i>mgf</i>) and Cumulative Generating Function (<i>cgf</i>). Properties of <i>mgf</i> and <i>cgf</i>. Definitions of variance, standard deviation (SD) and Coefficient of variation (CV) of univariate probability distribution, effect of change of origin and scale on mean, variance and SD. Definition of raw, central and factorial raw moments of univariate probability Distributions and their interrelations (without proof). Coefficients of skewness and kurtosis based on moments.			
4	Mathematical Expectation for Bivariate Frequency Distribution	9	CO4, CO5, CO6
Definition of raw and central moments, <i>mgf</i> and <i>cgf</i> . Theorems on expectations of sum and product of two jointly distributed random variables. Conditional expectation. Definitions of conditional mean and conditional variance. Definition of covariance, coefficient of correlation, independence and uncorrelatedness of two variables. Variance of linear combination of variables, $Var(aX + bY)$, $Var(aX + bY + C)$ and its generalization.			
Reference Books			
1. Gupta, S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi. 2. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentice Hall of India, New Delhi. 3. Hoel, P. G. (1971). Introduction to Mathematical Statistics, John Wiley and Sons, New York. 4. Hogg, R. V. and Craig, R. G. (1989). Introduction to Mathematical Statistics, Ed. MacMillan Publishing Co., New York. 5. Mayer, P. (1972). Introductory Probability and Statistical Applications, Addison Wesley Publishing Co., London. 6. Mood, A. M. and Graybill, F. A. and Boes D. C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company. 7. Rao, VLS Prakash (2008). First Course in Probability and Statistics, New Age International Publishers, New Delhi. 8. Ross S. (2002). A First Course in Probability, Sixth Edition, Pearson Education, Inc. & Dorling Kindersley Publishing, Inc.			

SavitribaiPhule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

DS-154-P : Lab Course on DS-153-T (Discrete Probability and Probability Distributions)

No. of Credits: 02	Teaching Scheme Practical: 4 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks
Prerequisites Basics of Set theory, Mathematical operations		
Objectives - To understand the concept of random variable - To study probability distribution (univariate and bivariate) discrete random variables, expectation and moments of probability distribution - To find marginal distribution and conditional distribution of bivariate frequency distribution - To find conditional mean of bivariate frequency distribution - To find variance, covariance and correlation of bivariate frequency distribution		
Course Outcomes On Completion of this course, student will be able to – CO1: Find the probabilities of events and its expectation, mean, variance, etc. CO2:Distinguish between random and non-random experiments CO3:Identify the nature of distribution CO4: Find marginal distribution and conditional distribution CO5: Find mean of marginal distribution and conditional mean of bivariate frequency distribution CO6: Find correlation of bivariate frequency distribution		
Sr.No.	List of Practical Assignments	Hours
1	Calculation of probability for different events based on real life situations.	4
2	Calculation of mathematical expectation and variance.	4
3	Obtain marginal and conditional distribution of bivariate probability distribution	4
4	Calculation of conditional expectation and conditional variance	4
5	Calculation of variance of linear combination.	4
6	Checking the independence of the probabilities of the events	4
7	Calculation of the correlation coefficient based on bivariate probability distribution	8
8	Model sampling from the given probability distributions	12
9	Small Project equivalent to 4 practical.	16
Reference Books		

1. Gupta, S. C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi.
2. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentce Hall of India, New Delhi.
3. Hoel, P. G. (1971). Introduction to Mathematical Statistics, John Wiley and Sons, New York.
4. Hogg, R. V. and Craig, R. G. (1989). Introduction to Mathematical Statistics, Ed. MacMillan Publishing Co., New York.
5. Mayer, P. (1972). Introductory Probability and Statistical Applications, Addison Wesley Publishing Co., London.
6. Mood, A. M. and Graybill, F. A. and Boes D. C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.
7. Rao, VLS Prakash (2008). First Course in Probability and Statistics, New Age International Publishers, New Delhi.
8. Ross S. (2002). A First Course in Probability, Sixth Edition, Pearson Education, Inc. & Dorling Kindersley Publishing, Inc.

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

DS-155-T : Graph Theory

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites Basics of mathematics, Set Theory			
Objectives - To introduce students about graph, graph models, types of graph, connectivity, applications of graph theory. - To know how to find shortest path for different Eulerian and Hamiltonian circuit. - To introduce students about Trees, applications of trees, binary tree, tree traversal , spanning trees. - To know how to find minimum spanning trees. - To make students familiar with the use of all these concepts as tools in other areas of the course curriculum.			
Course Outcomes On Completion of this course, student will be able to : CO1: Understand the graph, and graph models, terminology of graph. CO2: Students can solve examples on adjacency and incidence matrix. CO3: Identify the Euler tours and Hamiltonian cycle and find shortest path. CO4: Able to Compute the shortest spanning trees. CO5: Students can solve the problems on tournaments and traffic flow.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Graphs and Graph Models	4	CO1, CO5
Graph: Definition, basic terminology of Graph, Graph Models, Social networks, Communication networks, Information networks ,Software Design Applications, Transportation networks, Biological networks, Tournaments.			
2	Graph Isomorphism	5	CO2
Handshaking lemma, Special Types of Graph, Directed graph, Matrix representation of graph, Definition of isomorphism,Examples on isomorphism of graphs.			
3	Connected Graph	8	CO3
Walk, trail, path, cycle, connected graph, disconnected graph, component, Cut edge, Cut vertex, Cut set, Vertex connectivity, edge connectivity, Minimal degree of a graph, Relation between Vertex connectivity, edge connectivity and Minimal degree of a graph,Weighted graph, Shortest path algorithm, Dijkstra's algorithm			
4	Eulerian and Hamiltonian Graphs .	5	CO3
The Konigsberg Seven Bridge problem, Euler’s path, Euler’s circuit, Eulerian graph, Fleury’s algorithm, Hamilton path, Hamilton Circuit, Hamiltonian graph, Applications of Eulerian and Hamiltonian graph: Traveling Salesman problem, Chinese Postman problem.			

5	Trees	8	CO4
Definition of tree, basic terminology of tree, properties of trees, Eccentricity of a vertex, Centre, diameter, radius of a tree, Spanning Tree, Chords and branches of Spanning Tree, Shortest spanning tree, Kruskal's algorithm, M-ary tree, binary tree, Tree traversal, Ordered rooted tree, polish notation, arborescence.			
Reference Books			
1. Kenneth Rosen, Discrete Mathematics and It's Applications, Tata McGraw Hill, Seventh Edition. 2. Narsingh Deo, Graph Theory with applications to computer science and engineering, Prentice Hall. 3. Dougals B. West, Introduction to Graph Theory, Pearson Education, Second edition.			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
DS-156-P : Lab Course on DS-155-T (Graph Theory)

No. of Credits: 2	Teaching Scheme Practical: 4 Hours/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Basics of mathematics, Set Theory		
Objectives - To introduce students about graph, graph models, types of graph, connectivity, applications of graph theory. - To know how to find shortest path for different Eulerian and Hamiltonian circuit. - To introduce students about Trees, applications of trees, binary tree, tree traversal , spanning trees. - To know how to find minimum spanning trees. - To make students familiar with the use of all these concepts as tools in other areas of the course curriculum. - To know how to use Maxima software.		
Course Outcomes On Completion of this course, student will be able to : CO1: Understand the graph, and graph models, terminology of graph. CO2: Students can solve examples on adjacency and incidence matrix. CO3: Identify the Euler tours and Hamiltonian cycle and find shortest path. CO4: Able to Compute the shortest spanning trees. CO5: Students can solve the problems on tournaments and traffic flow. CO6: Students can solve the problems on theory using Maxima Software.		
Sr.No.	List of Practical Assignments	Hours
1	Problem Solving on Unit 1: Graphs and Graph Models (Written)	4
2	Problem Solving on Unit 2: Graph Isomorphism (Written)	4
3	Problem Solving on Unit 3: Connected Graph (Written)	4
4	Problem Solving on Unit 4: Eulerian and Hamiltonian Graphs (Written)	4
5	Problem Solving on Unit 5: Trees (Written)	4
6	Problem Solving on Unit 1: Graphs and Graph Models (Using Maxima Software)	8
7	Problem Solving on Unit 2: Graph Isomorphism (Using Maxima Software)	8
8	Problem Solving on Unit 3: Connected Graph (Using Maxima Software)	8
9	Problem Solving on Unit 4: Eulerian and Hamiltonian Graphs (Using Maxima Software)	8
10	Problem Solving on Unit 5: Trees (Using Maxima Software)	8
Reference Books		

1. Kenneth Rosen, Discrete Mathematics and It's Applications, Tata McGraw Hill, Seventh Edition.
2. Narsingh Deo, Graph Theory with applications to computer science and engineering, Prentice Hall.
3. Dougals B. West, Introduction to Graph Theory, Pearson Education, Second edition.

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II

SEC-151-DS : Lab Course on Excel and Advanced Excel

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basic Computer Skills and Mathematics Skill.			
Objectives - To familiarize the student in introducing and exploring MS excel. - To provide different ways of representation and exploratory data analysis in excel. - To prepare the students to use excel in their project works - Analyze data like a professional.			
Course Outcomes On Completion of this course, student will be able to - CO1: To Implement fundamental concept of Microsoft Excel CO2: Perform calculations in excel and apply excel functions. CO3: Represent data using charts and diagrams CO4: Design advanced graphic presentations on stored data. CO5: Perform various advanced data tools and data analytics.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Microsoft Excel	7	CO1
- Concepts of Work book &Work sheets - Various Data Types - Using different features with data, Cell and Texts - Inserting, Removing & Resizing of Columns & Rows - Working with Data and Ranges - Entering data into worksheet - Saving & quitting worksheet - Opening and moving around in an existing worksheet - Toolbars and menu, keyboard shortcuts - Working with single and multiple workbook- copying, renaming, moving, adding and deleting, copy in gentries and moving between work books - Different Views of Work sheets - Column Freezing, Labels, Hiding, Splitting etc. - Using different features with Data and Text; Advanced paste special techniques			
2	Formulas ,Functions and charts in Excel	7	CO1,CO2,CO3
Use of Formulas			

• Calculations and Functions • Chart Tools • Different types of charts and their use • Logical Functions • Text Functions • Date and Time Functions • Lookups.			
3	Advance Data Tools	4	CO5
• What-if-Analysis- Goal Seek, Data Table • Scenario Manager • Formatting Charts, 3D Graphs			
4	Advanced Graphing and Charting	5	CO3,CO4
• Formatting and customizing Pivot tables • Using advanced options of Pivot tables, Pivot charts • Line, Bar and Pie charts • Scatter plots • Histograms.			
5	Analytics using Excel	7	CO5
• Data analysis using normal chart • Regression in Excel • Correlation, stddev, average, ANOVA			
Reference Books			
1. Mastering MS Excel: Functions and Formulas, Webtech (Khanna Publications) 2. Microsoft Excel 2019 Data Analysis and Business Modeling, Wayne Winston, 2019 3. Advance Excel 2016, training Guide, By Ritu Arora			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)

Semester-II

Open Elective(OE)/General Elective(GE)

To be offered to faculty other than Science & Technology

OE-151-DS-T : Office Automation II

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites ● Previous knowledge of Computer concepts is assumed. ● Knowledge of Computer as operational tool is required.			
Objectives ● To introduce the foundations of office automation especially Presentation Skills. ● To develop the ability to prepare the well formatted PowerPoint presentations. ● To prepare the presentations using PowerPoint presentation tools such as tables, figures, shapes, images, audio, video etc. ● To prepare the presentations using advanced automated features such as animation, slide shows, etc.			
Course Outcomes On Completion of this course, student will be able to - CO1: Prepare the professional presentations CO2: Explore various tools in the PowerPoint presentation software CO3: Develop documents using PowerPoint advanced tools			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Presentation	5	CO1
1.1. Opening new presentation, 1.2. Different presentation templates, 1.3. Setting backgrounds, 1.4. Selecting presentation layouts			
2	Creating and Formatting Presentation	8	CO1
2.1 Presentation style, 2.2 Adding text to the Presentation. 2.3 Adding style 2.4 Colour, gradient fills 2.5 Arranging objects 2.6 Adding Header & Footer 2.7 Slide Background, Slide layout			
3	Adding Graphics and effects to Presentation	12	CO1,CO2
3.1 Inserting pictures, movies, tables etc into presentation,			

3.2. Drawing Pictures using Draw 3.3. Setting Animation 3.4 Transition Effect			
4	Printing and showing a presentation	5	CO2, CO3
4.1 Printing Handouts 4.2 Generating Standalone Presentation viewer 4.3 Presenting the presentation using various styles			
Reference Books			
1. Microsoft PowerPoint by James Holler 2. PowerPoint for dummies, office 2021 Edition Doug Lowe 3. Learn Microsoft Office 2019 by Linda Foulkes			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-152-DS-T : Computer Fundamentals

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites ● Basic knowledge of Computer concepts is assumed. ● Knowledge of Computer as operational tool is required.			
Objectives To introduce the fundamental concepts of computers. ● To study the basics of Computer System ● To introduce the computer peripherals and other devices ● To learn how to configure computer devices ● To Learn Basic Commands of Operating system and application software			
Course Outcomes On Completion of this course, student will be able to - CO1: Use the computers for the day to day life CO2: Learn the fundamental concepts of computer science CO3: Explore various applications available in the computers CO4: Explain the needs of hardware and software required for a computation task			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Computers	7	CO1,CO2
1.1 Introduction, Characteristics of Computers, 1.2 Block diagram of computer 1.3 Computer Generations 1.4 Types of computers and features- Mini Computers, Micro Computers, Mainframe Computers, Super Computers, Laptops and Tablets 1.5 Types of Programming Languages- Machine Languages, Assembly Languages, High Level Languages			
2	Introduction to Computer Peripherals	8	CO2,CO4
2.1 Primary And Secondary storage devices 2.2 Primary storage devices – RAM, ROM, PROM, EPROM 2.3 Secondary Storage Devices - CD, HD, Pen drive 2.4 I/O Devices- Scanners, Digitizers, Plotters, LCD, Plasma Display 2.5 Pointing Devices –Mouse, Joystick, Touch Screen			
3	Number System	6	CO2,CO3

3.1 Introduction to Binary Number System 3.2 Introduction to Octal Number System 3.3 Introduction to Hexadecimal Number System 3.4 Addition, Subtraction, Multiplication, Division			
4	Operating Systems	4	CO2,CO3
4.1 Definition of Operating System 4.2 Functions of Operating System 4.3 Role of Operating System 4.4 Types of Operating System			
5	Introduction to Computer Networking	5	CO3,CO4
5.1 Network definition Common terminologies: LAN, WAN, Node, Host, Workstation, bandwidth, Interoperability 5.2 Network Administrator, 5.3 Network Security 5.4 Network Components: Servers, Clients 1.1 Communication Media 5.6 Types of network: Peer to Peer, Clients Server			
Reference Books			
1. Computer Fundamentals by P.K. Sinha & Priti Sinha, 3rd edition, BPB pub. 2. Fundamental of Computers – By V. Rajaraman B.P.B. Publications 3. Computer Networks – By Tannenbaum Tata MacGrow Hill Publication			
E-Books and Online Learning Material			
3. https://www.geeksforgeeks.org/computer-fundamentals-tutorial/ 4. https://www.javatpoint.com/computer-fundamentals			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-153-DS-T : Introduction to Google Apps II

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites ● Basic knowledge of Computer concepts is assumed. ● Knowledge of Computer as operational tool is required. ● Knowledge of Internet is required			
Objectives ● To introduce the specialized Google tools. ● To develop the ability to analyses and use the tools effectively and skilfully			
Course Outcomes On Completion of this course, student will be able to - CO1: Use the google tools for the day to day life CO2: Explore various applications available in the Google tools. CO3: Develop the skills to implement the skills available in the Google tools			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Creating You Tube Channel	10	CO1,CO2,CO3
1.1 Creating a you tube channel 1.2 Managing the channel, permissions, playlists, etc 1.3 Uploading the videos on the channel 1.4 Live Streaming			
2	Creating a website	10	CO1,CO2,CO3
2.1 Creating a website 2.2 Managing home page 2.3 Creating Menus on the website 2.4 Adding pages to the site 2.5 Setting up themes 2.6 Adding tools such as tables, placeholders, hyper linking, buttons, maps, etc 2.7 Publishing a site			
3	Google Classroom	10	CO1,CO2,CO3
3.1 Creating Google classroom 3.2 Creating a class 3.3 Streaming a class 3.4 Adding students to class			

3.5 Adding class work : assignments, quiz, question, material, etc
3.5 Giving grades to the assignments
3.6 joining a class (from student side)
Reference Books
1. Complete Beginners guide to Google Apps Script by Daniel Lawrie.
2. Google Apps made easy by James Bernstein
3. My Google Apps by Sherry Kinkoph Gunter

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-154-DS-T : Fundamentals of Computers II

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basic knowledge of Computer concepts is assumed			
Objectives To build an understanding of the fundamental concepts of computer networking. - To familiarize the student with the basic terminology of the computer networking area. - To understand computer network technology with various devices independently. - To identify the different types of network topologies and protocols.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the basic concepts of Networking and Cyber Security CO2: Describe Cyber Security Laws and concepts of Digital Signature CO3: Identify the different types of Network devices and their functions within a Network CO4: Elaborate the Internet Services and related terms of Internet CO5: Evaluate information security threats			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Computers Network and E-Commerce	15	CO1,CO3,CO4
1.1 Introduction Importance of Networking Computer Network (LAN, WAN, MAN) Network Components (Hub, Switch, Bridge, Gateway, Router, Modem). 1.2 Network Topology, Wireless Network Internet and Internet application Introduction, Internet evolution, Working of Internet, Use of Internet, Overview of World Wide Web (Web Server and Client). 1.3 IEEE802.11 -Wi-Fi: Types of Wi-Fi, Uses of Wi-Fi Near by Share : Applications of Nearby share			
2	Introduction to Internet & Cyber Security	15	CO1,CO2,CO4, CO5
2.1 Concept of Internet, Internet Service Providers(ISP), Services Provided by the Internet: E-mail, Search Engine 2.2 Information security overview – Background and current scenario Types of Attacks, Goals of security 2.3 Overview of security threats, Weak / Strong passwords and password cracking, Insecure			

Network connections, Digital Signature. 2.4 Cyber Security: Cybersecurity definition, Cybercrime: Classification of Cybercrime: a)Email spoofing b) Spamming c) Identity Theft d) Online Fraud e) ATM Skimming f) Credit Card Fraud 2.5 Overview of Indian Information Technology Act 2002
Reference Books
1. Computer Fundamentals by: Anita Goel, Pearson Education India ISBN: 9788131742136 2. Connecting with Computer Science, by Greg Anderson, David Ferro, Robert Hilton, Course Technology, Cengage Learning,ISBN:9781439080351 3. Fundamentals of Computer : For undergraduate courses in commerce and management, ITL Education Solutions Limited, Pearson Education, ISBN:9788131733349 4. Introduction to Computer Science, 2/e, ITL Education Solutions Limited, Pearson Education, ISBN:9788131760307 5. Frontiers of Electronic Commerce, Ravi Kalakota, Andrew B. Whinston, Pearson Education,ISBN:9788177583922 6. Internet: The Complete Reference, Margaret Levine Young, Tata McGraw Hill Education Private Limited, ISBN: 9780070486997 7. On the Way to the Web: The Secret History of the Internet and Its Founders, A. Banks, Apress Publication, ISBN: 9781430208693 8. Computers and Commerce: A Study of Technology and Management at Eckert-Mauchly Computer Company, Engineering Research Associates, and Remington, Arthur L. Norberg, MIT Press (MA),ISBN:9780262140904 9. Essential of E-commerce technology by V.Rajaraman, Prentice Hall India Learning Private Limited ISBN 9788120339378 10. Fundamentals of Computers by E. Balagurusamy, McGraw Hill 11. Computer Fundamentals by Priti Sinha, Pradeep K. Sinha, BPB Publications 12. Computer Networks - Andrew Tanenbaum (III Edition) 13. Complete Guide to Networking - Peter Norton 14. Data Communications & Networking - Behrouz Ferouzan (III Edition) 15. Cyber Security Understanding Cyber Crimes, Computer Forensics and Legal Perspectives – Nina Godbole, Sunit Belapure, Wiley: April 2011 India Publications Released. 16. Cyber Security Essentials- James Graham Richard Howard Ryan Olson Indian Information Technology Act 2002
Continuous Internal Evaluation – Max. Marks 15 Marks (Min. Passing Marks: 06)(Min. Passing Percentage: 40% of Max. Marks)
The colleges need to adopt any Two Methods out of the following Methods for Continuous Internal Evaluation: 1) Offline Written Examination 2) PowerPoint Presentations 3) Assignments / Tutorials 4) Oral Examination 5) Open Book Test 6) Offline MCQ Test 7) Group Discussion 8) Analysis of Case Studies

Semester End Examination: Max. Marks 35 and Duration of Examination is 2 Hours (Min. Passing Marks: 14)(Min. Passing Percentage: 40% of Max. Marks)
Instructions: 1. <i>Attempt all questions</i> Q. 1. Fill in the Blanks on all Units (5 Marks) Q. 2. Theory Question on Unit-1 OR Unit-2 (8 Marks) Q. 3. Numerical Problem on Unit-1 OR Unit-2 (14 Marks) Q. 4. Write Short Notes on all Units (Any 2 out of 4) (8 Marks)

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-155-DS-T : Introduction to Data Science

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basic knowledge of Computer concepts is assumed			
Objectives To build an understanding of the fundamental concepts of computer networking. - To understand need of Data Science - To Know role of Statistics in Data Science - To know Data Science Models and Tasks			
Course Outcomes On Completion of this course, student will be able to - CO1: Define Data Science Tasks and Models and Lifecycle CO2: Apply Pre-processing and visualization Techniques\			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction	6	CO1
What and why. Why learn Data Science?, Types of Data -structured, semi-structured, unstructured Data Applications of Data Science, The Data Science Lifecycle, Role of Data Scientists Data sources-Open Data, Social Media Data, Multimodal Data, standard datasets			
2	Statistics for Data Science	6	CO1
Data Objects and Attributes, Attribute Types: Nominal, Binary, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes, Role of statistics in Data Science Descriptive statistics - Measuring the Frequency, Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion: Range, Standard deviation, Variance, Inter quartile Range			
3	Data science Models and Tasks	6	CO1
Predictive and Descriptive Models, Introduction to Data Science Tasks – Classification, Prediction, Association, Clustering, Performing simple Data Science Tasks using WEKA / R			
4	Data Quality and Pre-processing	6	CO2
Data Quality: Why Pre-process the Data?, Data munging/wrangling operations			

Data Cleaning - Missing Values, Noisy Data Data Transformation – Rescaling, Normalizing, Data reduction and Data discretization			
5	Data Visualization	6	CO2
Introduction to Exploratory Data Analysis (EDA), Data visualization, Basic data visualization tools –Box Plots, Histograms, Bar charts/graphs, Scatter plots, Line charts, Area plots, Pie charts			
Reference Books			
1. Data Science Fundamentals and Practical Approaches, Gypsy Nandi, Rupam Sharma, BPB Publications, 2020. 2. Data Mining Concepts and Techniques, Third Edition, Jiawei Han, Micheline Kamber, Jian Pei, Morgan Kaufmann, 2012. 3. A Hands-On Introduction to Data Science, Chirag Shah, University of Washington Cambridge University Press			

Savitribai Phule Pune University
B.Sc. Data Science (Pattern 2024)
Semester-II
Open Elective(OE)/General Elective(GE)
(To be offered to faculty other than Science & Technology)
OE-156-DS-T : AI Tools for Business

No. of Credits: 02	Teaching Scheme Practical: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basic knowledge of Computer concepts is assumed			
Objectives To build an understanding of the fundamental concepts of computer networking. - To introduce students to AI applications in business - To familiarize students with popular AI tools for automation, marketing, and decision-making. - To develop hands-on skills in using AI-driven business solutions			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the role of AI tools in business operations CO2: Utilize AI-powered tools for marketing, automation, and decision-making CO3: Apply AI solutions to business challenges through practical use cases			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to AI in Business	5	CO1
1.1 Basics of Artificial Intelligence (AI) and Machine Learning (ML) 1.2 Role of AI in Business Decision Making 1.3 AI-powered Business Automation: Advantages & Challenges 1.4 Introduction to AI Ethics and Bias 1.5 Case Study: AI Implementation in E-commerce and Banking Sectors			
2	AI-Powered Tools for Productivity & Automation	10	CO2
2.1 AI for Communication & Content Creation: ChatGPT, Google Bard, Grammarly, Canva AI 2.2 AI for Meetings & Transcription: Otter.ai, Fireflies.ai			
3	AI in Marketing & Customer Engagement	8	CO1, CO2
3.1 AI for Digital Marketing & SEO: Copy.ai, Jasper AI, SurferSEO 3.2 Chatbots & AI-Powered Customer Service: Drift, HubSpot AI, Salesforce Einstein			
4	AI for Business Analytics	7	CO3
4.1 AI for Data Visualization & Business Insights: Tableau AI, Power BI			

4.2 Predictive Analytics & Decision Making: AI in Finance & Risk Analysis
Reference Books
1. Artificial Intelligence in Business: Opportunities and Challenges" – Péter Szeredi & Attila Kiss 2. AI Superpowers: China, Silicon Valley, and the New World Order" – Kai-Fu Lee 3. The AI Advantage: How to Put the Artificial Intelligence Revolution to Work" – Thomas H. Davenport 4. 4Human + Machine: Reimagining Work in the Age of AI" – Paul R. Daugherty & H. James Wilson 5. Artificial Intelligence for Marketing: Practical Applications" – Jim Sterne
Resource Material/Other Online Courses
1. Google AI for Business (Google AI) 2. AI for Everyone (Coursera – Andrew Ng) 3. LinkedIn Learning: AI in Business Strategy

Savitribai Phule Pune University (Formerly University of Pune)

Bachelors Degree in Data Science (Faculty of Science and Technology)



Syllabi for B.Sc. (Data Science)-Second Year Sem-III and IV

(For Colleges Affiliated to Savitribai Phule Pune University)

Choice Based Credit System (CBCS) Syllabus
Under National Education Policy (NEP)

To be implemented from Academic Year 2025-26

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2025-26

SY (Level 5.0) SEMESTER III

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	C E	E E	Total
Major Core	DS-201-MJ-T	Database Management System	2	-	2	-	15	35	50
	DS-202-MJ-T	Data Structure-I	2	-	2	-	15	35	50
	DS-203-MJ-P	Lab Course on DS-201-MJ-T and DS-202-MJ-T	-	2	-	4	15	35	50
VSC	DS-221-VSC-T	Foundations of Data Science	2	-	2	-	15	35	50
FP/ OJT/ CEP	DS-231-FP	Mini Project	-	2	-	4	15	35	50
Minor	DS-241-MN-T	Probability Distribution and Modelling	2	-	2	-	15	35	50
	DS-242-MN-P	Lab Course on DS-241-MN-T	-	2	-	4	15	35	50
IKS	DS-201-IKS	Indian Knowledge System in Computing	2	-	2	-	15	35	50
GE/OE *	OE-201-DS-T* OE-202-DS-T* OE-203-DS-T* OE-204-DS-T*	• E commerce-I / • Web Design-I/ • Digital Marketing-I/ • AI for Everyone-I	2	-	2	-	15	35	50
AEC	AEC-201-T	From University Basket	2	-	2	-	15	35	50
CC	CC-201-T	From University Basket	2	-	2	-	15	35	50
Total			16	6	16	12			550

** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.*

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2025-26

SY (Level 5.0) SEMESTER IV

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	C E	E E	Total
Major Core	DS-251-MJ-T	Relational Database Management System	2	-	2	-	15	35	50
	DS-252-MJ-T	Data Structure-II	2	-	2	-	15	35	50
	DS-253-MJ-P	Lab Course on DS-251-MJ-T and DS-252-MJ-T	-	2	-	4	15	35	50
VSC	DS-271-VSC-P	Data Analytics	-	2	-	4	15	35	50
FP/OJT/CEP	DS-281-FP	Mini Project	-	2	-	4	15	35	50
Minor	DS-291-MN-T	Testing of Hypothesis and Sampling Distributions	2	-	2	-	15	35	50
	DS-292-MN-P	Lab Course on DS-291-MN-T	-	2	-	4	15	35	50
SEC	SEC-251-DS-T	Software Engineering	2	-	2	-	15	35	50
GE/OE *	OE-251-DS-T* OE-252-DS-T* OE-253-DS-T* OE-254-DS-T*	• E commerce-II / • Web Design-II / • Digital Marketing-II/ • AI for Everyone-II	2	-	2	-	15	35	50
AEC	AEC-251-T	From University Basket	2	-	2	-	15	35	50
CC	CC-251-T	From University Basket	2	-	2	-	15	35	50
Total			14	8	14	16			550

*** These subjects are offered to other faculty students under GE/OE vertical. The students of B.Sc. (Data Science) will opt the subjects offered by other faculty given in University Basket.**

Exit option: Award of UG Diploma in B.Sc. (Data Science) with 88 credits and an additional 04 credits as per University guidelines or else continue with Major and Minor.

Continue option: Third year will be continued with Data Science as Major and Statistics as a Minor subject.

Detail Syllabus

B.Sc. (Data Science)

Semester-III

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III
DS-201-MJ-T : Database Management System

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites Basic Knowledge of file system, storing data in file system and Operations on sets			
Objectives - To explain the basic principles of files. - A comprehensive introduction to the core concepts of database management systems, - To study the basic concept of Entity relationship model. - The process of data normalization and decomposition based on functional dependencies to create well-structured, efficient, and consistent database systems.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand basic concepts of file operations and organization. CO2: Basic concepts of DBMS, Data Models and Relational Data Model terminologies. CO3: Construct an Entity-Relationship (E-R) model from specifications. CO4: Formulate Queries using SQL and Relational Formal Query Languages, working with tables, applying and modifying constraints, join queries. CO5: Design Normalize database.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction of DBMS	5	CO1, CO2
Introduction, Definition of file, Basic file Operation, Types of file (Logical & Physical) Basic Concept and Definitions (Data, Information, Database), Definition of DBMS, File System Vs. DBMS, Application of DBMS, Purpose of Database Systems, Levels of abstraction, data independence, Structure of DBMS, DBMS users, Advantages and disadvantages of DBMS			
2	Entity-Relationship Model	9	CO2, CO3
Introduction to Data Models (E-R Models, Relational Model, Network Model, Hierarchical Model), Basic Concepts (entities, attributes, entity sets, relations, relationship sets), Constraints (Mapping Constraints, key constraints, referential Constraints, Unique, NULL, Check constraint), Strong and weak entities, Extended ER Features (generalization, specialization, Aggregation), Structure of Relational Databases (concepts of a table), Examples on E-R Model (Case studies)			

3	Structured Query Language (SQL)	10	CO4
Introduction, Basic Structure, DDL Commands, DML Commands, Form a Basic SQL Query Set Operations (union / intersection / except), Aggregate & String Function, NULL Values, Aggregate operators (Group by, having), Nested Queries (Set Membership & Set Comparison operator), SQL mechanisms for joining relations (inner joins, outer joins and their types) case study on SQL			
4.	Relational Database Design	6	CO5
Introduction to Schema Refinement, Problems Caused by Redundancy, Use of Decompositions, Problems Related to Decomposition, Functional Dependencies (Closure of a Set of FDs, Attribute Closure), Normalization forms (1NF, 2NF, 3NF, BCNF), Decomposition (lossless join, dependency preserving property)			
Reference Books			
1. Book1 Database System Concepts by Henry Korth and A. Silberschatz 2. Book2 Introduction to SQL by Reck F. van der Lans by Pearson 3. Book 3 Database Management System by Raghu Ramakrishnan			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-202-MJ-T : Data Structure-I

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites • Basic knowledge of algorithms and problem solving • Knowledge of Python Programming Language			
Objectives • To Know the Basic Concept of Data Structures and Algorithm • To Understand different Sorting and Searching Methods with their Time Complexity • To learn various Application of Array, Linked List, Stack, Queue • To know the good choice of Data Structure for Data Science Application			
Course Outcomes On Completion of this course, student will be able to - CO1: Understanding basic concept of Data Structure CO2: To use well-organized Data Structures in solving various problems. CO3: Develop Algorithms Using Stacks and Queues for Real-Time Applications CO4: To efficiently implement the different Data Structures CO5: To apply linear data structures			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Data Structure and Algorithm Analysis	4 Hrs.	CO1
Overview of Data Structures Need of Data Structure, Definitions: Data and Information, Data Type, Data Object, ADT, Data Structure, Types of Data Structures: Linear vs. Non-linear, Static vs. Dynamic Complexity Analysis Space and Time Complexity, Asymptotic Notation (Big O, Omega, Theta), Best, Worst, and Average Case Analysis			
2	Array as a Data Structure	7 Hrs.	CO2
Introduction to Arrays Representation of Arrays in memory (Contiguous Memory Allocation), Types of Array, Operations: Access, Insertion, Deletion, Traversing Applications of Array Searching (Linear Search, Binary Search), Comparison of searching Methods Sorting Terminology: Internal, External, Stable, In-Place Sorting, Types of Sorting Study of Comparison Based Sorting-Efficient Bubble Sort, Insertion Sort, Selection Sort			

Divide and Conquer Strategy – Merge Sort, Quick Sort			
3	Linked List	7 Hrs.	CO2,CO5
Introduction to Linked Lists List as a Data Structure, Comparison of Array and Linked list Static and dynamic Implementation of Linked List Types of Linked list -Singly Linked List, Doubly Linked List, Circular Linked List Node structure and pointer-based representation Operations on Linked Lists Create, Traversing, Insertion, Deletion, Searching, Reverse, Concatenate, Merge, Time Complexity of Operations Application of linked List Online Shopping Cart (E-Commerce) Managing Large Datasets in Big Data Processing LRU Cache Implementation (Optimized Data Retrieval) Memory-Efficient Data Structures for NLP			
4	Stack	6 Hrs.	CO3,CO4
Introduction Stack Operations – init(), Push(), isEmpty(), isFull(), Peek(), Pop(), Time Complexity of operations, Implementation –Static and Dynamic with Comparison Applications of Stack Data Preprocessing (Undo/Redo in Data Cleaning), Function Call Stack in Recursive Algorithms (Tree Traversal), Expression Evaluation in Data Analysis (Parsing Expressions) Reversing a Dataset (Time Series Data)			
5	Queue	6 Hrs.	CO3,CO4
Introduction Queue Operation-init(), enqueue(), dequeue(), IsEmpty(), isFull, Peek(), Time Complexity of Operations, Comparison with Stack Implementation –Static and Dynamic with Comparison Types of Queue : Linear Queue, Circular Queue, Priority Queue Dequeue Applications of Queue Data Processing Pipelines (Stream Processing) , Task Scheduling in Machine Learning, Web Scraping, Chatbots, Cloud Computing			
Reference Books			
1. Book1: Data Structures and Algorithms in Python by Michael T. Goodrich 2. Book2: Python Data Structures and Algorithms by Benjamin Baka 3. Book3: Algorithms in Python" by Robert Sedgwick and Kevin Wayne			

Savitribai Phule Pune University

B.Sc. Data Science (2024 Pattern)

Sem-III

DS-203-MJ-P : Lab Course on DS-201-MJ-T and DS-202-MJ-T (Database Management System and Data Structure-I)

No. of Credits 2	Teaching Scheme Practical 4 Hrs. / Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Knowledge of Database Management System, storing data in file system • Basic knowledge of algorithms and problem solving • Knowledge of Python Programming Language		
Course Objectives • Understand basic database management operations. • Design E-R Model for given requirements and convert the same into database tables. • Understand Basic Concept of Searching and Sorting Techniques • Understand Concept of Linear Data structure(Linked list, Stack, Queue) and their application		
Course Outcomes:- On completion of this course, students will be able to : CO1: Create database tables in postgresQL. CO2: Write and execute simple, nested queries. CO3: Implement Sorting and Searching Techniques CO4: Implement Linear Data Structure dynamically with their application		
Operating Environment: For DBMS: • Operating System: Linux or any relevant Operating System • DBMS: PostgreSQL For Data Structures: Python is cross-platform and can run on: • Windows (Windows 10/11) • Linux (Ubuntu, Fedora, Debian, etc.) • macOS (Latest versions) To write and execute Python programs efficiently, use the following: • VS Code (Lightweight, supports extensions) • PyCharm (Best for large projects, auto-suggestions) • Jupyter Notebook (Great for data visualization & step-by-step execution) • IDLE (Comes pre-installed with Python, good for beginners)		

• Spyder (Best for scientific computing)
Database Management System Assignment
Assignment 1
To create simple tables with only the primary key (PK) constraint (as a table level constraint & as a field level constraint) (include all data types)
Assignment 2
To create more than one table, with referential integrity constraint, FK constraint.
Assignment 3
To create one or more tables with following constraints, in addition to the first two constraints (PK & FK) • Check constraint • Unique constraint • Not null constraint
Assignment 4
To drop a table, alter schema of a table, insert / update / delete records using tables created in previous Assignments. (use simple forms of insert / update / delete statements)
Assignment 5
Computation on table using , aggregate functions ,string functions, special clauses (order by, group by, Having)
Assignment 6
To query table, using set operations (union, intersect) set membership operator (in ,not in)
Assignment 7.
To query tables using nested queries (use of 'Except', exists, not exists, all clauses)

Data Structure-I Assignments
Assignment 1: Searching Algorithms
Implementation of searching algorithms to search an element using: Linear Search, Search, Binary Search (with time complexity)
Assignment 2: Sorting Algorithms - I
Implementation of sorting algorithms: Bubble Sort, Insertion Sort, Selection Sort
Assignment 3: Sorting Algorithms - II
Implementation of sorting algorithms:

Quick Sort, Merge Sort
Assignment 4: Singly Linked List
1. Dynamic implementation of Singly Linked List to perform following operations: Create, Insert, Delete, Display, Search, Reverse 2. Create a list in the sorted order.
Assignment 5: Doubly Linked List
Dynamic implementation of Doubly circular Linked List to perform following operations: Create, Insert, Delete, Display, Search
Assignment 6: Linked List Applications
1. Merge two sorted lists. 2. Create Simple Music Playlist using Linked List
Assignment 7: Stack
Static and Dynamic implementation of Stack to perform following operations: Init, Push, Pop, Peek, Isempty, Isfull
Assignment 8: Applications of Stack
Data Preprocessing(Undo/Redo in Data Cleaning) Expression Evaluation in Data Analysis Tokenization in Natural Language Processing
Assignment 9: Linear Queue and Circular Queue
1. Static and Dynamic implementation of linear Queue to perform following operations: Init, enqueue, dequeue Peek, IsEmpty, IsFull. 2. Implementation of circular queue
Assignment 10 Application of Queue
1. Web Scraping (Handling Multiple request)
2. Customer Support Chatbots (Message Queues)
Reference Books Book 1 : Database System Concepts by Henry Korth and A. Silberschatz Book 2 : Introduction to SQL by Reck F. van der Lans by Pearson Book 3 : Database Management System by Raghu Ramakrishnan Book 4: Data Structures and Algorithms in Python by Michael T. Goodrich Book 5: Python Data Structures and Algorithms by Benjamin Baka Book 6: Algorithms in Python" by Robert Sedgwick and Kevin Wayne

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-221-VSC-T : Foundations of Data Science

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites • Knowledge of basic programming concepts • Knowledge of basic concepts in Mathematics and statistics • Knowledge of basic concepts in databases			
Objectives • To gain a comprehensive understanding of the core concepts, tools, and Methodologies used in large dataset. • To know how to Collect, pre-process, and explore datasets to extract meaningful insights. • To study statistical techniques and models to analyze data and draw conclusions. • To Understand Data visualization techniques to effectively communicate insights from complex data sets through charts and graphs.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the basic principles and concepts of data science and its relevance to data analytics CO2: Apply statistical techniques and models to analyze data and draw conclusions. CO3: Explain how data is collected, managed and stored for data science. CO4: Build, and prepare data for use with a variety of statistical methods and models. CO5: Analyze Data using various Visualization techniques.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Data Science	6	CO1
Data Science: - Definition, History, Benefits, scope and Applications of Data Science. The 3 V's: Volume, Velocity, Variety The Data Science Lifecycle :- Overview – Defining research goals – Retrieving data – Data preparation - Exploratory Data analysis – build the model– presenting findings and building applications Why learn Data Science? Real life usage of Data Science Systems More about Data Science: Data Analysis Vs. Data Analytic, Qualitative Analysis Vs. Quantitative Analysis Data Mining - Data Warehousing			
2	Statistical Analysis in data science	8	CO2
What is Statistical Analysis? Types of Data in statistics:-Qualitative , Quantitative :-Discrete, continuous Types of Statistical Analysis			

Descriptive Statistical Analysis: Measuring the Frequency Measuring the Central Tendency: Mean, Median, and Mode Measuring the Dispersion: Range, Standard deviation, Variance, Interquartile Range Measure of Relationship:-Covariance, Correlation Inferential statistics:-Parameter estimation for Statistical Inference Hypothesis:-Types of hypothesis, Hypothesis Testing Statistical Tests:-Parametric Tests-Z-test ,t-test ,F-test ,ANOVA ,Chi-square test, Non-parametric Tests Regression analysis			
3	Data Collection and Preprocessing	8	CO3,CO4
Types of Data: Structured, semi-structured, Unstructured Data Data sources:-Internal Data, External data:-Open Data ,Social media data , Standard Dataset Data Formats Data preprocessing:-Steps in data preprocessing Data Collection and Import, Data Exploration ,Data Cleaning:- Handling missing data, Removing duplicates, Correcting errors, Dealing with outliers Data Transformation:- Normalization or standardization, Encoding categorical variable, Feature engineering, Handling skewed data Data Reduction:- Dimensionality reduction, Numerosity Reduction, Data Compression Data Validation, Data Discretization, Imbalanced Dataset, Uses of Data Preprocessing, Advantages and Disadvantages of Data preprocessing.			
4	Visualization tools and software	8	CO5
Introduction to Data visualization Why Data visualization important? Data Scientist's Toolbox:- Python ,R-Programming ,SAS, Tableau public ,Microsoft Excel ,Rapid Miner, Tensor flow Types of visualization:- Chart, Graph, Geospatial, Info graphic , Dashboards ,Temporal data, Hierarchical data ,network data ,multi-dimensional data ,Univariate data, multivariate data Data visualization Software:- Tableau ,Looker studio , Zoho Analytics , Sisense , IBM Watson Analytics , Qlik Sense , Domo , Microsoft Power BI, Klipfolio , Dundas BI , Infogram, Chart Blocks Data visualization liabraries: - Matplotlib library, seaborn library, ggplot library, Bokeh library, plotly library, pygal library, geoplolib library, Altair library, Holoviews libraries. Graphs and Chart for Data Visualization:-Bar Graph , Line Graph, Histogram, scatter plot, Area chart, Pie charts, Bubble chart, Boxplot, Donut chart, Heat map, Radar Charts, Treemap Charts, Waterfall Charts			
Reference Books			
1.Data Science Fundamentals and Practical Approaches-Dr.Gypsi Nandi,Dr.Rupam Kumar Sharma 2. Data science concept and practice-Vijay kotu and bala Deshpande 3. Introducing Data Science Big Data, Machine Learning, And More, Using Python Tools-Davy Cielen ,Arno D. B. Meysman, Mohamed Ali 4. Hand-On Introduction to Data Science-Chirag Shah			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-231-FP : Mini Project (Using Excel & Advanced Excel)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Knowledge of Excel & Advanced Excel		
Objectives - Develop hands-on experience in using Excel and Advanced Excel for real-world applications. - Improve proficiency in data entry, formatting, and worksheet management. - Learn how to use functions and formulas in excel. - Learn data analysis and visualization using charts and pivot tables. - Learn use of macros for repetitive tasks. - Learn use of Excel for financial modeling, budgeting, and forecasting.		
Course Outcomes On Completion of this course, student will be able to - CO1: Develop Excel & Advanced Excel Skills. CO2: Format and manage worksheet. CO3: Use functions and formulas in excel. CO4: Use macros for repetitive tasks. CO5: Analyze & visualize data using charts and pivot tables CO6: Understand use of Excel for financial modeling, budgeting, and forecasting .		
Building an Excel project involves planning, structuring, and implementing various Excel features to meet specific requirements. The expectations for an Excel project depend on its complexity and purpose. Students can find data for analysis from variety of sources including online platforms like Data.gov, Kaggle and Google Dataset Search as well as collect data from Government agencies, research institutions, concern organizations and industries.		
Sr. No.	Guidelines for Mini Project using Excel & Advanced Excel	
1	Define the Purpose - What problem will the project solve? - Is it for data analysis, automation, financial modeling, or reporting? - Who will use the file (yourself, a team, or a client)?	
2	Data Structure & Formatting - Organize data using proper headers and columns. - Use tables to manage structured data.	

	• Apply consistent formatting for readability. • Use data validation to prevent incorrect entries.
3	Formulas & Functions • Implement basic formulas (SUM, AVERAGE, IF) Use advanced functions (FILTER, COUNTIF, SUMIF, TEXT, DATE functions).
4	Automation Using Macros • Automate repetitive tasks with macros.
5	Data Visualization & Dashboards • Use charts (bar, pie, line, pivot charts) for visualization. • Create interactive dashboards using slicers and pivot tables. • Apply conditional formatting to highlight key insights.
6	Security & Protection • Protect sensitive sheets and cells. • Use passwords to restrict access if needed. • Prevent accidental modifications using data validation and locked cells.
7	Documentation & User Guide • Provide clear instructions for users. • Add comments or notes to explain complex formulas. • Include a help sheet if required.
Examples of Mini Project using Excel & Advanced Excel	
1. Personal Budget Tracker 2. Customer Feedback Analysis 3. Meal Planner & Shopping List 4. Grade Book for Teachers 5. Event Planning Organizer 6. Inventory Management System 7. Sales Report and Forecasting Tool 8. Task List with Priority Levels And many other	

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-241-MN-T: Probability Distribution and Modelling

No. of Credits: 02	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basics of Probability, discrete and continuous random variable			
Objectives - To study probability distribution of (univariate and bivariate) continuous random variables, expectation and moments of probability distribution - To understand the concept of discrete and continuous probability distributions - To calculate the probabilities of discrete and continuous probability distributions			
Course Outcomes On Completion of this course, student will be able to – CO1: Identify the given function is probability density function (pdf) or not CO2: Calculate cumulative distribution function (cdf), mean, variance, Moment generating function (MGF), Cumulant generating function (CGF), skewness and kurtosis for univariate continuous random variable CO3: Calculate cdf, mean and variance of (XY), joint moments, MGF, CGF, skewness and kurtosis for bivariate continuous random variable CO4: Obtain marginal and conditional expectation and distribution CO5: Identify the distribution CO6: Calculate the probabilities of discrete and continuous probability distributions CO7: Fit distribution to the particular data CO8: Interpret the results after fitting			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Continuous Univariate Distributions	04	CO1, CO2
Continuous sample space: Definition, illustrations.			
Continuous random variable: Definition, pdf, cdf, properties of cdf (without proof), probabilities of events related to random variable.			
Expectation of continuous r.v., expectation of function of r.v. $E[g(X)]$, mean, variance, geometric mean, harmonic mean, raw and central moments, skewness, kurtosis, mean deviation about mean.			
MGF: Definition, properties (without proof). CGF, Mode, partition values			
Probability distribution of function of r. v. : $Y = g(X)$ using i) Jacobian of transformation for $g(.)$ monotonic function and one-to-one, on to functions, ii) Distribution function for $Y = X^2$, $Y = X $ etc., iii) M.G.F. of $g(X)$.			
2	Continuous Bivariate Distributions	04	CO3, CO4

Continuous bivariate random vector or variable (X, Y): Joint p. d. f., joint c. d. f., properties (without proof), probabilities of events related to random variables. Marginal and conditional distributions. Expectation of r.v. (X, Y), expectation of function of r.v. $E[g(X, Y)]$, joint moments, $Cov(X, Y)$, $Corr(X, Y)$, conditional mean, conditional variance, $E[E(X Y = y)] = E(X)$ & $E[E(Y X = x)] = E(Y)$. Theorems on expectation (without proof): i) $E(X + Y) = E(X) + E(Y)$, ii) $E(XY) = E(X)E(Y)$, if X and Y are independent, generalization to k variables. iii) $E(aX + bY + c), Var(aX + bY + c)$. Independence of random variables X and Y and also its extension to k random variables. MGF for bivariate distribution and its properties without proof. Probability distribution of transformation of bivariate r. v. $U = \phi_1(X, Y), V = \phi_2(X, Y)$.			
3	Standard Discrete Distributions	10	CO5, CO6, CO7, CO8
Uniform Distribution: Pmf of uniform distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of uniform probabilities, drawing a random sample from uniform distribution Bernoulli Distribution: Pmf of Bernoulli distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of Bernoulli probabilities Binomial Distribution: Pmf of binomial distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of binomial probabilities, drawing a random sample from binomial distribution, fitting of binomial distribution Poisson Distribution: Pmf of poisson distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of poisson probabilities, drawing a random sample from poisson distribution, fitting of poisson distribution Geometric Distribution: Pmf of geometric distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of geometric probabilities			
4	Standard Univariate Continuous Distribution	12	CO5, CO6, CO7, CO8
Uniform or Rectangular Distribution: Pdf of uniform distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of uniform probabilities Normal Distribution: Pdf of normal distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of normal probabilities, fitting of normal distribution Exponential Distribution: Pdf of exponential distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on calculation of normal probabilities, drawing a random sample from exponential distribution Gamma Distribution: Pdf of gamma distribution, mean, variance, moments, skewness, kurtosis, standard results and theorems, examples on transformation of random variable			
Reference Books 1. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta. 2. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics,			

(Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002.

3. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
4. Hogg, R. V. and Craig, A. T. , Mckean J. W. (2012), Introduction to Mathematical Statistics (Tenth Impression), Pearson Prentice Hall.
5. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
6. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
7. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentce Hall of India, New Delhi.
8. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East West Press.
9. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-242-MN-P : Lab Course on DS-241-MN-T
(Probability Distribution and Modelling)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Basics of Probability, discrete and continuous random variable		
Objectives Practical Implementation of - Probability distribution of (univariate and bivariate) continuous random variables, expectation and moments of probability distribution - Discrete and continuous probability distributions		
Course Outcomes On Completion of this course, student will be able to - CO1: Fit suitable discrete and continuous probability distributions to the data. CO2: Identify the suitable probability model for the population. CO3: Generate random samples from the discrete and continuous probability distributions.		
List of Assignments		
Sr. No.	Assignment	
1	Fitting of Binomial distribution and computation of expected frequencies.	
2	Fitting of Poisson distribution and computation of expected frequencies.	
3	Applications of Binomial distribution	
4	Application of Poisson and geometric distribution.	
5	Model sampling from Poisson and Binomial distributions.	
6	Fitting of normal distribution and computation of expected frequencies.	
7	Model sampling from exponential distribution using distribution function	
8	Generating random samples from normal distribution using (i) distribution function (ii) Box-Muller transformation.	
9	Applications of normal and exponential distribution.	
10	Problems on Bivariate Probability distributions.	
Reference Books		
10. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta.		

11. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002.
12. Gupta, S. P. (2002), Statistical Methods (Thirty First Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
13. Hogg, R. V. and Craig, A. T. , McKean J. W. (2012), Introduction to Mathematical Statistics (Tenth Impression), Pearson Prentice Hall.
14. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
15. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
16. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentce Hall of India, New Delhi.
17. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East West Press.
18. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

DS-201-IKS: Indian Knowledge System in Computing

No. of Credits: 02	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Objectives - To introduce Vedic mathematical techniques and their relevance to modern computational methods. - To understand Nyaya’s logical framework and its application in reasoning and AI. - To explore the algorithmic structure of Panini’s grammar and Chandasastra’s binary system in computational linguistics and mathematics. - To explore real-world applications of IKS concepts in computational sciences.			
Course Outcomes On Completion of this course, student will be able to – CO1: Understand the computational foundations of Indian Knowledge Systems by applying Vedic mathematical techniques in problem-solving. CO2: Use Nyaya’s logical reasoning in AI and decision-making. CO3: Explore the connection between Panini’s grammar and NLP technologies. CO4: Recognize the applications of IKS in modern computing fields.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Vedic Mathematics & Computational Thinking	8	CO1
1.1 Introduction to Vedic Mathematics: Origins and importance in ancient India, Sutras and their logical foundation			
1.2 Basic Arithmetic using Vedic Methods: Addition, subtraction, multiplication, and division tricks			
1.3 Algebraic Applications of Vedic Mathematics: Squaring, square roots, cube roots, and factorization			
2	Introduction to Nyaya (Indian Logic)	8	CO2
2.1 Introduction to Nyaya Philosophy: Introduction to Nyaya (Indian Logic), Overview of Indian philosophical schools, Importance of Nyaya in logical reasoning, Types of reasoning (Anumana, Pramana, etc.)			
2.2 Nyaya’s Four Sources of Knowledge (Pramāṇa): Perception, inference, comparison, verbal testimony			

2.3 Types of Argumentations in Nyaya Vada (truth-based), Jalpa (debate-focused), Vitanda (criticism)			
2.4 Applications in AI & Machine Learning: Logical reasoning models, expert systems, and rule-based AI			
3	Panini's Astadhyayi & Chandasāstra	8	CO3
3.1 Introduction to Panini's Astadhyayi: Historical background and linguistic importance 3.2 Rule-Based System of Sanskrit Grammar: Sutras, meta-rules, recursion, and transformations 3.3 Chandasastra's Binary logic and combinatorial techniques			
4	Applications of IKS in Computer Science	6	CO4
4.1 Mind and cognition in Samkhya and Yoga: AI insights 4.2 Machine Learning and Indian philosophies: Understanding of human cognition in Indian philosophical schools (Advaita, Samkhya and Yoga) 4.3 Cryptography and Security: Ancient cryptographic methods in Kautilya's Arthashastra, protecting information: analogies from Indian traditions			
Reference Books			
1. Vedic Mathematics, Jagadguru Swami Bharati Krishna Tirtha, Motilal Banarsidass Publishing House, New Delhi. 2. "The Power of Vedic Maths" – Atul Gupta, JAICO publishing 3. Nyaya Theory of Knowledge" – S.C. Vidyabhusana 4. "A Primer of Indian Logic" – Kuppuswami Sastri, Hassell Street Press.2021 5. "Indian Logic: A Reader" – Jonardon Ganeri 6. "Aṣṭādhyāyī of Pāṇini" (Volumes 1 & 2) – Rama Nath Sharma, Munshirm Manoharlal publication 7. "Panini: His Work and Its Traditions" – George Cardona, Motilal Banarsidass Publishing House 8. "The Mathematics of Metre" – Satyanarayana Das 9. "Samkhya and Science" – Debabrata Sen Sharma 10. Explores the cognitive science aspects of Samkhya and Yoga in AI research. 11. "AI and Indian Philosophy" – Sangeet Kedia 12. "Kautilya's Arthashastra" – R. Shamasastri (Translation) 13. "History of Indian Cryptography" – Subhash Kak 14. Discusses coded messages, steganography, and security concepts in ancient India. 15. Saubhagya Vardhan, AI in Land of Vedas, Notion Press, 2023			

Savitribai Phule Pune University

B.Sc. Data Science (2024 Pattern)

Sem-III

Course Type: GE/OE Course Code: OE-201-DS-T

Course Title: Ecommerce I

(To be offered to faculty other than Science and Technology)

Teaching Scheme 2 Hours /Week	No. of Credits: 2	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks
----------------------------------	-------------------	---

Prerequisites

- Basic Computer and Internet Knowledge
- Fundamentals of Business Commerce and Digital Marketing Basics
- Financial and Payment Systems

Course Objectives

- To understand basic concepts about e-Commerce.
- To understand the applications of e-Commerce.
- To learn business model knowledge.
- To enable knowledge about E-payment system.
- To get a general idea of M-commerce

Course Outcomes

On completion of the course, student will be able to–

CO1: Implementation of basic concepts and application of e-Commerce

CO2: Use of electronic payment system.

CO3: Implementation of Business model knowledge

CO4:M-commerce technology use

Course Contents

Chapter 1	E- Commerce and Business Model Concepts	7 Hours
1.1. Main Activities of E Commerce 1.2. Definition 1.3. Goals 1.4. Technical Components 1.5. Functions 1.6. Status 1.7. Prospects 1.8. Significance 1.9. Advantages 1.10. Disadvantages E-Commerce Business Models 1.11. Major Business to Consumer (B2C)Business Model Portal, E-tailor 1.12. Major Business to Business (B2B) Business Model 1.13. E Distributor, E-Procurement, Exchanges 1.14. Business models in Emerging E-Commerce Areas - C2C, P2P, and B2G.		
Chapter 2	E-Marketing and E- Commerce Application	7 Hours
2.1. Identifying Goals 2.2. Browsing Behaviour Model		

2.3. Online Marketing 2.4. e-Commerce and retailing 2.5. e-Commerce and banking, 2.4. E Advertising 2.5. Internet Marketing Trends 2.6. Target Markets 2.7. E-Branding 2.8. Marketing Strategies 2.9. Consumer Online: The Internet Audience and Consumer Behaviors 2.10. E-cycle of Internet Marketing		
Chapter 3	E-commerce Payment Method	6 Hours
3.1. The requirements of an electronic payment system, 3.2. The constraints of a traditional payment system, and an overview of electronic payment technology. 3.3. Electronic payment gateways 3.4. B2B electronic payments. 3.5. Third-party payment processing electronic or digital currency, its characteristics, and its operation. 3.6. System for online credit card payments and smart cards.		
Chapter 4	Title E-Commerce Laws & Regulations	5 Hours
4.1.Introduction to E-Commerce Laws 4.2.Information Technology (IT) Act, 2000 (India) 4.3.GDPR (General Data Protection Regulation) – EU 4.4.Other Global E-Commerce Laws.		
Chapter 5	Title Future Trends in E-Commerce	5 Hours
5.1. AI & Chatbots in E-Commerce 5.2. AR/VR Shopping Experiences 5.3. Personalization & Data Analytics 5.4. The Role of IoT in E-Commerce		
Reference Books:		
1. Kenneth C. Laudon, E-Commerce : Business, Technology, Society, 4th Edition, Pearson 2. S. J. Joseph, E-Commerce: an Indian perspective, PHI 3 E-Commerce Law: National and Transnational Topics– Alan Davidson 4. Artificial Intelligence in E-Commerce– Richard Boire		

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

Course Type: GE/OE Course Code : OE-202-DS-T

Course Title: Web Design I

(To be offered to faculty other than Science and Technology)

Teaching Scheme 02 Hrs/ week	No. of Credits 2	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks
---------------------------------	---------------------	---

Prerequisites

- Basic computer knowledge and the ability to work with files.
- Knowledge of other programming languages will enhance your abilities with HTML and web development.
- Knowledge and understanding of Internet.

Course Objectives

- To learn HTML tags and programming concepts and techniques.
- To develop the ability to logically plan and develop web pages.
- To learn to write, test, and debug web pages using HTML.
- To learn to design Style sheets.
- To know to link and publish Web pages.

Course Outcomes

On completion of the course, student will be able to—

- Learn and use the HTML Tags.
- Understand and resolves written HTML codes.
- Design and develop the page using HTML codes.

Course Contents

Chapter 1	Introduction to Web Design	8
1.1 Introduction 1.2 Working of the Internet 1.3 Role of Web Servers, Clients(Communication) 1.4 Web Browsers 1.5 Working of the Internet , Intranet and WWW 1.6 E-Mail Servers and Protocols 1.7 E-mail Clients and Web Based Mail Access using Browser 1.8 Messenger Services and Clients(Chat) 1.9 Advantages and Disadvantages of Internet 1.10 Concept of effectiveWeb Design (Web site, classification of website, Advantages and Disadvantages. Of website) 1.11 Fundamental Principles of Web page design and issues		

Chapter 2	Getting Started with HTML	6
2.1 Introduction to scripting Languages 2.2 HTML Editing Tools 2.3 WYSISYG Authoring Tools 2.4 My First HTML Script 2.5 Basic HTML Document Structure 2.6 Common HTML Tags and it's attributes 2.7 Design HTML Tags 2.8 Text Formatting and Styles 2.9 Images and Graphics 2.10 Button, Formatting and Style 2.11 Lists 2.12 Hyperlinks 2.13 Multimedia 2.14 Frames 2.15 HTML Forms 2.16 Linking Web pages 2.17 Publishing Web Pages		
Chapter 3	Tables	6
3.1 The summary of Table tags 3.2 Introduction to tables 3.3 The Basic table structure 3.4 Affecting table appearance 3.5 Table troubleshooting 3.6 Tips and tricks 3.7 Standard table templates 3.8 Multipart images in tables		
Chapter 4	Frame / Forms	6
4.1.Introduction to frames 4.2.The basic frameset structure 4.3.The frame function and appearance and Targeting frames 4.4.The Inline (Floating) frames and Frame design tips and tricks 4.5.Introduction to Forms and The basic form (FORM) 4.6. The FORM elements and FORM attributes 4.7.Unconventional use of FORM elements 4.8.Demystifying CGI 4.9.Retrieving the parameter value using getParameter () method		
Case Studies		4
Case study 1: Creation of forms, small case study to create HTML pages using all the above learnt techniques.		

Case study 2: Creation of Forms layout designing by using div element with CSS property

Case study 3: Create Multiple Web pages link them to publish a small website.

Reference Books:

1. Computer Programming For Beginners:Learn The Basics Of HTML5-Joseph Connor
2. The Complete Reference HTML & CSS-Fifth Edition-Thomas A.Powell
3. Learning Web Design: A beginner's Guide To HTML, CSS, Javascript, and Web Graphics - Jennifer Robbins
4. HTML5: The Missing Manual - Matthew MacDonald.

Reference Link :

1. [HTML reference - HTML: HyperText Markup Language | MDN](#)
2. [HTML Standard](#)

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

Course Type: GE/OE. Course Code: OE-203-DS-T

Course Title: Digital Marketing I

(To be offered to faculty other than Science and Technology)

Teaching Scheme 02 Hours /Week	No. of Credits 2	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks
-----------------------------------	------------------	---

Prerequisites

- Creative & Logical thinking ability,
- Digital devices operational skills and Knowledge

Course Objectives

- To understand Digital Marketing as the most powerful marketing tool.
- Learn to create digital marketing artworks.
- Learn how to use email campaigns, blogging to produce worthwhile, pertinent material that draws in and engages a target audience.

Course Outcomes

On completion of the course, student will be able to–

CO1: Students will have improved their visualization power to explore new ideas.

CO2: Students will learn to develop their Marketing skills.

CO3: Student will making sure that tactics are in line with corporate objectives and produce quantifiable outcomes from e-payment mechanisms.

CO4: Student will be in line with target demographics and company objectives, increasing brand awareness and boosting conversions.

Course Contents

Chapter1	History of Digital Marketing	8
1.1.What is Digital Marketing ? 1.2.Offline digital marketing - Electronic billboards, Radio marketing, T.V. marketing, Phone marketing 1.3.Online marketing - Search Engine Optimization (SEO), Social media marketing, E-mail marketing 1.4.Difference between conventional marketing and online marketing.		
Chapter2	Internet Marketing	8
2.1. Structure of Website - Team 2.2. Types of website - Static Website, Dynamic website, Personal, Commercial, Governmental, Non- profit organisation 2.3. Web Portals - Type of Portals		
Chapter3	Classification of e-Commerce	8
3.1. Business to Business (B2B) Model 3.2. Business to Consumer(B2C) Model		

3.3. Consumer to Consumer(C2C) Model	
3.4. Consumer to Business (B2B) Model	
Case Study	6
Case Study 1 Case Study 2 Case Study 3	
Reference Books:	
1. SEO 2025: Learn Search Engine Optimization with Smart Internet Marketing Strategies" by Adam Clarke 2. Digital Marketing : Nitin Kamat, ChinmayKamat (Himalaya Publishing House) 3. "Made to Stick: Why Some Ideas Survive and Others Die" by Chip Heath and Dan Heath 4. "Digital Marketing: Strategy, Implementation, and Practice" by Dave Chaffey and Fiona Ellis-Chadwick	
Reference Links:	
1. https://www.coursera.org/browse/business/digital-marketing 2. https://learndigital.withgoogle.com/digitalgarage	

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-III

Course Type: GE/OE. Course Code: OE-204-DS-T

Course Title: AI for Everyone-I
(To be offered to faculty other than Science and Technology)

Teaching Scheme: 2 hours / week	No. of Credits: 2	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks
------------------------------------	----------------------	---

Course Objectives: -

1. Understand the basics of artificial intelligence and its subfields.
2. Explore real-world applications of AI across different industries.
3. Gain insights into the ethical, social, and economic implications of AI.
4. Develop an appreciation for the potential of AI to drive innovation and transformation.

Course Outcomes: -On completion of the course, student will be able to–

CO1: Define and explain the fundamental concepts and subfields of AI.
CO2: To understand the potential of AI to drive innovation and transformation in different domains.
CO3: Identify real-world applications of AI across various industries.
CO4: Analyse the ethical, social, and economic implications of AI.

Course Contents

Unit 1	Introduction to Artificial Intelligence	8 hours	CO1
1.1 Definition and scope of AI 1.2 Historical overview and key milestones 1.3 Differentiating AI from human intelligence 1.4 Types of AI tools: Text, image, audio, video, coding, and automation. 1.5 Where to find free AI tools? (Google AI, OpenAI, Hugging Face, etc.)			
Unit 2	AI Subfields	6 hours	CO2
2.1 Machine learning: Supervised, unsupervised, and reinforcement learning 2.2 Deep learning and neural networks 2.3 Natural language processing (NLP) and computer vision			
Unit 3	Applications of AI	8 hours	CO3
3.1 AI in healthcare: Diagnosis, treatment, and medical imaging 3.2 AI in finance: Fraud detection, algorithmic trading, and risk assessment 3.3 AI in transportation: Autonomous vehicles and traffic optimization 3.4 AI in customer service and chatbots 3.5 AI in education: Personalized learning and intelligent tutoring systems			

Unit4	Ethical and Social Implications of AI	8 hours	CO4
4.1 Bias and fairness in AI systems. 4.2 Privacy and data protection concerns 4.3 Impact of AI on employment and the workforce 4.4 AI and social inequality			
Reference Books:			
1. Artificial Intelligence: A Guide for Thinking Humans" – Melanie Mitchell 2. The AI Revolution in Medicine: GPT-4 and Beyond" – Peter Lee, Carey Goldberg, Isaac Kohane 3. AI 2041: Ten Visions for Our Future" – Kai-Fu Lee, Chen Qiufan 4. The Business of AI: AI Technologies and How to Leverage Them for Business Success" – Anirudh Koul 5. AI-Powered Marketing: Harness the Future of Marketing with AI" – Peter Gentsch 6. The AI Marketing Handbook" – Ryan McKenzie			

Detail Syllabus

B.Sc. (Data Science)

Semester-IV

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-251-MJ-T: Relational Database Management System

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites Basic Knowledge of DBMS, Knowledge of SQL Queries, Basics of relational design			
Objectives - To teach the essential concepts of RDBMS with PL/PgSQL - Understand the need of transaction processing and learn techniques for controlling the consequences of concurrent data access. - Learn how different recovery techniques restore the database to a consistent state after failures			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the basic syntax and structure of PL/SQL language. CO2: Analyze database transactions and can control them by applying ACID properties. CO3: Demonstrate the basic concepts of transaction processing and concurrency control. CO4: Understand concurrent transactions affect the recovery process. CO5: Analyze various databases.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to PL/SQL	10	CO1
PL/SQL: Structure, Elements, Data types, Control Structure, Iterative Control, Views, Cursors - Stored Procedure and Function, Exceptional Handling and Errors, Triggers.			
2	Transaction Management	5	CO2
Define Transaction, Properties of Transaction, States of Transaction , The problems associated with concurrently executing transactions What is Schedule Types of Schedule Concept of Serializability, Precedence graph for serializability			
3	Concurrency Control Techniques	7	CO3
Ensuring serializability by locks, Lock Based Protocols ,Different locks modes, 2PL & it's Variation, Timestamp Based Protocol, Thomas Writes rule, Multiple Granularity Phantom problem, Deadlock & Deadlock handling technique.			

4.	Database Recovery Technique	6	CO4
Recovery concepts, Log based Recovery technique (Deferred and Immediate), Checkpoints, Shadow paging, Aries Algorithm, Recovery with concurrent Transactions. Database backup and Recovery from Catastrophics failures.			
5	Advanced Topics in Database	2	CO5
Parallel Database, Distributed Database, Multimedia Database , Mobile Database , Web Database , Multidimensional Database, Data Warehouse , OLTP Vs OLAP, Bigdata, NoSQL Databases			
Reference Books			
1. Book1 - Database Management Systems, Raghu Ramakrishnan and Johannes Gehrke 2. Book2 - “Fundamentals of Relational Database Management System”, Sumathi, S. Esakkirajan 3. Book 3 - Database System Concepts by Henry Korth and A. Silberschatz 4. Book 4 - Practical Postgresql , By Joshua D. Drake, John C Worsley (O’Reillypublications) 5. Book 5 - PostgreSQL, Korry Douglas, ISBN:9780672327568, Sams			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-252-MJ-T : Data Structure II

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites • Basic knowledge of Python Programming. • Fundamental Knowledge of Array, Linked List, Stack, Queue Data Structure • Familiarity with basic Searching and Sorting techniques			
Objectives • To understand advanced data structures and their applications in Data Science. • To analyze and implement complex algorithms using Python. • To improve problem-solving skills by applying data structures to real-world problems. • To introduce students to efficient searching, sorting, and graph-based techniques. • To enhance the ability to optimize performance in large-scale data-driven applications			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the Fundamental Concepts of Non-Linear Data Structures CO2: Develop optimized algorithms for data science applications CO3: Apply graph algorithms for solving real-world problems. CO4: Evaluate the efficiency of algorithms using Time and Space complexity.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Tree	10	CO1
Concept and Terminologies Types of Binary Tree-Binary Tree, Skewed tree, Strict Binary Tree, Full Binary Tree, Complete Binary Tree, Expression tree, Binary Search Tree, Heap			
Representation Static and Dynamic Representation of tree			
Implementation and Operations on Binary Search Tree Create, Insert, Delete, Search Tree Traversal-preorder, inorder, postorder ,Level order traversal using queue Counting Leaf node, Non leaf nodes, Copy, Mirror			
Application of Trees Heap Sort Introduction to Greedy Strategy, Huffman encoding (Implementation Using Priority Queue)			
Applications of tree in Data Science Decision tree in machine learning			

Hierarchical clustering in data analysis			
2	Efficient Search Tree	4	CO1,CO2
Terminology- Balanced trees-AVL Trees, Red Black tree, Splay Tree, Lexical Search tree-Trie, Multiway search Trees-B and B+ tree AVL Tree- Concept and Rotations			
3	Graph	10	CO1,CO2,CO3, CO4
Concept and Terminology Graph representation-Adjacency Matrix, Adjacency List Graph Traversals-Breadth First Search, Depth First Search (with implementation) Applications of Graph - Topological Sorting Use of Greedy Strategy in Minimal Spanning Trees(Prim's and Kruskal Algorithm) Single Source Shortest Path-Dijkstra's Algorithm Dynamic Programming Strategy, All Pair shortest Path(Floyd Warshall Algorithm) Applications of Graph in Data Science Use of Graph in Social Network Graph based Machine Learning Transportation & Route Optimization			
4	Hash table	6	CO1, CO2
Concept of Hashing Terminologies-Hash Table, Hash Function, Bucket, Hash address, Collision, Synonym, Overflow Properties of good Hash function Hash Function-Division function, MID Square, Folding Method Collision Resolution Techniques Open Addressing-Linear Probing, Quadratic Probing, Rehashing Chaining: Coalesced, Separate Chaining Applications of Hash Table in Data Science Efficient Lookup in Datasets : Searching specific values or attributes in large datasets Counting Frequencies : Counting frequencies of words in large datasets Caching Results : Storing computed results of expensive operations for reuse			
Reference Books			
5. Data Structures and Algorithms in Python Authored by Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser 6. Data Structures Using Python by Shriram K. Vasudevan, 7. Data Structures and Algorithm Analysis in Python by Mark Allen Weiss 8. Python Data Structures and Algorithms by Benjamin Baka			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-253-MJ-P : Lab Course on DS-251-MJ-T and DS-252-MJ-T
(Relational Database Management System and Data Structure-II)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Course Objectives - To solve real world computational problems. - To perform operations on relational database management systems. - Understand the Fundamental Concepts of Non-Linear Data Structures - Develop optimized algorithms for data science applications		
Course Outcomes On completion of this course students will able to: CO1: To use PostgreSQL (PL/SQL.) CO2: To perform advanced database operations (View ,Function, cursor, Trigger) CO3: Implement Fundamental Concepts of Non-Linear Data Structures CO4: Implement nonlinear Data Structure dynamically with their application		
Operating Environment: For DBMS: - Operating System: Linux or any relevant Operating System - DBMS: PostgreSQL For Data Structures: - Python is cross-platform and can run on: Windows (Windows 10/11) Linux (Ubuntu, Fedora, Debian, etc.) macOS (Latest versions) To write and execute Python programs efficiently, you can use: VS Code (Lightweight, supports extensions) PyCharm (Best for large projects, auto-suggestions) Jupyter Notebook (Great for data visualization & step-by-step execution) IDLE (Comes pre-installed with Python, good for beginners) Spyder (Best for scientific computing)		
Course Contents:		
Relational Database Management Systems Assignments		
Assignment 1: Views		

1) To Create and Drop Views
Assignment 2 : Stored Procedure
1) Simple Stored Procedure 2) Stored Procedure with IN, OUT and IN/OUT parameter
Assignment 3: Stored Function
1) Simple Stored Function 2) Stored Function that returns
Assignment 4 : Cursors
1) Simple Cursor 2) Parameterize Cursor
Assignment 5 : Exception Handling
1) Simple Exception- Raise Debug Level Messages 2) Simple Exception- Raise Notice Level Messages 3) Simple Exception- Raise Exception Level Messages
Assignment 6 : Triggers
1) Before Triggers (insert, update, delete) 2) After Triggers (insert, update, delete)
Assignments for Data Structures – II
Assignment 1: Binary Search Tree and Traversals
1. Implement Binary Search Tree (BST) to perform following operations on BST– Create, Recursive Traversals - Inorder, Preorder, Postorder 2. Perform following operations: insert, delete
Assignment 2: Binary Search Tree Operations
1. Implement Binary Search Tree (BST) to perform following operations on BST– Copy and mirror image of BST, counting leaf, non-leaf and total nodes. 2. Level-order traversal of binary search tree using queue.
Assignment 3: Applications of Binary Tree
1. Sort set of elements using Heap sort 2. Encode a set of characters using Huffman encoding
Assignment 4: Graph implementation
1. Implement Graph as adjacency matrix and adjacency list 2. Calculate indegree and out degree of vertices 3. Graph traversals: BFS and DFS.
Assignment 5: Graph Applications - I
1. Implementation of Topological sorting 2. Implementation of Prims/Kruskals Minimum spanning tree algorithm
Assignment 6: Graph Applications - II
1. Implementation of Dijkstra's shortest path algorithm for finding Shortest Path from a given source vertex using adjacency cost matrix. 2. Implementation of Floyd Warshall algorithm for all pairs shortest path.
Assignment 7: Hash Table
1. Implementation of static hash table with Linear Probing. 2. Implementation of static hash table with chaining

Books
1. Database Management Systems, Raghu Ramakrishnan and Johannes Gehrke 2. “Fundamentals of Relational Database Management System”, Sumathi, S. Esakkirajan 3. Database System Concepts by Henry Korth and A. Silberschatz 4. Practical Postgresql , By Joshua D. Drake, John C Worsley (O’Reillypublications) 5. PostgreSQL, Korry Douglas, ISBN:9780672327568, Sams 6. Data Structures and Algorithms in Python Authored by Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser, 7. Data Structures Using Python by Shriram K. Vasudevan, 8. Data Structures and Algorithm Analysis in Python by Mark Allen Weiss 9. Python Data Structures and Algorithms by Benjamin Baka

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-271-VSC-P: Data Analytics

No. of Credits: 02	Teaching Scheme Practical: 4 Hours/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks
Prerequisites • Basic knowledge of logic and Python programming concepts • Knowledge of problem-solving tools like algorithms, flowcharts and pseudo codes will be an added advantage		
Objectives • To explore the fundamental concepts of data analytics. • To understand the various methods and tools for data analytics. • Design a simple database. • Query a database using SQL in Python		
Course Outcomes On Completion of this course, student will be able to - CO1: Use different data analytics functions to solve a problem. CO2: Apply different data cleaning functions on a given data. CO3: Explain basic excel formulas and functions and apply them on a data set. CO4: Design a simple database with DDL and DML commands. CO5: Write sub queries and join operations for retrieving data from various tables. CO6: Use the aggregations and group operations for data analysis in python.		
Guidelines: Operating Environment: Excel, PostgreSQL, Python		
List of Assignments:		
Excel: Assignment 1 : Getting Started with Excel • Creation of spread sheets, Insertion of rows and columns, Drag & Fill, use of Aggregate functors. Assignment 2 : Working with Data • Importing data, Data Entry & Manipulation, Sorting & Filtering, Data Validation, Pivot Tables & Pivot Charts Assignment 3 : Data Analysis Process • Conditional Formatting, What-If Analysis, Data Tables, Charts & graphs Assignment 4 : Cleaning Data with Text Functions and Cleaning Data Containing Date and Time Values • Use of UPPER and LOWER, TRIM function, Concatenate, use of DATEVALUE function, DATEADD and DATEDIF, TIMEVALUE functions. Assignment 5 : Conditional Formatting • Formatting, parsing, and highlighting data in spreadsheets during data analysis.		

Assignment 6 : Working with Multiple Sheets • Work with multiple sheets within a workbook is crucial for organizing and managing data, perform complex calculations and create comprehensive reports Assignment 7 : Generation of report & presentation using Autofilter & macro
SQL (Using Python) Connect to PostgreSQL database using Python and Write SQL queries. • Filtering and sorting data (SELECT, WHERE, ORDER BY), Aggregation (SUM(), AVG(), COUNT()), Joining multiple tables (JOIN), Fetch results and analyze them in Pandas.
Python: Assignment 1: Load a dataset and perform some primary operations. • info(), describe(), shape, size, loc, sort_values, value_counts() used to describe the dataframe, Handle missing values, Remove duplicate records, Standardize column names • Getting introduced to essential packages like NumPy, SciPy, and pandas from the Data Analytics point of view. Assignment 2 : Basic statistical operations • Apply basic statistical operations on a dataset- Compute the mean, median, mode, range, quartiles, and variance for one or more attributes. Assignment 3 : Data pre-processing Apply data pre-processing techniques • Describing data set • Shape the data set • Displaying specific rows from the data set • Partition them into appropriate number of bins by equal-frequency as well as equal-width partitioning. Assignment 4 : Data Visualization • View the data using various 2-D, 3-D plots and charts to detect and handle outliers
Reference Books 1. How to solve it by Computer, R.G. Dromey, Pearson Education. 2. Programming Python, O`Reilly, 4th Edition, 2010

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-281-FP : Mini Project
(Software Engineering Project)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Fundamental Knowledge of Software Engineering, RDBMS		
Objectives - Apply software engineering principles to design, develop a small-scale software project. - Demonstrate understanding of software development life cycles, methodologies, and tools. - Develop teamwork, communication and problem-solving skills.		
Course Outcomes On Completion of this course, student will be able to - CO1: Understand software engineering discipline. CO2: The project should have a clear set of requirements, design documents. CO3: The project can be developed using any process model.		
Unit No.	Name of Unit	
1	Project Proposal	
	- Project title and description - Project goals and objectives - Technical specifications and requirements	
2	Problem Definition and Scope	
	- Problem Description - Study of Existing system(Manual or computerized) - Drawbacks of Existing system - Scope of the Proposed System	
3	Feasibility Study	
	- Technical Feasibility - Economical Feasibility - Operational Feasibility	
4	Fact Finding Techniques	
	- Interview - Questionary - Survey - Observations	

5	Entity Relationship Diagram
	• Identify entities & attributes • Identify relations in entities
6	UML Diagrams
	• Class Diagram • Use case Diagram • Sequence Diagram • Activity Diagram • Component Diagram • Deployment Diagram
7	Data Dictionary
	• Designing the Database • Normalized to 3NF
8	Designing queries related to Functional requirements
	• Simple Queries • Nested Queries(using Aggregate Functions) • Views • Functions • Cursors • Trigger
Reference Books	
1. Book1:Software Engineering”, by Roger Pressman, 8 th Edition 2. Book2: Practical PostgreSQL”, O’Reilly Publications	

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-291-MN-T : Testing of Hypothesis and Sampling Distributions

No. of Credits: 02	Teaching Scheme Theory: 2 Hours/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites Basics of Distribution			
Objectives - To study the hypothesis - To perform the large sample test - To overview the sampling distribution - To understand the difference between small sample and large sample test - To study hypothesis testing based on sampling distributions			
Course Outcomes On Completion of this course, student will be able to – CO1: Identify the difference between statistic, estimate and parameter. CO2: Formulate the null and alternative hypotheses CO3: Calculate type-I and type-II error, p-value CO4: Apply small, large sample tests in real life problems CO5: Derive probability distribution function of chi-square, t, F distribution CO6: Explains interrelation between the above distributions and their properties. CO7: Get familiar with statistical tests of hypothesis and are able to apply in real life situations in various fields			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Testing of Hypothesis	08	CO1, CO2, CO3
Statistic, estimate and parameter. Sampling distribution of a statistic, standard error of a statistic with illustrations. Statistical Inference: Introduction to problem of Estimation and testing of hypothesis. Difference between estimator and estimate. Concept of unbiased estimator. Point and interval estimation. Statistical hypothesis, null and alternative hypothesis, simple and composite hypothesis, one sided and two-sided alternative hypothesis, critical region, type-I and type-II error, level of significance, p -value. Two-sided confidence interval. Tests of hypotheses using i) critical region approach, ii) p -value approach and iii) confidence interval approach. Tests for population means (large sample / approximate tests): Testing of population mean (μ) with a specified value (μ_0) for the known variance. Testing equality of two population means ($\mu_1 = \mu_2$) for the known variance. Construction of two-sided confidence interval for population mean μ and for the mean difference $\mu_1 - \mu_2$.			

Testing of population proportion (P) with a specified value (P_0). Testing equality of two population proportions ($P_1 = P_2$). Construction of two-sided confidence interval for P and $P_1 - P_2$.			
2	Chi-square Distribution	08	CO5
Definition of chi-square random variable as a sum of squares of independent and identical standard normal variables. Derivation of the p.d.f. of Chi-square variable with n degrees of freedom (d.f.) using MGF. Mean, variance, MGF, CGF, central moments skewness, kurtosis, mode, additive property. Use of chi-square tables for calculations of probabilities. Normal approximation: $\frac{\chi_n^2 - n}{\sqrt{2n}}$ (statement only) Distribution of $\underline{X}$ and $\frac{nS^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \underline{X})^2$ for a random sample from a normal distribution using orthogonal transformation, independence of $\underline{X}$ and S^2 .			
3	t -Distribution	04	CO5
Definition of r.v. with n d.f. in the form of $= \frac{U}{\sqrt{\frac{V}{n}}}$, where $U \sim N(0,1)$ and V is chi-square with n d.f., where U and V are independent random variables. Derivation of the p.d.f of distribution, nature of probability curve, mean, variance, moments, mode. Use of t -tables for calculations of probabilities, Normal approximation (statement only).			
4	Snedecore's F -Distribution	05	CO5
Definition of r.v. with n_1 and n_2 d.f. as $F_{n_1, n_2} = \frac{X_1/n_1}{X_2/n_2}$ where X_1 and X_2 are independent chi-square variables with n_1 and n_2 d.f. Derivation of the p.d.f, nature of probability curve, mean, variance, moments, mode. Distribution of $\frac{1}{F_{n_1, n_2}}$, use of tables for calculation of probabilities. Interrelationship between Chi-square, t and F -distributions.			
5	Tests based on Sampling Distributions	05	CO6, CO7
Tests based on chi-square distribution: Test for independence of two attributes, Test for goodness of fit, Test for variance for known and unknown mean. Tests based on t-distribution: Tests for population means, Paired t -test for one-sided and two-sided alternatives. Test based on F-distribution: Test for $H_0: \sigma_1^2 = \sigma_2^2$ against one-sided and two-sided alternatives when i) means are known and ii) means are unknown. Take $F = \frac{S_1^2}{S_2^2}$.			
Reference Books			
1. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta. 2. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002. 3. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan, Pune 411029 4. Medhi, J., Statistical Methods, Wiley Eastern Ltd., 4835/24, Ansari Road, Daryaganj, New Delhi – 110002. 5. Meyer, P. L., Introductory Probability and Statistical Applications, Oxford and IBH Publishing Co. New Delhi. 6. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.			

7. Mood, A. M., Graybill F. A. and Bose, F. A. (1974), Introduction to Theory of Statistics (Third Edition, Chapters II, IV, V, VI), McGraw - Hill Series G A 276
8. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
9. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
10. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East West Press.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

DS-292-MN-P : Lab Course on DS-291-MN-T
(Testing of Hypothesis and Sampling Distributions)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Basics of Distribution		
Objectives Practical Implementation of - Hypothesis - Performing the large sample test - The sampling distribution - The difference between small sample and large sample test - Hypothesis testing based on sampling distributions		
Course Outcomes: At the end of this course, students will be able to: CO1: To draw conclusions and determining probability of making errors in hypothesis tests using critical values. CO2: To conduct various tests of significance like averages, population proportions, independence of attributes, variance etc.		
List of Assignments		
Sr. No.	Assignment	
1	Test for proportions (one sample problem and two sample problem)	
2	Test for means (large sample test).	
3	Test for means (one sample problem) (small sample test).	
4	Test for means (two sample problem) (small sample test).	
5	Paired t-test.	
6	Test for independence of attributes	
7	χ^2 goodness of fit test.	
8	Test based on χ^2 distribution : $H_0: \sigma^2 = \sigma_0^2$ for μ unknown	
9	Tests for population variance (two sample problem only).	
Reference Books		
19. Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta.		

20. Gupta, S. C. and Kapoor, V. K. (2002), Fundamentals of Mathematical Statistics, (Eleventh Edition), Sultan Chand and Sons, 23, Daryaganj, New Delhi, 110002.
21. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. (1999), Common Statistical Tests, Satyajeet Prakashan, Pune 411029
22. Medhi, J., Statistical Methods, Wiley Eastern Ltd., 4835/24, Ansari Road, Daryaganj, New Delhi – 110002.
23. Meyer, P. L., Introductory Probability and Statistical Applications, Oxford and IBH Publishing Co. New Delhi.
24. Mood, A. M. and Graybill, F. A. and Boes D.C. (1974). Introduction to the Theory of Statistics, Ed. 3, McGraw Hill Book Company.
25. Mood, A. M., Graybill F. A. and Bose, F. A. (1974), Introduction to Theory of Statistics (Third Edition, Chapters II, IV, V, VI), McGraw - Hill Series G A 276
26. Neil, A. Weiss, (2016). Introductory Statistics, Tenth Edition, Pearson.
27. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
28. W. and Cochran W. G. (1989). Statistical Methods, Eighth Ed. East West Press.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

SEC-251-DS-T: Software Engineering

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks	
Prerequisites Fundamentals of programming language, Databases, ER Modeling			
Objectives 1. To get knowledge and understanding of software engineering discipline. 2. To learn analysis and design principles for software project development. 3. To become familiar with software development life cycle models and their applications. 4. To understand software testing strategies and techniques.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand software engineering discipline. CO2: Compare and choose a process model for a software project development. CO3: Identify requirements analyze, prepare models and Agile Models. CO4: Apply design concepts and metrics in software development. CO5: Implement testing strategies to ensure software quality.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Software Engineering Fundamentals	3	CO1
1.1 Definition of Software and the evolving role of software 1.3 Changing nature of software 1.4 Software myths. 1.5 Software Process 1.5.1 The Process Framework 1.5.2 Umbrella Activities 1.5.3 Process Adaptation			
2	Software Development Life Cycle (SDLC)	5	CO2
2.1 Generic Process Model 2.2 Prescriptive Process Models 2.2.1 The Waterfall Model 2.2.2 Incremental Process Models 2.2.3 Evolutionary Process Models 2.2.4 Concurrent Models 2.2.5 The Unified Process 2.6 Reverse Engineering			
3	Agile Development	6	CO3
3.1 Agility and Agile Process			

3.2 Agility Principles 3.3 The Politics of Agile Development 3.4 Human Factors 3.5 Extreme Programming (XP) 3.5.1 XP Values 3.5.2 XP Process 3.6 Adaptive Software Development (ASD) 3.7 Scrum 3.8 Dynamic System Development Model (DSDM) 3.9 Agile Unified Process (AUP)			
4	Software Requirement Analysis and Design Using UML	12	CO3,CO4
4.1 Requirement Elicitation 4.2 Software requirement specification (SRS) 4.3 Building the Analysis Model 4.4 Negotiating Requirements 4.5 Validating Requirements 4.6 Introduction to UML 4.7 Structural Modeling 4.7.1 Use case model 4.7.2 Class model 4.8 Behavioral Modeling 4.8.1 Sequence model 4.8.2 Activity model 4.8.3 Collaboration model 4.9 Architectural Modeling 4.9.1 Component model 4.9.2 Deployment model			
5	Software Testing	4	CO5
5.1 Introduction to Software Testing 5.2 Software Testing Life Cycle (STLC) 5.3 Software Testing Strategies & Techniques 5.3.1 Black box and White box Testing 5.4 Level of Testing 5.4.1 Unit Testing 5.4.2 Integration Testing 5.4.3 System Testing			
Reference Books			
1. Software Engineering : A Practitioner's Approach - Roger S. Pressman, McGraw hill(Eighth Edition) ISBN-13: 978-0-07-802212-8, ISBN-10: 0-07-802212-6 2. The Unified Modeling Language Reference Manual - James Rumbaugh, Ivar Jacobson, Grady Booch ISBN 0-201-30998-X 3. Effective Methods of Software Testing, William E Perry, 3rd Edition, Wiley Publishing Inc			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

Course Type: GE/OE. Course Code: OE-251-DS-T
Course Title: Ecommerce II
(To be offered to faculty other than Science and Technology)

Teaching Scheme 02 Hours /Week	No. of Credits 2	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks
Prerequisites • Knowledge of basic e-commerce concepts, business models and payment systems. • Understanding of digital marketing, financial systems, and internet technologies		
Course Objectives • To understand the technical and security aspects of e-commerce. • To explore data-driven decision-making and analytics in e-commerce. • To study supply chain and logistics management in e-commerce. • To gain insights into global e-commerce trends and challenges. • To learn about the integration of AI, Blockchain, and Cloud Computing in e-commerce.		
Course Outcomes On completion of the course, student will be able to– CO1: Implement secure e-commerce transactions and protect user data. CO2: Apply analytics tools to track and enhance e-commerce performance. CO3: Manage e-commerce logistics and understand global trends. CO4: Use emerging technologies such as AI, Blockchain, and Cloud for e-commerce applications.		
Course Contents		
Chapter 1	E-Commerce Logistics and Supply Chain Management	6 Hours
1.1. E-Commerce Supply Chain Overview 1.2. Inventory Management in E-Commerce 1.3. Role of Warehousing and Fulfillment Centers 1.4. Last-Mile Delivery Challenges 1.5. Reverse Logistics and Return Policies 1.6. Green and Sustainable Supply Chains		
Chapter 2	Data Analytics and Consumer Behavior in E-Commerce	7 Hours
2.1. Importance of Data Analytics in E-Commerce 2.2. Key Performance Indicators (KPIs) and Metrics		

2.3. Web Analytics (Google Analytics, Heatmaps, A/B Testing)		
2.4. Predictive Analytics & Customer Insights		
2.5. Recommendation Engines & Personalization		
2.6. Conversion Rate Optimization (CRO)		
2.7. Fraud Detection Using AI in E-Commerce		
Chapter 3	E-Commerce Security and Privacy	6 Hours
3.1. Importance of Security in E-Commerce		
3.2. Threats to E-Commerce (Phishing, Fraud, Cyber Attacks)		
3.3. Cryptography & Secure Transactions (SSL/TLS, Encryption)		
3.4. Digital Signatures & Certificates		
3.5. Firewalls & Intrusion Detection Systems		
Chapter 4	Advanced E-Commerce Technologies	5 Hours
4.1. Machine Learning for Product Recommendations		
4.2. Blockchain for Secure Transactions and Smart Contracts		
4.3. Cloud Computing and SaaS Platforms for E-Commerce		
4.4. The Role of 5G in E-Commerce Growth		
Chapter 5	Global Trends of E-Commerce	6 Hours
5.1. Cross-Border E-Commerce and Global Expansion		
5.2. Mobile Commerce (M-Commerce) Innovations		
5.3. Subscription-Based E-Commerce Models		
5.4. Social Commerce (Instagram, Facebook Shops, TikTok Commerce)		
5.5. Ethical and Sustainable E-Commerce Practices		
5.6. Future Challenges in E-Commerce		
Reference Books:		
1. E-Commerce Analytics: Analyze and Improve the Impact of Your Digital Strategy, Judah Phillips, Pearson		
2. Cloud Computing: Principles and Paradigms, Rajkumar Buyya, James Broberg, Andrzej Goscinski, Wiley		
3. Global E-Commerce: Theory and Case Studies, Jie Lin, Fei Gao, Springer		
4. Logistics and Supply Chain Management, Martin Christopher, Pearson		

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

Course Type: GE/OE Course Code: OE-252-DS-T

Course Title: Web Design II

(To be offered to faculty other than Science and Technology)

Teaching Scheme 02 Hrs/ week	No. of Credits 2	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks
---------------------------------	---------------------	--

Prerequisites

- Knowledge of other programming languages will enhance your abilities with HTML and web development.
- Knowledge and understanding of Internet.

Course Objectives

- To learn to design Style sheets.
- To know to link and publish Web pages.
- To prepare the learners with the fundamentals of HTML programming and scripting languages.
- Learners should know how to create HTML documents, applying styles using CSS.

Course Outcomes

On completion of the course, student will be able to–

- Learn and use the CSS to design Web Pages.
- Link and publish Web pages.
- Develop and create a webpage by planning, designing, and implementing everything themselves.
- Understand and develop a dynamic web pages using JavaScript (client side programming) .

Course Contents

Chapter 1	CSS	8
1.1. Introduction of CSS and its Syntax 1.2. Ways to Insert CSS and Background image handling 1.3. Background colour management using CSS 1.4. Text and Font management using CSS 1.5. Managing Hyperlinks and List using CSS 1.6. Designing Borders and Outline 1.7. Setting Page Margin using CSS		
Chapter 2	XML	8
2.1. XML Namespaces and Infoset and Document Type 2.2. Definitions (DTDs)		

2.3.XML Schemas and XML-Parser		
2.4.Data Modeling , Document and Object Model (DOM)		
2.5.Displaying XML with XSLT		
Chapter 3	Introduction to JavaScript	8
3.1. Concept of Script, Types of Scripts		
3.2. Introduction to JavaScript.		
3.3. Variables, identifier and Operator, Control structure.		
3.4. Examples on JavaScript Operators.		
3.5. Functions		
3.6. Event Handling in JavaScript with examples.		
Case Study		6
Case study 1: Creation of forms, small case study to create HTML pages using all the above learnt techniques.		
Case study 2: Redesigning the Website of a Small Business.		
Case study 3: Create a Styled Web Page for a Coffee Shop.		
Reference Books:		
1. Learning Web Design: A beginner's Guide To HTML, CSS, Javascript, and Web Graphics - Jennifer Robbins		
2. HTML5: The Missing Manual - Matthew MacDonald		
3. HTML and JavaScript – Ivan Bayross		
4. Mastering HTML, CSS &Javascript Web Publishing		
Reference Link:		
1. JavaScript Guide - JavaScript MDN		
2. JavaScript reference - JavaScript MDN		
3. XML.com		

Savitribai Phule Pune University B.Sc. Data Science (2024 Pattern) Sem-IV Course Type: GE/OE Course Code: OE-253-DS-T Course Title: Digital Marketing- II <i>(To be offered to faculty other than Science and Technology)</i>		
Teaching Scheme 02 Hours /Week	No. of Credits 2	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks
Prerequisites - Digital marketing requires creativity and problem-solving abilities. - Experience with social media platforms (Facebook, Instagram, Twitter, LinkedIn, etc.) is beneficial, as digital marketing		
Course Objectives - To understand Digital Marketing as the most powerful marketing tool. - To Learn to create digital marketing artworks. - To use social media sites like Facebook, Instagram, Twitter, LinkedIn, and others to raise sales, engage customers, and establish your brand.		
Course Outcomes On completion of the course, student will be able to– CO1: Students will be able to communicate marketing strategies and results effectively to stakeholders. CO2: Student will be able to assess and enhance digital marketing campaigns' return on Investment. CO3: Students will gain practical experience with industry-standard digital marketing tools. CO4: Student will be adept at using a variety of social media channels to create and interact with communities, raise awareness of a brand.		
Course Contents		
Chapter 1	Online Consumer Behaviour Analysis	8
1.1 Consumer Behavior 1.2 Segmentation and targeting online customers 1.3 Psychological Responses 1.4 Social Trends		
Chapter 2	Social Media Marketing	8
2.1. Social Media Sites 2.2. Influence of Social Media Marketing 2.3. Power of Social Media 2.4. Monetization through Social Media		
Chapter 3	Future of Digital Marketing	8
3.1. Use of Artificial Intelligence (AI) in Digital Marketing. 3.2. Common use of household gadgets for online marketing.		

3.3. Digital Marketing strategies.	
Case Study	6
Case Study 1 : Experiential Learning : Creating a website. Case Study 2 : Online Consumer Behavior Analysis for an E-Commerce Fashion Brand Case Study 3	
Reference Books:	
1. Digital Marketing : Nitin Kamat, Chinmay Kamat (Himalaya Publishing House) 2. "Digital Marketing for Dummies" by Ryan Deiss and Russ Henneberry 3. "Influence: The Psychology of Persuasion" by Robert B. Cialdini 4. "Social Media Marketing Workbook: How to Use Social Media for Business" by Jason McDonald	
Reference Links:	
1. https://www.socialmediaexaminer.com/ 2. https://www.marketingprofs.com/	

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-IV

Course Type: GE/OE Course Code: OE-254-DS-T

Course Title: AI for Everyone- II

(To be offered to faculty other than Science and Technology)

Teaching Scheme: 2hours / week	No. of Credits: 2	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks	
Course Objectives: - 1. Understand the basics of artificial intelligence and its subfields. 2. Explore real-world applications of AI across different industries. 3. Gain insights into the ethical, social, and economic implications of AI. 4. Develop an appreciation for the potential of AI to drive innovation and transformation.			
Course Outcomes: - On completion of the course, student will be able to– CO1: To understand different types of AI Models CO2: To understand content optimization using AI. CO3: To understand Animations and motions in AI CO4: To Understand uses of AI tools.			
Course Contents			
Unit 1	Advanced AI Fundamentals	6 hours	CO1
1.1 Deep Dive into AI, Machine Learning & Deep Learning 1.2 Types of AI Models: Generative AI, NLP, Computer Vision, Reinforcement Learning 1.3 Latest AI Trends: AGI, Large Language Models (LLMs), and multimodal AI 1.4 Exploring AI Frameworks & APIs: Open AI, Hugging Face, Google AI			
Unit 2	AI for Advanced Text & Content Creation	8 hours	CO2
2.1 AI for Long-form Writing & Reports 2.2 Automating Research & Citation Management 2.3 AI for SEO & Content Optimization 2.4 Using AI for Professional Emails & Business Writing			
Unit 3	Advanced AI for Image & Video Processing	8 hours	CO3
3.1 AI Image Generation Beyond Basics 3.2 Deepfake Technology & Ethical Concerns 3.3 AI Video Editing & Creation			

3.4 AI Animation & Motion Capture			
Unit 4	AI Tools	8 hours	CO4
4.1 Chat GPT (Open AI) 4.2 Google Gemini (Bard AI) 4.3 Canva, Beautiful. AI, Gamma, Slides AI 4.4 Rytr, Grammarly			
References: 1. Artificial Intelligence: A Modern Approach – Stuart Russell & Peter Norvig. 2. Practical AI for Business Leaders – Anand S. Rao 3. AI-Powered Automation Handbook – Will Kelly 4. AI for Content Creators: How to Use AI Tools for Writing and Marketing – Rob Lennon 5. Human Compatible: Artificial Intelligence and the Problem of Control – Stuart Russell 6. https://www.grammarly.com/ai/ai-writing-tools/article-writer			

Savitribai Phule Pune University

(Formerly University of Pune)

Bachelors Degree in Data Science

(Faculty of Science and Technology)



Syllabi for B.Sc. (Data Science)-Third Year Sem-V and VI

(For Colleges Affiliated to Savitribai Phule Pune University)

**Choice Based Credit System (CBCS) Syllabus
Under National Education Policy (NEP)**

To be implemented from Academic Year 2026-27

Savitribai Phule Pune University

Syllabus Structure as per NEP Guidelines

B.Sc. (Data Science) from 2026-27

TY (Level 5.5) SEMESTER V

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	C E	E E	Total
Major Core	DS-301-MJ-T	Data Visualization and Modelling	4	-	4	-	30	70	100
	DS-302-MJ-T	R Programming	2	-	2	-	15	35	50
	DS-303-MJ-T	Foundations of Artificial Intelligence	2	-	2	-	15	35	50
	DS-304-MJ-P	Lab Course on DS-301-MJT (Data Visualization and Modelling)	-	2	-	4	15	35	50
	DS-305-MJ-P	Lab Course on DS-302-MJT (R Programming)	-	2	-	4	15	35	50
Major Elective	DS-310-MJ-T	Business Analytics	2	-	2	-	15	35	50
	DS-311-MJ-P	Lab Course on DS-310-MJT (Business Analytics)	-	2	-	4	15	35	50
	OR					-			
	DS-312-MJ-T	Social Media Analytics	2	-	2	-	15	35	50
	DS-313-MJ-P	Lab Course on DS-312-MJT (Social Media Analytics)	-	2	-	4	15	35	50
VSC	DS-321-VSC-P	Business Intelligence and Dashboarding Tools		2	-	4	15	35	50
FP/CEP	DS-331-FP	Project on Data Analytics & Visualization	-	2	-	4	15	35	50
Minor	DS-341-MN-T	Categorical and Multivariate Data Analysis	2	-	2	-	15	35	50
Total			12	10	12	20			550

Savitribai Phule Pune University
Syllabus Structure as per NEP Guidelines
B.Sc. (Data Science) from 2026-27
TY (Level 5.5) SEMESTER VI

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			T H	P R	TH	PR	CE	E E	Total
Major Core	DS-351-MJ-T	NoSQL databases	4	-	4	-	30	70	100
	DS-352-MJ-T	Artificial Intelligence in Data Science	2	-	2	-	15	35	50
	DS-353-MJ-T	Data Privacy and Security	2	-	2	-	15	35	50
	DS-354-MJ-P	Lab Course on DS-351-MJ-T (NoSQL databases)	-	2	-	4	15	35	50
	DS-355-MJ-P	Lab Course on DS-352-MJ-T (Artificial Intelligence in Data Science)	-	2	-	4	15	35	50
Major Elective	DS-360-MJ-T	HR Analytics	2	-	2	-	15	35	50
	DS-361-MJ-P	Lab Course on DS-360-MJ-T (HR Analytics)	-	2	-	4	15	35	50
	OR								
	DS-362-MJ-T	Financial Analytics	2	-	2	-	15	35	50
	DS-363-MJ-P	Lab Course on DS-362-MJ-T (Financial Analytics)	-	2	-	4	15	35	50
VSC	DS-371-VSC-P	Lab Course on Statistical Analysis using PSPP	-	2	-	4	15	35	50
OJT	DS-381-OJT	On Job Training	-	4	-	8	30	70	100
Total			10	12	10	24			550

Detail Syllabus

B.Sc. (Data Science)

Semester-V

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-301-MJ-T : Data Visualization and Modelling

No. of Credits: 4	Teaching Scheme Theory: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 30 Marks End Semester : 70 Marks	
Prerequisites - Fundamental understanding of programming concepts using Python - Basic knowledge of statistics			
Objectives - To introduce fundamental concepts of data visualization and data handling - To develop skills in loading, cleaning, and preprocessing data using Python libraries - To provide knowledge of exploratory data analysis techniques for understanding data - To build a foundation in predictive modelling and machine learning algorithms - To enable students to apply data science techniques to real-world datasets			
Course Outcomes On Completion of this course, student will be able to – CO1: Explain the need, importance, and applications of data visualization techniques CO2: Load, clean, preprocess, and visualize data using Python-based data science libraries CO3: Perform exploratory data analysis and interpret data patterns and correlations CO4: Build, evaluate, and interpret regression and classification models for prediction CO5: Apply unsupervised learning and association rule mining techniques to analyze datasets.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Data Visualization and Data Handling	11	CO1, CO2
1.1 Introduction to Data Visualization Need for Data Visualization, Importance of Data Visualization, Applications of Data, Visualization, Challenges of Data Visualization.			
1.2 Data Loading and Preprocessing Loading datasets from CSV and Excel files, Data loading using Pandas, Data cleaning techniques, Data preprocessing methods.			
1.3 Introduction to Data Science Libraries NumPy, Pandas, Matplotlib, Seaborn, Plotly, SciPy, Statsmodels, Scikit –learn, XGBoost			
1.4 Basic Charts and Graphs Bar Chart ,Line Chart, Histogram, Pie Chart, Scatter Plot, Spike Plot			

2	Advanced Data Visualization and Exploratory Data Analysis	10	CO1, CO2, CO3
2.1 Advanced Visualization Techniques Box Plot / Box Whisker Plot, Violin Plot, Joint Plot, Pair Plot, Heatmap, Word Cloud, Bubble Chart, Area Chart, Subplots 2.2 Exploratory Data Analysis (EDA) Concept and importance of EDA, Identifying data patterns and trends, Correlation analysis using visualization, Visual techniques for exploratory analysis.			
3	Predictive Modelling and Machine Learning	20	CO3, CO4
1.1 Introduction to Machine Learning Definition of Machine Learning, Types of Machine Learning (Supervised Learning, Unsupervised Learning, Reinforcement Learning), Applications of Machine Learning. 1.2 Introduction to Statistical Modelling Model building process, Training and testing data concepts. 1.3 Simple Linear Regression Model formulation, Assumptions of Linear Regression, Interpretation of regression coefficients. 1.4 Multiple Linear Regression Model formulation, Interpretation of regression coefficients, Multicollinearity 1.5 Model Evaluation for Regression Models R-square, Adjusted R-square, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Residual analysis, Interpretation and comparison of regression models. 1.6 Regularization Techniques Ridge Regression, Lasso Regression. 1.7 Practical Applications of Regression Models			
4	Classification and Ensemble Learning	10	CO4
1.1 Classification Algorithms Logistic Regression, K -Nearest Neighbors (KNN), Decision Tree Classifier, Naïve Bayes Classifier. 4.2 Ensemble Learning Techniques Concept of Bagging and Boosting, Random Forest, Gradient Boosting, Ada Boost. 4.3 Model Evaluation Techniques for Classification Confusion Matrix, Accuracy, Precision, Recall, F1 Score, ROC Curve and AUC, Interpretation and comparison of classification models. 4.4 Model Performance Issues Bias–Variance Trade-off, Overfitting, Underfitting.			
5	Unsupervised Learning and Association	9	CO5

	Analysis		
5.1 Introduction to Unsupervised Learning 5.2 Clustering Techniques Concept of clustering, Distance and similarity measures, K-Means Clustering, Applications of clustering, Hierarchical Clustering, DBSCAN. 5.3 Dimensionality Reduction Principal Component Analysis (PCA). 5.4 Association Rule Mining Concept of Association Rules, Apriori Algorithm, Market Basket Analysis.			
Reference Books			
1. Book1 : Han, Kamber & Pei, <i>Data Mining: Concepts and Techniques</i> 2. Book2 : Learning Python, 4th Edition by Mark Lutz 3. Book3: Data Science Essentials in Python: Collect, Organize, Explore, Predict, Value. Dmitry Zinoriev, The Pragmatic Programmers LLC, 2016 4. Book4: James et al., <i>An Introduction to Statistical Learning</i> 5. Book5: Géron, <i>Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow</i>			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-302-MJ-T: R Programming

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester: 35 Marks	
Prerequisites ● Basic knowledge of Statistics ● Basic knowledge of Computer Software			
Objectives ● Understand R basics, set up R Studio, and customize the environment ● Develop proficiency in using R data structures: vectors, matrices, lists, and data frames ● Use R – software to find the summary statistics. ● Compute probabilities of the distributions using R – software. ● Represent statistical data by diagrammatically and graphically using R-software			
Course Outcomes On Completion of this course, student will be able to - CO1: Apply basic R programming commands and data management techniques using R/RStudio. CO2:Create and interpret diagrammatic and graphical representations for data visualization in R CO3: Analyze measures of central tendency, dispersion, and perform matrix computations in R CO4: Summarize datasets using frequency distributions and advanced statistical summaries. CO5: Perform correlation and regression analysis to study relationships between variables. CO6: Simulate and analyze discrete and continuous probability distributions and conduct hypothesis testing using R.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to R Programming	5	CO1
1.1 Introduction to R, History, Features and Applications 1.2 Installing R and RStudio, Running Simple Programs 1.3 Data Types, Variables, Operators and Type Conversion 1.4 Meaning of package, base (standard package) package, to install a package, to load a package, to delete a package, learning about a package, getting help. 1.5 Operators & expressions in R – Arithmetic, logical, relational operators & expressions, precedence rule.			
2	Working with Vectors	5	CO1, CO3
2.1 Vectors: creating a vector, modifying a vector, deleting a vector. 2.2 Working with vectors: the functions to be discussed-c(), rep(), rev(), sort(), diff(), max(), min(), colon operator(:), seq(), scan(), cut(), cat(), table(), which(), unique(), is.vector(), as.vector(), print() 2.3 Mathematical functions: abs(), sqrt(), ceiling(), floor(), trunc(), log(), log10(), exp(), sin(),			

cos(), tan(), atan(), sign(), round(), range(), length(), prod(), sum(),. 2.4 Character functions: nchar(), substr(), grep(), sub(), paste(), strsplit(), toupper(), tolower(). Meaning of data frame, creation of data frame, modifying a data frame, deleting data frame, extracting elements from a data frame, use of \$ sign. Functions to be discussed: subset(), transform(), attach(), detach(), with(), data.entry(), edit(), is.data.frame(), as.data.frame()			
3	Matrix Operations	4	CO1, CO3
3.1 Creating a matrix, scalar multiplication of matrix, matrix addition, subtraction and multiplication, transpose of a matrix, inverse of a matrix, solving system of linear equations, finding row and column sums and means. 3.2 The functions to be discussed: matrix(), colMeans(), colSums(), rowMeans(), rowSums(), solve(), t(), diag(), is.matrix(), as.matrix()			
4	Data Handling and Visualization	5	CO2, CO4
4.1 Reading and Writing Data Files, Data Preprocessing and Cleaning 4.2 Packages used for Data Visualization 4.3 Diagrams - Simple bar diagram, Subdivided bar diagram, multiple bar diagram, Pie diagram, Stem and leaf diagram 4.4 Graphs - Boxplot for one and more than one variables, spike plot, histogram for both equal and unequal class intervals, frequency polygon, ogive curves			
5	Statistical Analysis and Predictive Modelling	6	CO3, CO5
5.1 Measures of Central tendency - mean, mode, median, Partition values - quartiles, deciles, percentiles, geometric mean and harmonic mean. 5.2 Dispersion: variance, standard deviation, coefficient of variation, mean deviation. 5.3 Summary Statistics using fivenum() and summary() for quick data summaries. 5.4 Skewness: Bowley's coefficient and Karl Pearson's coefficient of skewness. 5.5 Moments: Computations of raw and central moments, measures of skewness and kurtosis. 5.6 Correlation and Regression Analysis			
6	Probability Distributions and Hypothesis Testing	5	CO6
6.1 Discrete Probability Distributions a) Simulating data from Bernoulli, Binomial, and Poisson distributions. b) Calculating probabilities and percentiles using dbinom, pbinom, dpois, etc. 6.2 Continuous Probability Distributions a) Visualization of normal, Chi square, t-distribution, f-distribution using R b) Applications of standard normal distribution (Z-scores) 6.3 Hypothesis Testing : t test, Chi square test, f Test for testing variances, One Way ANOVA			
Reference Books			
1. Book1 : Richard Cotton, "Learning R", SPD O'Reilly Publications 2. Book2 : Gupta, S. C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi.			
Important Links			
https://www.w3schools.com/r/r_math.asp https://www.stats4stem.org/pdfjs/web/viewer.php?file=using_r_for_introductory_statistics.pdf			

Savitribai Phule Pune University

B.Sc. Data Science (2024 Pattern)

Sem-V

DS-303-MJ-T: Foundations of Artificial Intelligence

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester: 35 Marks	
Prerequisites Foundational knowledge of programming, data structures, discrete mathematics, probability and statistics, and data science.			
Objectives - To introduce students to the fundamental concepts, scope, and evolution of Artificial Intelligence. - To develop understanding of problem solving, knowledge representation, reasoning, and learning in AI. - To bridge Data Science foundations with Artificial Intelligence concepts. - To create awareness of ethical, social, and practical implications of AI.			
Course Outcomes On Completion of this course, student will be able to – CO1: Understand the basic concepts, goals, and applications of Artificial Intelligence. CO2: Formulate problems and analyze AI-based problem-solving techniques. CO3: Describe knowledge representation and logical reasoning mechanisms. CO4: Understand learning paradigms and the role of expert systems in Artificial Intelligence. CO5: Analyze ethical issues and future trends in Artificial Intelligence.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Artificial Intelligence	5	CO1
- What is Artificial Intelligence? - Scope and importance of Artificial Intelligence - Brief History and Evolution of AI - Types of Artificial Intelligence - AI Problems and Basic Concepts - Application of Artificial Intelligence - Game Playing - Natural Language Processing - Computer Vision - Speech Processing			
2	Problem Solving Techniques in Artificial Intelligence	5	CO2
- Problem definition and problem space - Characteristics of AI Problem - State space representation - Production systems - Problem reduction			

Issues in the Design of Search Programs			
3	Knowledge Representation	6	CO3
- Role of knowledge in Artificial Intelligence - Knowledge representation and mapping - Approaches to knowledge representation - Issues in knowledge representation - Rule-based systems - Procedural versus Declarative knowledge			
4	Logic-Based Representation and Reasoning under Uncertainty	8	CO3
- Propositional Logic - Predicate Logic: - Representing facts and relationships - Modus Ponens - Inference techniques: - Resolution - Natural Deduction - Forward chaining and Backward chaining - Structured knowledge representation: - Semantic Networks - Frames - Reasoning under uncertainty: - Probabilistic reasoning (conceptual overview) - Certainty factors - Comparison of logical and probabilistic reasoning			
5	Learning and Expert Systems	4	CO4
Learning in Artificial Intelligence - Concept of learning in AI - Types of learning: - Rote learning - Learning by taking advice - Learning in problem solving - Learning by induction - Explanation-based learning Expert Systems - Need and justification of expert systems - Components of an expert system - Limitations of expert systems			
6	Ethical Issues in AI	2	CO5

- Ethical challenges in AI
- Bias and fairness in AI systems
- Transparency and Explainability
- Privacy and data protection
- Social impact of Artificial Intelligence
- Responsible and trustworthy AI

Reference Books

Book1 : Artificial Intelligence by Elaine Rich & Kevin Knight

Book2 : Artificial Intelligence: A Modern Approach by Stuart Russell & Peter Norvig

Book 3: Artificial Intelligence: Foundations of Computational Agents by David Poole & Alan Mackworth

Book 4: Ethics of Artificial Intelligence and Robotics by Edited by Patrick Lin, Keith Abney & Ryan Jenkins

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-304-MJ-P: Lab Course on DS-301-MJ-T (Data Visualization and Modelling)

No. of Credits: 2	Teaching Scheme Practical : 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites • Basic knowledge of Statistics • Basic knowledge of Python			
Objectives • To introduce essential libraries for data analysis and visualization. • To develop practical skills in data loading, data cleaning, preprocessing, and exploratory data analysis using Python. • To understand and apply various data visualization techniques for effective graphical representation of data. • To introduce basic data modelling techniques such as regression, classification, and clustering using Python. • To enable students to build, evaluate, and interpret simple data models for real-world datasets.			
Course Outcomes On Completion of this course, student will be able to - CO1: Implement Python programs for data loading, preprocessing, visualization and exploratory data analysis. CO2: Create effective data visualizations using Python libraries for graphical representation of data. CO3: Apply statistical techniques and basic data modelling methods to real-world datasets. CO4: Build, evaluate, and interpret simple data models for analysis and prediction. CO5: Integrate visualization and modelling results to support data-driven decision making.			
Operating Environment: • Windows (Windows 10/11) • Linux (Ubuntu, Fedora, Debian, etc.) • macOS (Latest versions) Python Programs can be executed using: • VS Code • Jupyter Notebook			
Sr No.	Name of Practical	Hours	CO Targeted
1	Overview of Python Programming Environment and Data Visualization Libraries	12	CO1
Overview of Python environment, installation and use of Jupyter Notebook.			
Introduction to data visualization libraries: NumPy, Pandas, Matplotlib, ggplot2 and Seaborn			

2	Data Cleaning, and Preprocessing using Pandas	4	CO1
Loading datasets from CSV/Excel files. Handling missing values, data type conversion, filtering, sorting, and data preprocessing			
3	Basic and Advanced Data Visualization	8	CO2
Creating basic plots such as line plot, bar chart, histogram, and scatter plot using data visualization libraries such as Matplotlib. Advanced visualizations including boxplot, joint plot, violin plot, pair plot, word cloud, bubble chart, area chart, pie chart, and heatmap.			
4	Exploratory Data Analysis (EDA) and Correlation Analysis with Visualization	12	CO1, CO2
Performing Exploratory Data Analysis to understand data patterns and trends. Correlation analysis using correlation matrix and visualization using heatmaps.			
5	Simple Linear Regression	8	CO3, CO4
Implementation of Simple Linear Regression model. Interpretation of regression coefficients, analysis of residual plots and visualization of regression line.			
6	Multiple Linear Regression and Model Evaluation	8	CO4, CO5
Building Multiple Linear Regression model using Python. Evaluation of model performance using metrics such as R^2 , MAE, MSE, and RMSE.			
7	Logistic Regression and K-Means Clustering	8	CO4, CO5
Building Logistic Regression model using Python for classification. Application of K-Means Clustering technique for grouping unlabeled data. Selection of appropriate number of clusters and visualization of clusters using Python.			
Reference Books			
1. Book1: Python Programming: An introduction to computer science, John Zelle, 3 rd Edition. 2. Book2: Statistics and Machine Learning in Python by Edouard Duchesnay, Tommy Löfstedt, Feki Younes 3. Book3: Practical Statistics for Data Scientists, Second Edition by Peter Bruce, Andrew Bruce, and Peter Gedeck, O'Reilly, 2020			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-305-MJ-P : Lab Course on DS-302-MJ-T (R Programming)

No. of Credits: 2	Teaching Scheme Practical : 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Basic knowledge of Statistics • Basic knowledge of computers and data handling		
Objectives • To familiarize students with the R programming environment and RStudio • To develop practical skills in handling data using R data structures • To perform statistical computations using R software • To visualize data using diagrams and graphical techniques in R • To apply R programming for statistical analysis, probability distributions, and hypothesis testing		
Course Outcomes On Completion of this course, student will be able to - CO1: Use basic R commands, functions, and packages for data handling and computation CO2: Create and interpret diagrammatic and graphical representations of data using R CO3: Compute and analyze measures of central tendency, dispersion, and matrix operations using R CO4: Perform summary statistics, skewness, kurtosis, and correlation analysis using R CO5: Apply R programming to probability distributions, hypothesis testing, regression, and ANOVA		
List of Practical Assignments:		
Practical No.	Title of Practical	Hours
1	Introduction to R and RStudio: Writing and executing basic R programs, understanding R environment and packages	4
2	Working with R Data Types and Operators: Variables, type conversion, arithmetic, logical and relational operations	4
3	Creating and Manipulating Vectors in R using built-in functions	4
4	Mathematical and Character Functions in R	4
5	Creating and Managing Data Frames: Data extraction, modification, and summary using R	8

6	Matrix Operations in R: Matrix creation, arithmetic operations, transpose, inverse, and solving linear equations	8
7	Diagrammatic Representation of Data using R (Bar diagrams, Pie charts, Stem-and-Leaf)	4
8	Graphical Representation of Data using R (Histogram, Boxplot, Frequency Polygon, Ogive Curves)	4
9	Measures of Central Tendency and Dispersion using R	4
10	Summary Statistics, Skewness, and Kurtosis Computation using R	4
11	Correlation and Regression Analysis using R	4
12	Probability Distributions and Hypothesis Testing using R (Binomial, Normal, t-test, Chi-square test, ANOVA)	8
Reference Books		
2. Book1 : Introduction to Probability and Statistics Using R by G. Jay Kerns 3. Book2 : An R Companion to Linear Statistical Models by Christopher Hay-Jahans 4. Book3 : Introductory Statistics with R, Second Edition by Peter Dalgaard, Springer		

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-310-MJ-T: Business Analytics

No. of Credits: 02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites • Basic Mathematical operations • Fundamental understanding of Data Science workflows • Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage			
Objectives • To acquaint students with the process of translating business problems into analytical queries • To introduce data mining processes (CRISP/SEMMA) and data preparation techniques • To identify the nature of business data across different organizational domains • To apply advanced statistical tools for trend analysis and prediction.			
Course Outcomes On Completion of this course, student will be able to - CO1: Identify types of analytics (Descriptive to Prescriptive) and translate vague business problems into data queries. CO2: Prepare data for analysis using cleaning tools while following ethical and privacy laws (GDPR/DPDP). CO3: Perform business benchmarking and value-based profiling using Pareto (80/20) and ABC analysis. CO4: Measure relationships between business drivers using Correlation and Regression models. CO5: Predict future trends using Time Series forecasting and evaluate accuracy with error metrics (MAE/MSE).			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	INTRODUCTION	05	CO 1
1.1 Definition of analytics, Evolution of analytics, Concept of analytics, Types of Analytics.			
1.2 The Growing Role of Business Analytics, the distinction between Business Analysis (Process), BI (Reporting), and Data Science (Algorithms).			
1.3 Applications of Business Analytics: Marketing and Sales, Human Resource, Healthcare, Product Design, Service Design, finance, operation, Customer Service and Support.			
1.2 Categorization of Analytical methods and models: - Descriptive (What happened?), Diagnostic (Why?), Predictive (What will happen?), and Prescriptive (How to make it happen?).			
1.3 The process of translating a vague business problem into a structured data query.			
1.4 The analytical decision-making process vs. intuitive decision-making vs. evidence-based decision making.			

2	DATA BASES, DATA WAREHOUSING	05	CO 2
2.1 Importance of data in business analytics, Differences between data, information and knowledge, Quality of data. 2.2 Types of Data Sources- Structured Vs. Semi structured Vs. Unstructured data, 2.3 Data Warehouse Vs. Databases, Relational Database Vs. Non-Relational Database, RDBMS Data structures, and Columnar Data structures. 2.4 5Vs (Volume, Velocity, Variety, Veracity, Value) of Big Data, Big Data Collection and Ethics. 2.5 Ethical collection methods, and data privacy laws (GDPR/Digital Personal Data Protection basics). 2.6 Data preparation for knowledge discovery: Data understanding and data cleaning tools, Data transformation, Data Discretization, Data Visualization. 2.7 Data Mining Process: CRISP and SEEMA, Supervised and unsupervised learning techniques.			
3	DESCRIPTIVE AND DIAGNOSTIC BUSINESS ANALYTICS (focus must on "how to interpret and apply" using Python.)	10	CO 3
3.1 Comparative Analysis: Benchmarking performance across categories (Region, Product, and Department) or fixed time periods. Statistical Tools: Mean, Median, Variance, Standard Deviation, and Coefficient of Variation (for stability analysis). Visual Tools: Side-by-Side Bar Charts and Box-and-Whisker plots for distribution comparison. 3.2 Value-Based Analysis (Impact Analysis): Identifying high-impact business segments and resource optimization. Statistical Tools: Pareto Analysis (80/20 Rule), ABC Analysis (Inventory/Customer classification), and Cumulative Frequency Distribution. 3.3 Trend & Growth Analysis (Historical): Identifying secular, seasonal, and cyclical shifts; distinguishing "Noise" from "Signal." Statistical Tools: Moving Averages (Simple & Exponential), Percentage Change, YoY Growth, and CAGR (Compound Annual Growth Rate).			
4	PREDICTIVE AND RELATIONSHIP ANALYTICS (focus must on "how to interpret and apply" using Python.)	10	CO 4, CO 5
4.1 Correlation & Relationship Analysis: Measuring the strength and direction between business drivers (e.g., Marketing Spend vs. Revenue). Statistical Tools: Pearson's Correlation (Linear), Spearman's Rank (Non-linear), and Scatter Plot Matrices. 4.2 Regression Analysis (Causal Modeling): Predicting continuous business outcomes based on one or more predictors. Statistical Tools: Simple and Multiple Linear Regression. Diagnostics: R^2 (Goodness of fit), P-values (Significance), and Residual Analysis for error checking. 4.3 Time Series Forecasting (Predictive): Analyzing interval-based data to project future business metrics. Components: Trend, Seasonality, Cyclicity, and Irregularity. Models: Decomposition models, Autoregression (AR), and Holt-Winters Exponential Smoothing. Accuracy Metrics: Mean Absolute Error (MAE) and Mean Squared Error (MSE).			

Reference Books

1. Business Analytics: The Science of Data-Driven Decision Making by U. Dinesh Kumar (Wiley India). **(Unit Coverage: Excellent for Unit 1 (Foundations)), and Unit 4 (Regression & Time Series))**
2. Business Analytics: Data Analysis & Decision Making by Albright & Winston (Cengage Learning). **(Unit Coverage: The good standard for Unit 4 (Methodologies))**
3. Business Intelligence, Analytics, and Data Science: A Managerial Perspective by Ramesh Sharda, Dursun Delen, & Efraim Turban (Pearson). **(Unit Coverage: Best for Unit 1 and Unit 2).**
4. Data Science for Business by Foster Provost & Tom Fawcett. **(Unit Coverage: Unit 2).**
5. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends by Michael Minelli. **(Unit Coverage: Unit 2).**
6. Python for Data Analysis by Wes McKinney (O'Reilly). **(Unit Coverage : Unit 3)**
7. Business Analytics by Sahil Raj (Pearson Education). **(Unit Coverage : Unit 3)**
8. Forecasting: Principles and Practice by Hyndman & Athanasopoulos (O'Reilly/Online). **(Unit Coverage : Unit 4)**
9. Introduction to Linear Regression Analysis by Montgomery & Peck. **(Unit Coverage : Unit 4)**
10. Hands-on Machine Learning with Scikit-Learn by Aurélien Géron. **(Unit Coverage : Unit 4)**

OTHERS

1. **Competing on Analytics: The New Science of Winning** by *Thomas H. Davenport*.
2. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling** by *Ralph Kimball*.
3. **Big Data: A Revolution That Will Transform How We Live, Work, and Think** by *Viktor Mayer-Schönberger*.
4. **Data Mining for Business Analytics: Concepts, Techniques, and Applications** by *Galit Shmueli et al. (Wiley)*.
5. **Introduction to Data Mining** by *Pang-Ning Tan, Michael Steinbach, & Vipin Kumar*.
6. **Forecasting: Principles and Practice** by *Rob J. Hyndman & George Athanasopoulos*.
7. **Naked Statistics: Stripping the Dread from the Data** by *Charles Wheelan*.

VIDEO REFERENCES LINK

1. Analytics Overview & Types: [Business Analytics Definition, Types, and Applications](#)
2. BI vs. BA vs. Data Science: [Business Intelligence vs. Business Analytics](#)
3. Problem Framing: [Translating Business Problems into Analytical Questions](#)
4. Data Architecture: [Database vs. Data Warehouse vs. Data Lake Explained](#)
5. Big Data Fundamentals: [The 5 Vs of Big Data](#)
6. Mining Processes: [CRISP-DM & SEMMA Methodologies](#)
7. Supervised vs. Unsupervised Learning: [Machine Learning Categories Explained](#)
8. Forecasting: [Time Series Analysis and Forecasting for Beginners](#)
9. Full Course Reference: [Business Analyst Full Course 2026](#)
10. [Business Analyst Full Course](#)

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-311-MJ-P : Lab Course on DS-310-MJ-T (Business Analytics)

No. of Credits: 02	Teaching Scheme Practical: 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Basic Mathematical operations • Fundamental understanding of Data Science workflows • Knowledge of problem solving tools like algorithms, flowcharts and pseudo codes will be an added advantage		
Objectives • To acquaint students with the process of translating business problems into analytical queries. • To introduce data mining processes (CRISP/SEMMA) and data preparation techniques. • To identify the nature of business data across different organizational domains. • To apply advanced statistical and machine learning tools for trend analysis and prediction		
Course Outcomes On Completion of this course, student will be able to - CO1: Execute classification, clustering, and association rule mining to solve business use cases. CO2: Perform time series forecasting and trend analysis to predict future business outcomes. CO3: Apply comparative and value-based analysis to optimize business performance. CO4: Evaluate model performance using professional metrics like Confusion Matrix and Accuracy.		
Sr. No.	List of Practical Assignments	Teaching Hours
1	Write a program to identify missing values, duplicate records, and outliers in a business dataset using pandas and seaborn .	04
2	Create a script to convert continuous variables (e.g., Age) into categorical bins (e.g., Age Groups) and apply One-Hot Encoding to unstructured text categories.	04
3	Use Python to generate histograms and density plots to analyze the Volume and Variety of a dataset.	04
4	Write a program to compare sales across different regions using Groupby operations and visualize the consistency using Side-by-Side Bar Charts and Box-and-Whisker plots .	04
5	Implement a script to calculate cumulative contribution of products to total revenue and identify the top 20% "Vital Few." By using Pareto (80/20) analysis .	04
6	ABC Inventory Classification: Create a program to categorize inventory/customers based on usage value into A, B, and C classes.	04

7	Trend & Growth Calculator: Write a script to calculate Year-over-Year (YoY) growth and Compound Annual Growth Rate (CAGR) for a 5-year financial dataset.	04
8	Apply Simple and Exponential Moving Averages to smooth out daily fluctuations in stock or sales data	04
9	Correlation Matrix & Scatter Plot Matrix: Write a program using <code>seaborn.pairplot()</code> to measure the strength of relationships between multiple business drivers like "Marketing Spend," "Social Media Reach," and "Sales."	04
10	Simple & Multiple Linear Regression: Implement a regression model using <code>statsmodels</code> to predict sales and interpret the R-squared and P-values to check model significance.	04
11	Time Series Decomposition: Create a script to decompose a business time-series into its Trend, Seasonality, and Residual components.	04
12	Holt-Winters Forecasting: Use the <code>ExponentialSmoothing</code> library to forecast future demand and calculate accuracy using Mean Absolute Error (MAE) .	04

Reference Books

1. "Business Analytics" by Sahil Raj (Pearson Education).
2. "Storytelling with Data: A Data Visualization Guide for Business Professionals" by Cole Nussbaumer Knaflic.
3. "Data Mining for Business Analytics: Concepts, Techniques, and Applications in R/Python" by Galit Shmueli, Peter C. Bruce, et al. (Wiley).
4. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
5. "Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.
6. "Forecasting: Principles and Practice" by Rob J. Hyndman and George Athanasopoulos.
7. "Business Analytics: Data Analysis & Decision Making" by S. Christian Albright and Wayne L. Winston.
8. Forecasting: Principles and Practice by Hyndman & Athanasopoulos.
9. Hands-On Machine Learning (Scikit-Learn) by Aurélien Géron.
10. An Introduction to Statistical Learning by James, Witten, Hastie, Tibshirani.

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-312-MJ-T: Social Media Analytics

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites ● Graph Theory, Data Mining, Python programming			
Objectives ● Familiarize the learners with the concept of social media. ● Familiarize the learners with the concept of social media analytics and understand its significance. ● Familiarize the learners with different tools of social media analytics. ● Familiarize the learner with different visualization techniques for Social media analytics. ● Examine the ethical and legal implications of leveraging social media data.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand the concept of Social media CO2: Understand the concept of social media Analytics and its significance. CO3: Learners will be able to use different Social media analytics tools effectively and efficiently. CO4: Learners will be able to use different effective Visualization techniques to represent social media analytics. CO5: Acquire the fundamental perspectives and hands-on skills needed to work with social media data			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Social Media Analytics: An Overview	6	CO1, CO2
1.1 Core Characteristics of Social Media, Types of Social Media. 1.2 Need for Social Media Analytics (SMA), SMA in small & large organizations. 1.3 Social Media vs. Traditional Business Analytics. 1.4 Seven Layers of Social Media Analytics. 1.5 Types of Social Media Analytics, 1.6 Social Media Analytics Cycle 1.6 Challenges to Social Media Analytics 1.7 Social Media Analytics Tools			

2	Social Network Structure, Measures & Visualization	8	CO3, CO4, CO5
2.1 Basics of Social Network Structure - Nodes, Edges & Tie			
2.2 Describing the Networks Measures - Degree Distribution, Density, Connectivity, Centralization			
2.3 Centrality: Degree Centrality, Eigenvector Centrality, Katz Centrality, PageRank, Betweenness, Centrality, Closeness Centrality, Group Centrality			
2.4 Network Visualization - Graph Layout, Visualizing Network features, Scale Issues.			
2.5 Social Media Network Analytics - Common Network Terms, Common Social Media Network Types			
2.6 Types of Networks, Common Network Terminologies, Network Analytics Tools.			
3	Social Media Text, Action and Hyperlink Analytics	10	CO3,CO5
3.1 Social Media Text Analytics- Types of Social Media Text, Purpose of Text Analytics, Steps in Text Analytics, Social Media Text Analysis Tools			
3.2 Social Media Action Analytics -What Is Actions Analytics? Common Social Media Actions, Actions Analytics Tools			
3.3 Social Media Hyperlink Analytics-Types of Hyperlinks, Types of Hyperlink Analytics, Hyperlink Analytics Tool			
4	Social Media Location & Search Engine Analytics	6	CO4, CO5
4.1 Location Analytics - Sources of Location Data, Categories of Location Analytics, Location Analytics and Privacy Concerns, Location Analytics Tools			
4.2 Search Engine Analytics - Types of Search Engines, Search Engine Analytics, Search Engine Analytics Tools			
Reference Books			
1. Gohar F. Khan. Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine and Location Data. ISBN: 1507823207, 2015.			
2. Analyzing the Social Web 1st Edition by Jennifer Golbeck			

Savitribai Phule Pune University

B.Sc. Data Science (2024 Pattern)

Sem-V

DS-313-MJ-P: Lab Course on DS-312-MJ-T (Social Media Analytics)

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites Types of Graphs, Data Mining, Data Analytics			
Objectives - To understand the fundamental concepts of social media networks. - To learn various social media analytics tools and evaluation matrices. - To collect and store social media data. - To analyze and visualize social media data - To design and develop social media analytics models.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand characteristics and types of social media networks. CO2: Use social media analytics tools for business. CO3: Collect, monitor, store and track social media data. CO4: Analyze and visualize social media data from multiple platforms. CO5: Design and develop content and structure based social media analytics models.			
Unit No.	Practical Assignment	Teaching Hours	CO Targeted
1	Study various - i) Social Media platforms (Facebook, X, YouTube etc) ii) Social Media analytics tools (Facebook insights, google analytics etc) iii) Applications of Social media analytics for business. e.g. Google Analytics: https://marketingplatform.google.com/about/analytics/	4	CO1
2	Select two social media platforms (e.g., Facebook & Instagram OR Twitter/X & LinkedIn). Study their network characteristics , types of users, and interaction patterns. - Identify network type (ego network, follower network, community network) - Identify nodes, edges, and ties - Compare user engagement mechanisms	4	CO2, CO4
3	Data Cleaning and Storage- Preprocess, filter and store social media data for business (Using Python, MongoDB, etc)	4	CO3

4	Analyze engagement metrics of a selected social media page/account. Tasks: □ Compute likes, shares, comments, reach	4	CO2, CO4
	• Identify best-performing content • Suggest business insights		
5	Plot the given graph in python and calculate its Degree centrality, density, connectivity	4	CO1
6	Analyze public sentiment on a trending topic (e.g., budget, movie release, product launch). Tasks: • Clean text data • Perform sentiment analysis • Generate word cloud	4	CO3, CO4
7	Data Collection-Select the social media platforms of your choice (X, Facebook, LinkedIn, YouTube, Web blogs etc), connect to and capture social media data for business (scraping, crawling, parsing).	4	CO3, CO4
8	Use Google Trends to study interest patterns for a topic/product. Tasks: • Compare keyword trends • Analyze regional popularity • Correlate with social media trends	4	CO5
Reference Books			
1. Gohar F. Khan. Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine and Location Data. ISBN: 1507823207, 2015. 2. Analyzing the Social Web 1st Edition by Jennifer Golbeck 3. Python Social Media Analytics: Analyze and visualize data from Twitter, YouTube, GitHub, and more Kindle Edition by Siddhartha Chatterjee , Michal Krystyanczuk 4. Matthew A. Russell. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, Github, and More, 2nd Edition, O'Reilly Media, 2013			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-321-VSC-P : Business Intelligence and Dashboarding Tools

No. of Credits: 2		Teaching Scheme Practical : 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Basic understanding of statistics • Fundamental knowledge of data concepts (rows, columns, datasets) • Familiarity with spreadsheets (Excel/CSV files) • Basic exposure to programming concepts is desirable			
Objectives • To introduce students to advanced data visualization and business intelligence tools • To develop practical skills in Tableau and Power BI for data analysis • To enable students to perform statistical analysis using BI tools • To create interactive dashboards for decision making • To expose students to integration of Python/R scripts with BI tools.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understand concepts of business intelligence and data visualization CO2: Create interactive visualizations and dashboards using Tableau Public CO3: Perform data transformation and modelling using Power BI Desktop CO4: Apply DAX for basic statistical calculations and analytical measures CO5: Integrate Python/R scripts in Power BI for advanced analytics			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Tableau and Data Visualization	12	CO1, CO2
➤ Introduction to Business Intelligence (BI) and Data Visualization ➤ Importance of visualization in data science and analytics ➤ Overview of Tableau: features, architecture, and applications ➤ Installation of Tableau Public / Tableau Desktop ➤ Tableau interface: worksheets, dashboards, stories ➤ Understanding data types, measures, and dimensions ➤ Connecting Tableau to Excel, CSV, and text files for data visualization.			
2	Data Analysis and Dashboarding using Tableau	12	CO2
➤ Data preparation and data cleaning in Tableau ➤ Creating basic and advanced visualizations ➤ Formatting and customization of charts ➤ Filters, sorting, and grouping ➤ Calculated fields and table calculations ➤ Parameters for dynamic analysis ➤ Designing interactive dashboards			

➤ Dashboard actions and story creation ➤ Best practices for effective dashboard design			
3	Introduction to Power BI and Data Modeling	16	CO3, CO4
➤ Introduction to Power BI and its components ➤ Installation and overview of Power BI Desktop ➤ Connecting Power BI to various data sources ➤ Introduction to Power Query Editor ➤ Data cleaning and transformation ➤ Data modelling concepts ➤ Creating and managing relationships ➤ Understanding cardinality and star schema			
4	Data Analysis and Reporting using Power BI	12	CO4
➤ Introduction to DAX (Data Analysis Expressions) ➤ DAX syntax and structure ➤ Calculated columns and measures ➤ Aggregation and statistical DAX functions ➤ Basic statistical calculations using DAX • Aggregation functions: SUM(), AVERAGE(), MIN(), MAX(), COUNT() • Logical functions: IF(), SWITCH() • Statistical functions: MEDIAN(), STDEV.P(), VAR.P() ➤ Filter context and row context ➤ Using CALCULATE() function ➤ Time intelligence functions • TOTALYTD(), TOTALMTD(), TOTALQTD() ➤ Creating interactive reports using filters and slicers			
5	Advanced Power BI and Integration	8	CO5
➤ Introduction to Python and R integration in Power BI ➤ Installing and configuring Python and R environments for Power BI ➤ Using Python scripts in Power BI for data preprocessing and visualization ➤ Executing R scripts in Power BI for statistical analysis and modelling ➤ Creating Python/R visuals within Power BI reports ➤ Advantages and limitations of using Python and R scripts in Power BI Desktop			
Reference Books			
1. Book1: Microsoft Power BI Cookbook by Brett Powell 2. Book2: Tableau Your Data by Daniel G. Murray, ISBN: 978-1-118-61204-0, John Wiley & Sons, Inc., Indianapolis, Indiana 3. Book3: Mastering Microsoft Power BI by Brett Powell, ISBN 978-1-78829-723-3, Packt Publishing Ltd. 4. Book4: Learning Tableau 2019, Third Edition by Joshua N Milligan, ISBN 978-1-78883-952-5 5. Book5: Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-331-FP: Project on Data Analytics & Visualization

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Knowledge of programming concepts • Knowledge of statistics fundamentals • Knowledge of database concepts • Knowledge of spreadsheet handling • Knowledge of Python/R Programming and Power BI/ Tableau		
Objectives • Understand the complete data science lifecycle from data collection to visualization. • Apply programming concepts using Python or R for data analysis and pre-processing. • Perform exploratory data analysis and statistical analysis on real-world datasets. • Develop analytical and problem-solving skills using data science techniques. • Interpret data patterns, trends, and predictive insights for decision-making. • Enhance project documentation		
Course Outcomes On Completion of this course, student will be able to - CO1: Identify and formulate real-world data science problems. CO2 : Collect, clean, pre-process, and organize datasets using Python or R. CO3 : Perform exploratory data analysis and statistical interpretation. CO4 : Apply suitable machine learning or analytical techniques for problem-solving. CO5 : Design interactive dashboards and visualizations using Power BI or Tableau. CO6 : Present analytical findings through reports, dashboards, and presentations.		
Software & Tools Programming Tools • Python • RStudio Visualization Tools • Microsoft Power BI • Tableau Supporting Tools • Jupyter Notebook • Microsoft Excel • CSV / JSON datasets		
Sr. No.	Guidelines for Project	
1	Problem Statement • Define the project problem • Identify objectives and scope	

	• Determine expected outputs Students should choose a real-world problem from areas such as: • Healthcare • Banking • Education • Agriculture • Retail • Finance • Social Media • Sports Analytics • Transportation and many more
2	Data Collection • Collect data from reliable sources • Understand dataset structure and attributes
3	Data Preprocessing Using Python/R perform: • Missing value handling • Duplicate removal • Data cleaning • Data transformation • Feature selection
4	Exploratory Data Analysis (EDA) Students should analyze: • Trends • Relationships • Correlations • Outliers • Statistical summaries Visualization Techniques • Bar Charts • Histograms • Pie Charts • Scatter Plots • Heat Maps • Line Graphs
5	Model Building (Optional) • Using Regression, Classification or Clustering techniques of statistics
6	Dashboard Development Using Power BI or Tableau: • Create interactive dashboards • Add filters and slicers • Include KPI indicators • Display summary reports and charts

7	Result Interpretation Students should: • Explain analytical findings • Compare outcomes • Draw conclusions from data
8	Documentation and Presentation • Students should prepare: • Project report (Hardcopy & Softcopy) • PPT presentation • Dashboard demonstration

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-V

DS-341-MN-T: Categorical and Multivariate Data Analysis

No. of Credits: 02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: Marks 20 End Semester: 30 Marks	
Prerequisites ● Fundamental knowledge of Probability Theory ● Basic Statistical Inference ● Introductory Linear Algebra (matrices, eigenvalues, eigenvectors)			
Objectives ● Introduce statistical techniques for analyzing categorical (qualitative) data. ● Develop understanding of logistic and log-linear modeling frameworks. ● Provide foundations of multivariate statistical methods. ● Enable students to apply dimension reduction and classification techniques.			
Course Outcomes On Completion of this course, student will be able to - CO1: Explain the structure and properties of categorical and multivariate data. CO2: Apply measures of association, and trend tests to analyze contingency tables. CO3: Fit and interpret logistic regression and log-linear models for categorical outcomes. CO4: Compute and interpret multivariate summary measures. CO5: Perform and interpret Principal Component Analysis and Factor Analysis. CO6: Apply classification and clustering techniques.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Categorical Data Analysis	08	CO1, CO2
Types of Categorical Data: Nominal vs Ordinal variables, Binary data, multinomial data, Contingency tables (2-way and $r \times c$ tables) Measures of Association: Odds and Odds Ratio, Relative Risk, Yule’s Q, Phi coefficient, Cramér’s V Chi-Square Methods: Pearson’s Chi-square test, Likelihood ratio test, Test of independence, Test of homogeneity Exact & Trend Tests: Fisher’s Exact Test, Mantel–Haenszel Test, Test for linear trend (ordinal data)			
2	Log-linear Models for Categorical Data	07	CO3
Log-linear Models: Poisson regression model, Two-way log-linear model, Interaction terms, Model selection (Backward/Forward approach) Logistic Regression: Binary logistic regression, Interpretation of coefficients, Odds ratio interpretation, Goodness-of-fit measures Multinomial & Ordinal Logistic Models: Baseline category logit model, Proportional odds model (overview)			

3	Introduction to Multivariate Data Analysis	07	CO4, CO5
Multivariate Concepts: Random vectors, Mean vector & covariance matrix, Correlation matrix, Mahalanobis distance Multivariate Normal Distribution: Definition and properties (without proof) Principal Component Analysis (PCA): Eigenvalues and eigenvectors, Variance explained, Dimensionality reduction Factor Analysis: Model formulation, Factor loadings, Communality, Varimax rotation (conceptual overview)			
4	Dimension Reduction and Classification	08	CO6
Discriminant Analysis: Linear Discriminant Analysis (LDA), Classification rule, Misclassification probability Cluster Analysis: Hierarchical clustering, K-means clustering, Distance measures Introduction to Canonical Correlation Analysis: Concept, Interpretation, Applications in data science			
Reference Books			
1. Agresti, A. (2018). <i>Categorical data analysis</i> (3rd ed.). Wiley. 2. Anderson, T. W. (2003). <i>An introduction to multivariate statistical analysis</i> (3rd ed.). Wiley. 3. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). <i>Multivariate data analysis</i> (8th ed.). Cengage Learning. 4. Johnson, R. A., & Wichern, D. W. (2018). <i>Applied multivariate statistical analysis</i> (6th ed.). Pearson. 5. McCullagh, P., & Nelder, J. A. (1989). <i>Generalized linear models</i> (2nd ed.). Chapman & Hall. 6. Bruce, P., Bruce, A., & Gedeck, P. (2020). <i>Practical statistics for data scientists: 50+ essential concepts using R and Python</i> (2nd ed.). O'Reilly Media. 7. Fox, J., & Weisberg, S. (2019). <i>An R companion to applied regression</i> (3rd ed.). Sage. 8. Hastie, T., Tibshirani, R., & Friedman, J. (2009). <i>The elements of statistical learning: Data mining, inference, and prediction</i> (2nd ed.). Springer.			

Detail Syllabus

B.Sc. (Data Science)

Semester-VI

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-351-MJ-T: NoSQL Databases

No. of Credits: 4	Teaching Scheme Theory:4 Hrs./Week	Examination Scheme Continuous Evaluation: 30 Marks End Semester: 70 Marks	
Prerequisites ● Knowledge of DBMS concepts ● Basics of SQL- data retrieval and aggregation ● Knowledge of file system concepts ● A firm foundation of any RDBMS package ● Fundamentals data structures for efficient data storage and analysis			
Objectives ● Gives a basic understanding of NoSQL technology. ● Provide an insight into different types of NoSQL databases. ● Provide hands-on experience with popular NoSQL databases ● Helps students choose the right database based on application needs. ● Enables students to use NoSQL databases through suitable frameworks.			
Course Outcomes On Completion of this course, student will be able to - CO1: To know principles, and architecture of NoSQL databases CO2: Analyze and design appropriate aggregation models CO3: Apply distribution models for scalable data systems CO4: Basics of Key-Value and Column-Family databases CO5: CRUD Operation- Document and Graph databases CO6: Apply Polyglot Persistence and MapReduce techniques CO7: Use NoSQL databases in real-world Data Science applications CO8: Select suitable NoSQL databases for data science and big data applications			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to NOSQL	8	CO1
1.1 Evolution of data management systems			
1.2 Limitations of RDBMS for Big Data			
1.3 Characteristics of Big Data (5Vs)			
1.4 Introduction to NoSQL databases			
1.5 CAP Theorem and BASE properties			
1.6 ACID vs BASE			
1.7 Types of NoSQL databases			
1.8 Schema-agnostic nature of NoSQL databases			
1.9 Use cases in Data Science and Big Data ecosystems			

2	Aggregation Models	8	CO2
2.1 Concept of aggregation in NoSQL systems 2.2 Aggregate-oriented databases 2.3 Key-Value aggregation model 2.4 Document aggregation model 2.5 Column-family aggregation model 2.6 Aggregate design principles 2.7 Query-driven data modelling 2.8 Comparison with relational normalization 2.9 Schema flexibility and schema evolution in aggregate models			
3	Distribution Models	10	CO3
3.1 Need for distributed data storage 3.2 Data partitioning strategies 3.3 Sharding techniques: Range-Based Sharding, Hashed Sharding, Zone-Based Sharding (Tag-Aware Sharding) 3.4 Replication models: Master-Slave Replication, Multi-Master Replication, Peer-to-Peer architectures, Log-Based/Incremental Replication 3.6 Consistency models: strong consistency, weak consistency (including eventual consistency), and causal consistency 3.8 Fault tolerance and availability 3.9 Version stamps and conflict resolution in distributed systems			
4	NoSQL Databases : Key Value & Column Family	8	CO4
4.1 Key-Value databases: concepts and architecture 4.2 Operations and data access patterns 4.3 In-memory databases 4.4 Redis architecture and commands 4.5 Riak overview 4.6 Column-family databases 4.7 Cassandra architecture and data model 4.8 HBase architecture 4.9 Data modelling and querying 4.10 Use cases in analytics, caching, and logging			
5	NoSQL Databases: Document Store & Graph Database	10	CO5
5.1 Document-oriented databases 5.2 JSON, BSON, and XML document formats 5.3 MongoDB architecture 5.4 CRUD operations and indexing 5.5 Aggregation framework in MongoDB 5.6 Graph database fundamentals 5.7 Property graph model 5.8 Neo4j architecture 5.9 Graph traversal and queries 5.10 Use cases: Social Networks, Recommendation Engines			

6	Polyglot Persistence, Map Reduce	8	CO6,CO8
6.1 Introduction to Polyglot Persistence 6.2 Benefits and challenges 6.3 Database selection strategies 6.4 Integrating multiple data stores 6.6 Introduction to MapReduce 6.7 Map and Reduce functions 6.8 Hadoop ecosystem overview 6.9 Database Performance Monitoring tools 6.10 Introduction to cloud databases: Amazon RDS, Google Cloud Firestore, Microsoft Azure SQL Database 6.11 Introduction to AI Databases: Pinecone, Milvus, Weaviate, Qdrant			
7	NoSQL in Data Science Applications	8	CO7,CO8
7.1 Role of NoSQL in Data Science lifecycle 7.2 Data ingestion and preprocessing 7.3 NoSQL for real-time analytics 7.4 Handling unstructured and semi-structured data 7.5 NoSQL integration with Python and R 7.6 Case studies: • Recommendation systems • Social media analytics • IoT and sensor data analysis 7.7 Mini case study discussion			
Reference Books			
1 NoSQL Distilled: Pramod J. Sadalage and Martin Fowler, Addison-Wesley, Pearson 2 NoSQL for Mere Mortals, Dan Sullivan, Addison-Wesley 3 Professional NoSQL, Shashank Tiwari, Wrox Press 4 MongoDB: The Definitive Guide, Kristina Chodorow, O'Reilly Media. 5 Seven Databases in Seven Weeks, Eric Redmond and Jim R. Wilson, Pragmatic Bookshelf. 6 Ian Robinson, Jim Webber, and Emil Eifrem, Graph Databases, O'Reilly Media.			
Online Learning Resources			
2 MongoDB University – Practical labs and case studies 3 Apache Cassandra Documentation – Industry-grade reference 4 Redis University – Key-Value store fundamentals 5 Neo4j GraphAcademy – Hands-on graph database learning			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)

Sem-VI

DS-352-MJ-T : Artificial Intelligence in Data Science

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation:15 Marks End Semester : 35 Marks	
Prerequisites • Fundamental concepts, scope, and evolution of Artificial Intelligence. • Knowledge of Programming Language • Data Analytics Skill			
Objectives • Introduce students to AI problem-solving techniques relevant to data-driven applications. • Enable understanding and application of search and heuristic algorithms for decision making. • Familiarize students with game theory, constraint satisfaction, and planning techniques used in AI systems. • Develop the ability to analyse AI strategies and evaluate their performance in data science problems. • Provide exposure to modern AI trends, including Generative AI and Explainable AI. • Prepare students to integrate AI techniques with data science workflows through practical implementation.			
Course Outcomes On Completion of this course, students will be able to – CO1: Apply uninformed and informed search techniques to solve data-driven AI problems. CO2: Represent problems and knowledge using suitable AI representations. CO3: Design AI-based decision-making solutions using game search and CSP techniques. CO4: Analyse and evaluate AI model behaviour and search performance. CO5: Analyse and implement AI-based planning and scheduling techniques for data science applications under constraints and uncertainty. CO6: Understand recent trends and applications of AI in Data Science			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Problem Solving Using Search Techniques	9	CO1,CO2
1.1 Problems Solving methods			
1.2 Problem-Solving Agents			
1.3 Example Problems			
1.4 Search Algorithms			

1.5 Blind Search Techniques: - Breadth-first search(BFS) - Depth-first search(DFS) - Uniform cost Search. 1.6 Heuristic search techniques: - - Hill Climbing - Best First search - Constraint Satisfaction - A*			
2	Game Theory and Decision Making	9	CO3,CO4
2.1 Introduction Game Theory - Types of Game 2.2 Optimal Decisions in Games - Application of Optimal Decisions in Games 2.3 Heuristic Alpha–Beta Tree Search - Application of Heuristic Alpha–Beta Tree Search 2.4 Monte Carlo Tree Search - Application of Monte Carlo Tree Search 2.5 Constraint Satisfaction Problems (CSP). - Application of Constraint Satisfaction Problem 2.6 Limitations of Game Search Algorithms			
3	Planning and Scheduling for Data Science Applications	7	CO5
3.1 Introduction to Classical and Automated Planning 3.2 Planning Algorithms and State-Space Models 3.3 Heuristic-Based Planning Techniques 3.4 Hierarchical Task Planning 3.5 Planning under Uncertainty and Resource Constraints 3.6 Comparative Analysis of Planning Approaches			
4	Recent Trends and Role of AI in Data Science	5	CO6
4.1 Role of AI in Data Science 4.2 Information Retrieval and Information Extraction 4.3 Introduction to Generative AI 4.4 Explainable AI and Ethical Considerations 4.5 Computer Vision Applications 4.6 AI Applications in Social Media and Finance 4.7 AI Application in Business and HR Analytics			
Reference Books			
Book1 : Artificial Intelligence: A Modern Approach by Stuart Russell & Peter Norvig Book2 : Artificial Intelligence by Elaine Rich & Kevin Knight Book3 : Artificial Intelligence: Foundations of Computational Agents by David Poole & Alan Mackworth Book4 : Artificial Intelligence Basics: A Non-Technical Introduction, Apress by Tom Taulli			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-353-MJ-T : Data Privacy and Security

No. of Credits: 2	Teaching Scheme Theory: 2 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites • Basic knowledge of computer networks and operating systems • Introductory knowledge of data science workflows • Understanding of database management systems			
Objectives • To introduce students to the principles of data security and privacy in digital systems • To explain cryptographic techniques and their applications in securing data • To familiarize students with privacy laws, frameworks, and ethical considerations • To develop skills for designing secure data science systems • To expose students to emerging trends like blockchain and differential privacy.			
Course Outcomes On Completion of this course, student will be able to - CO1: Understands the concepts of confidentiality, integrity, and availability in data security. CO2: Apply basic cryptographic methods (encryption, hashing, digital signatures) to secure data. CO3: Understand privacy principles and compliance with data protection laws. CO4: Design secure mechanisms for data storage, transmission, and access control in data Science systems. CO5: Understand role of emerging technologies: (Blockchain, differential privacy, AI-driven security) in data protection.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Introduction to Data Security	4	CO1
1.1 Introduction: CIA triad: confidentiality, integrity, availability 1.2 Types of threats: malware, phishing, insider threats 1.3 Vulnerabilities: software bugs, weak passwords, misconfigurations 1.4 Risk assessment and mitigation strategies 1.5 Security policies: organizational rules, compliance requirements			
2	Cryptography Basics	8	CO2
2.1 Introduction: Plain text and cipher Text 2.2 Substitution techniques 2.3 Transposition techniques			

2.4 Symmetric and Asymmetric key cryptography (Only introduction of DES,AES,RSA,SHA,MD-5) 2.4 Steganography 2.5 Digital signatures and certificates			
3	Privacy Concepts	6	CO3
3.1 Definitions: privacy vs. security 3.2 Legal frameworks: GDPR, Indian IT Act, HIPAA basics 3.3 Data anonymization: k-anonymity, l-diversity 3.4 Pseudonymization and masking techniques 3.5 Privacy in data publishing and sharing			
4	Secure Systems in Data Science	6	CO4
4.1 Access control models: RBAC (Role Based), ABAC (Attribute Based) 4.2 Authentication methods: passwords, biometrics, multifactor authentication 4.3 Secure storage: encryption at rest, database security 4.4 Secure transmission: Introduction to SSL/TLS, VPNs 4.5 Cloud security: shared responsibility model, data isolation 4.6 Big data security: Hadoop/NoSQL vulnerabilities, secure pipelines			
5	Emerging Topics	6	CO5
5.1 Differential privacy: noise addition, privacy budget 5.2 Blockchain: distributed ledger, immutability, smart contracts 5.3 AI-driven threat detection: anomaly detection, intrusion detection systems 5.4 Security challenges in IoT 5.5 Ethical considerations in data security and privacy			
Reference Books			
Book1: William Stallings, Cryptography and Network Security Book2: Daniel J. Solove, Understanding Privacy Book3: Arvind Narayanan et al., A Primer on Differential Privacy Book4: Atul Kahate, Cryptography and Network Security			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-354-MJ-P: Lab Course on DS-351-MJ-T (NoSQL Databases)

No. of Credits: 2	Teaching Scheme Practical: 4 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites ● Basic knowledge of database concepts (tables, records, keys) ● Understanding of programming fundamentals ● Familiarity with JSON data format ● Basic knowledge of data structures and relationships		
Objectives ● Introduce NoSQL database technologies using MongoDB and Neo4j ● Perform CRUD operations and manage data efficiently in MongoDB ● Design and query complex document structures using JSON/BSON formats ● Understand indexing and aggregation techniques for performance optimization ● Model real-world applications using document-oriented and graph databases ● Implement graph data models and traversal queries using Neo4j ● Integrate document and graph databases for analytical applications		
Course Outcomes On Completion of this course, student will be able to - CO1: Perform CRUD operations in MongoDB using standard queries. CO2: Design and manage nested MongoDB documents using JSON/BSON. CO3: Optimize MongoDB queries using indexing and aggregation. CO4: Design real-world applications using MongoDB document schemas. CO5: Create and manage nodes, relationships, graph traversal and constraints in Neo4j. CO6: Develop graph-based applications such as social networks and recommendation systems.		
Assignment 1 : Perform CRUD operations using MongoDB		CO1
Perform CRUD operations in MongoDB: create a collection, insert documents, read data with queries, update records, and delete documents.		
Assignment 2 : Create MongoDB documents using JSON/BSON formats		CO2
Insert and manipulate nested documents in MongoDB using JSON/BSON formats. Query the nested fields effectively.		
Assignment 3 : Create indexes on MongoDB collections		CO3
Create indexes on MongoDB collections and perform queries using filters, projections, and sorting. Measure query performance before and after indexing.		
Assignment 4 : Perform MongoDB aggregation		CO3
Implement MongoDB aggregation pipelines: filter, group, and summarize data from a collection. Demonstrate aggregation using multiple stages		

Assignment 6 : Design a small application using MongoDB	CO4
Design and implement a document schema in MongoDB for an application (e.g., e-commerce or blogging). Populate with sample data and perform queries.	
Assignment 7 : Create nodes and relationships	CO5
In Neo4j, create nodes and relationships representing a domain like users, products, posts . Use Cypher queries to insert data.	
Assignment 8 : Graph traversal queries	CO5
Perform graph traversal queries in Neo4j: find all connected nodes, shortest paths, and neighbors up to N levels.	
Assignment 9 : Design a small application using Neo4j	CO6
• Model a social network in Neo4j with users, friendships, posts, and likes. Write queries to find friends-of-friends and user activity. • Build a recommendation system in Neo4j using graph algorithms or traversal queries to suggest products or connections	
Assignment 10 : Integrate MongoDB and Neo4j	CO6
Integrate MongoDB and Neo4j: Store transactional data in MongoDB and relationships in Neo4j. Perform cross-database queries for analytics.	
Assignment 11 : Perform advanced Neo4j operations	CO6
Perform advanced Neo4j operations: create labels, constraints, indexes, and execute complex Cypher queries to extract insights.	
Reference Books: 1. MongoDB: The Definitive Guide – Kristina Chodorow – O’Reilly Media 2. MongoDB Applied Design Patterns – Shannon Bradshaw, Eoin Brazil, Kristina Chodorow – O’Reilly Media 3. Graph Databases – Ian Robinson, Jim Webber, Emil Eifrem – O’Reilly Media 4. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence – Pramod J. Sadalage, Martin Fowler – Addison-Wesley 5. Learning Neo4j – Rik Van Bruggen – Packt Publishing 6. Graph Algorithms: Practical Examples in Apache Spark and Neo4j – Mark Needham, Amy E. Hodler – O’Reilly Media 7. NoSQL Distilled – Jesse Anderson – Addison-Wesley	

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-355-MJ-P : Lab Course on DS-352-MJ-T (Artificial Intelligence in Data Science)

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs./Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites - Fundamentals of Artificial Intelligence - Data Structures and Algorithms - Python Programming - Basic Discrete Mathematics		
Objectives - To implement AI algorithms using Python - To analyse search, planning, and decision-making techniques - To apply AI methods to real-world data science problems		
Course Outcomes On Completion of this course, students will be able to – CO1: Implement and analyse uninformed and informed (heuristic) search techniques to solve state-space and graph-based problems relevant to data science CO2: Formulate and solve constraint satisfaction problems such as scheduling and resource allocation using appropriate AI techniques. CO3: Implement game-theoretic decision-making algorithms to achieve optimal strategies in competitive and adversarial environments. CO4: Apply heuristic planning techniques for intelligent task scheduling and resource optimization in data-driven applications. CO5: Design, implement, and evaluate rule-based intelligent systems, and analyse the performance of AI algorithms from a data science perspective.		
Operating Environment: Operating Systems: - Microsoft Windows (Windows 10 or higher) - Linux (Ubuntu 20.04 or higher) - macOS (macOS Catalina or higher) Programming Language: - Python 3.x Hardware Requirements: - Minimum 4 GB RAM - Processor: Intel i3 / AMD equivalent or higher - Minimum 10 GB free disk space		

Python Development Environments

To write, execute, and debug Python programs efficiently, students can use any of the following tools:

- Anaconda Distribution (Recommended for Data Science students)
- Python IDLE
- PyCharm Community Edition
- Visual Studio Code (VS Code)
- Jupyter Notebook / JupyterLab

Required Python Libraries (As applicable to experiments)

- numpy
- pandas
- matplotlib
- scikit-learn
- networkx (for graph-based experiments – optional)

Execution Environment

Programs may be executed using:

- Command Line / Terminal
- Integrated Development Environment (IDE)
- Jupyter Notebook environment

Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Assignment on Uninformed Search Techniques	6	CO1, CO5
• Implementation of Breadth First Search (BFS) • Implementation of Depth First Search (DFS) • Implementation of UCS for a weighted graph • Application Oriented Implementation: ○ Cycle Detection in data pipelines(DFS) ○ Network traversal (BFS) ○ Supply Chain Analytics using UCS			
2	Assignment on Informed (Heuristic)Search Techniques	8	CO1, CO4,CO5
• Implementation of A* Search Algorithm • Implementation of Hill Climbing • Implementation of Best First Search • Application Oriented Implementation: ○ Delivery Route Optimization using A* Search ○ Hill Climbing Algorithm for Feature Selection ○ Best First Search for Delivery Prioritization ○ Heuristic Planning for Task Scheduling using Best First Search			

3	Assignment on Constraint Satisfaction Problem	6	CO2, CO4
• Application-Oriented Implementation ○ Resource Allocation using Constraint Satisfaction Techniques ○ Scheduling using Constraint Satisfaction Techniques			
4	Assignment on Game Theory	4	CO3, CO5
• Implement of Alpha–Beta Pruning Algorithm • Application Oriented Implementation: ○ Tic-Tac-Toe game using Alpha–Beta Pruning			
5	Assignment on Rule based system	6	CO5
• Implementation of Simple Chabot • Implementation of Rule-based Spam Filter • Implementation of Rule-based Fraud Detection			
Reference Books			
Book1: Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Pearson Education.			
Book2: Elaine Rich, Kevin Knight, Artificial Intelligence, McGraw-Hill.			
Book3: Ivan Vasilev et al., Python Deep Learning, Packt Publishing.			
Book4: Python Data Science Handbook <i>O'Reilly Media</i>			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-360-MJ-T : HR Analytics

No. of Credits: 02	Teaching Scheme Theory: 2 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites ● Knowledge of Descriptive Statistics. ● Basic Knowledge of Data Science. ● Practical knowledge of R Programming.			
Objectives ● To explore evolving nature of HR analytics ● To understand analysis strategies required for HR analytics. ● To have hands on practice for Predictive HR analytics using R programming language.			
Course Outcomes On Completion of this course, student will be able to - CO1: Defined HR analytics and its metrics CO2: Preform HR data importing and merging in R environment CO3: Do statistical test on HR data using R language CO4: Perform analytics like diversity analytics, employee engagement and performance analytics, predictive analytics for employee turnover using R programming language.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	HR analytics :Introduction	5	CO1
1.1. Analytics in HR, Defined Predictive HR analytics 1.2. Introduction of HCM (Human Capital Management) Model, Metrics and Analytics 1.3. Understanding Descriptive Analytics and Perspective Analytics, 1.4. Understanding the need (and business case) for mastering and utilizing predictive HR analytic techniques 1.5. Human capital data storage and ‘big (HR) data’ manipulation 1.6. Predictors, prediction and predictive modelling 1.7. Current state of HR analytic capabilities and professional or academic training 1.8. Business applications of modelling 1.9. HR analytics and HR people strategy 1.10. Becoming a persuasive HR function			
2	HR information systems and data	5	CO2

1.1. Information sources 1.2. Analysis software options 1.3. Using artificial intelligence and large language models to aid analyses Using R 1.4. Importing and merging data 1.5. Big data			
3	Analysis strategies	12	CO3
1.1. From descriptive reports to predictive analytics 1.2. Statistical Significance 1.3. HR analytic metrics/measures 1.4. Data integrity 1.5. Types of data, Categorical and Continuous variables 1.6. Group level and individual level data 1.7. Dependent and Independent variables 1.8. Types of statistical tests 1.9. Statistical test for categorical data 1.10. Statistical test for continuous data/interval level data 1.11. Factor analysis and reliability analysis			
4	Case Studies in HR analytics	8	CO4
1.1. Diversity analytics: Equality, diversity and inclusion, Approaches to measuring and managing D&I, Testing the impact of diversity. 1.2. Employee attitude surveys: employee engagement and workforce perceptions, measure employee engagement, Interrogating the measures, 1.3. Predicting employee turnover: Descriptive turnover analysis, measuring turnover 1.4. Predicting employee performance: measures for performance, predict performance using multiple linear regression .			
Reference Books			
1. <i>Using R in HR Analytics: A Practical Guide to Analysing People Data</i> . Edwards, Martin R., Kirsten Edwards, and Daisung Jang.London, Kogan Page, 2024 2. <i>The Practical Guide to HR Analytics: Using Data to Inform, Transform, and Empower HR Decisions</i> , Waters, Shonna D., Valerie Streets, Lindsay McFarlane, and Rachael Johnson-Murray. Alexandria, VA: Society for Human Resource Management, 2018 3. <i>The New HR Analytics: Predicting the Economic Value of Your Company's Human Capital Investments</i> . New York. Fitz-Enz, Jac, AMACOM, 2010 4. <i>Leveraging Data for Competitive Advantage</i> . New Delhi, Soundararajan, Ramesh, and Kuldeep Singh. <i>Winning on HR Analytics</i> .Sage Publications / Atlantic Publishers & Distributors, 2019			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-361-MJ-P : Lab course on DS-360-MJ-T (HR Analytics)

No. of Credits: 02	Teaching Scheme Practical: 4 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites - R Programing - Knowledge of HR Analytics			
Objectives - To introduce students to the R programming environment for creating, importing, managing, and pre-processing HR datasets. - To develop students' ability to apply statistical techniques in R for analysing employee's diversity and inclusion data within organizations. - To enable students to analyse employee attitude and engagement data using appropriate analytical tools and techniques in R. - To equip students with the skills to build and interpret predictive models for employee turnover using R. - To provide hands-on experience in performing predictive analytics for evaluating and forecasting employee performance outcomes.			
Course Outcomes On Completion of this course, student will be able to - CO1: Create or import datasets in R environment CO2: Perform employee diversity analytics using R CO3: Perform Employee attitude analytics using R language CO4: Do Predictive analytics for employee turnover. CO5: Do Predictive analytics for employee performance.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	R Studio Environment	4	CO1
Explore R Studio Environment, Revision Data types, Data structures (vectors, lists, matrices, arrays), create Data set and Data frame, Make plots and visualize data (line, scatter, pie, bar), Importing and Merging Data.			
2	Diversity analytics	8	CO2

• Prepare and import data set with useful attributes for Gender and grade analysis • Write Program using R for Gender and job grade analysis using frequency tables and Chi square • Write Program using R for exploring ethnic diversity across teams descriptive statistics • Write Program using R for Multiple linear regression to model and predict ethnic diversity variation across teams.			
3	Employee Attitude Survey	16	CO3
• Prepare datasets for Employee survey to assess employee engagement. • Write Program using R for Exploratory and factor analysis • Write Program using R for Regression and descriptive analytics for survey data			
4	Predicting Employee Turnover	16	CO4
• Prepare datasets suitable for do the analytics regarding employee turnover including attributes like Gender, Age, LengthOf Service, AppraisalRating, Country, LeaverStatus etc. • Write Program in R language using frequency tables to explore regional differences in staff turnover • Write Program in R language using Chi-square analysis to explore regional differences in individual staff turnover • Write Program in R language for predicting individual turnover • Write Program in R language for comparing expected length of service for men vs women.			
5	Predicting Employee Performance	16	CO5
• Prepare datasets suitable for do the analytics regarding employee turnover • Write Program in R language uses multiple regression to predict sickness absence using employee attitude scores. • Write Program in R language uses multiple regression to explore patterns of performance with employee profile data in an organisation.			
Reference Books			
1. <i>Using R in HR Analytics: A Practical Guide to Analysing People Data</i>. Edwards, Martin R., Kirsten Edwards, and Daisung Jang. London, Kogan Page, 2024 2. <i>R for HR: An Introduction to Human Resource Analytics Using R</i>. Caughlin, David E. 2022. Online book. https://rforhr.com/			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-362-MJ-T : Financial Analytics

No. of Credits: 02	Teaching Scheme Theory: 02 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks	
Prerequisites ● Basic Statistics knowledge. ● Python or R proficiency. ● Understanding of Linear Algebra fundamentals and finance.			
Objectives ● To acquaint students with corporate financial modeling and insolvency prediction. ● To introduce portfolio optimization and risk-adjusted performance metrics. ● To identify market volatility patterns using advanced time-series frameworks. ● To apply technical indicators and visual analytics for trend detection.			
Course Outcomes On Completion of this course, student will be able to - CO1: Evaluate corporate solvency and project viability using financial modeling and bankruptcy analytics. CO2: Quantify market risk and price fluctuations using volatility models and Value at Risk (VaR). CO3: Construct optimized asset portfolios and assess performance using risk-adjusted metrics. CO4: Identify market trends and signals through technical indicators and visual candlestick analysis.			
Unit No.	Name of Unit	Teaching Hours	CO Targeted
1	Corporate Finance, Financial Modeling & Portfolio Management	15	CO 1, CO 3
1.1 Introduction to Finance & Accounting: Understanding financial statements (Balance Sheet, P&L, Cash Flow), data sources, and data cleaning for accounting analytics.			
1.2 Project Analysis & Modeling: Cash flow forecasting, Cost of Capital (WACC) using sensitivity analysis, and Financial Break-even modeling.			
1.3 Capital Budgeting Models: Payback Period, NPV, IRR, and MIRR.			
1.4 Bankruptcy & Fraud Analytics: Modeling corporate distress using Beaver’s t-test, Ohlson’s Logistic Regression, and the Altman Z-score. Introduction to Fraud detection in accounting data.			
1.5 Modern Portfolio Theory: Markowitz’s Mean-Variance Optimization and the Capital Asset Pricing Model (CAPM).			
1.6 Performance Metrics: Sharpe Ratio, Treynor Ratio, and Cluster Analysis for portfolio categorization.			

2	Financial Market & Time Series Analysis	15	CO 2, CO 4
2.1 Data Acquisition: Importing financial data from web portals (APIs), adjusting for stock splits and mergers, and handling time-series gaps. 2.2 Risk and Return: Estimation and prediction of returns for bonds and stocks. 2.3 Advanced Time Series Modeling: Examining data stationarity, EWMA (Exponentially Weighted Moving Average), and Value at Risk (VaR). 2.4 Volatility Modeling: Implementing ARMA, ARCH, and GARCH models to predict market fluctuations. 2.5 Technical Indicators: RSI (Relative Strength Index), ROC, MACD, and Moving Averages. 2.6 Visual Analytics: Candlestick chart patterns and volume analysis.			
Reference Books			
For (UNIT 1) 1. Financial Modeling by Simon Benninga. 2. Corporate Financial Analysis by Blaine Robertson. 3. Corporate Financial Distress and Bankruptcy by Edward Altman. 4. <i>Modern Portfolio Theory and Investment Analysis</i> by Edwin J. Elton & Martin J. Gruber. For (UNIT 2) 5. Analysis of Financial Time Series by Ruey S. Tsay. 6. Introductory Econometrics for Finance by Chris Brooks. 7. Forecasting: Principles and Practice by Hyndman & Athanasopoulos. 8. <i>Python for Finance: Mastering Data-Driven Finance</i> by Yves Hilpisch. 9. <i>Technical Analysis of the Financial Markets</i> by John J. Murphy. OTHERS 10. Options, Futures, and Other Derivatives by John C. Hull. 11. Investment Analysis and Portfolio Management by Reilly & Brown. 12. Quantitative Trading by Ernest P. Chan. 13. Python for Algorithmic Trading by Yves Hilpisch.			

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-363-MJ-P : Lab Course on DS-362-MJ-T (Financial Analytics)

No. of Credits: 02	Teaching Scheme Practical: 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites • Basic Statistics knowledge. • R Programming proficiency. • Financial Math fundamentals.		
Objectives • To acquaint students with R-based financial modeling and insolvency prediction. • To introduce portfolio optimization and risk-adjusted performance evaluation. • To identify market volatility patterns using statistical time-series frameworks. • To apply technical indicators and visual analytics for trend detection.		
Course Outcomes On Completion of this course, student will be able to - CO1: Execute corporate finance and project feasibility models using R. CO2: Quantify market risk and volatility using GARCH and VaR. CO3: Construct optimized portfolios using the Markowitz Efficient Frontier. CO4: Build automated technical dashboards for market signal interpretation.		
Sr. No.	List of Practical Assignments	Teaching Hours
1	Import raw CSV/Excel accounting data and use dplyr to calculate key liquidity and profitability ratios.	04
2	Write an R function to calculate NPV and IRR for uneven cash flows and perform sensitivity analysis on the discount rate.	06
3	Implement a script to compare multiple projects using Payback Period and MIRR (Modified Internal Rate of Return).	06
4	Apply the Altman Z-Score formula to real-time company data to classify financial distress zones.	04
5	Use the Portfolio Analytics package to plot the Markowitz Efficient Frontier and identify the minimum variance portfolio.	06
6	Calculate Beta (β) for a stock by regressing its returns against a market index (e.g., Nifty 50) to determine the Cost of Equity.	06

7	Compute and visualize the Sharpe and Treynor Ratios for a multi-asset portfolio using the Performance Analytics library.	04
8	Use quantmod to fetch live stock data via Yahoo Finance API and adjust for stock splits and dividends .	04
9	Perform the Augmented Dickey-Fuller (ADF) test on stock prices and transform them into log-returns for analysis.	04
10	Calculate the Value at Risk (VaR) using the Historical Simulation and Parametric (Gaussian) methods in R.	04
11	Fit ARCH and GARCH (1,1) models to capture volatility clustering in financial returns using the rugarch package.	04
12	Generate and plot RSI, MACD, and Moving Averages overlaid on interactive Candlestick charts using tidyquant .	04
13	Analyze the correlation between trading volume and price breakouts using visual scatter plots and ROC (Rate of Change) indicators.	04

Reference Books

1. **Reproducible Finance with R** by **Jonathan K. Regenstein** (Covers: Practicals 1, 5, 6, 7, and 8.)
2. **Analysis of Integrated Financial Statistics with R** by **Ronald D. Fricker & N.M. Lewis**. (Covers: Practicals 9, 10, and 11).
3. **Technical Analysis with R** by **Christopher Hare**. (Covers: Practicals 12 and 13).
4. **Financial Analytics with R** by **Mark J. Bennett & Dirk L. Hugen**. (Covers: Practicals 2, 3, and 4).
5. **Statistics and Data Analysis for Financial Engineering** by **David Ruppert & David S. Matteson**. (Covers: Practicals 5, 6, and 11).

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-371-VSC-P : Lab Course on Statistical Analysis using PSPP

No. of Credits: 02	Teaching Scheme Practical: 04 Hrs/Week	Examination Scheme Continuous Evaluation: 15 Marks End Semester : 35 Marks
Prerequisites Basic Statistics knowledge.		
Objectives - To understand the fundamentals of statistical data analysis using PSPP. - To perform data entry, coding, and management of research data effectively. - To apply descriptive and inferential statistical techniques for data analysis. - To interpret statistical results and prepare meaningful research reports. - To develop practical skills in using open-source statistical software		
Course Outcomes At the end of this course, students will be able to: CO1: Demonstrate competence in data management using PSPP, including data entry, variable coding, transformation, and preparation of datasets for analysis. CO2: Apply descriptive statistics, correlation, regression, and hypothesis testing techniques using PSPP and accurately interpret the statistical output. CO3: Analyze real-world datasets using appropriate statistical procedures in PSPP and present clear, evidence-based conclusions in report format.		
Sr. No.	List of Practical Assignments	Teaching Hours
1	Introduction to PSPP Interface and Data Entry: Understanding Data View and Variable View, Defining variables (numeric, string, labels, value labels), Saving and opening data files (.sav format)	06
2	Data Import and Coding: Importing data from Excel/CSV, Recoding variables (Recode into Same/Different Variables)	06
3	Data Visualization: Frequency tables, Diagram and graphs	06

4	Descriptive Statistics – Univariate Analysis: Computation of Mean, Median, Mode, Standard deviation, Variance	06
5	Cross Tabulation and Chi-Square Test: Creating contingency tables, Performing Chi-square test of independence, Interpretation of output	06
6	Correlation Analysis: Computation of Pearson's correlation coefficient, Spearman's rank correlation, Testing significance, Interpretation of correlation matrix	06
7	Fitting of Simple, Multiple and Logistic Linear Regression Model	06
8	Testing of hypothesis: Testing difference between two means, Checking homogeneity of variance, Interpretation of output	06
9	Parametric and Non-parametric Tests	06
10	Reliability Analysis: Cronbach's Alpha, Scale reliability, Interpretation for survey data	06

Savitribai Phule Pune University
B.Sc. Data Science (2024 Pattern)
Sem-VI

DS-381-OJT: On Job Training

No. of Credits: 4	Duration: 120 Hrs	Examination Scheme Continuous Evaluation: 30 Marks End Semester: 70 Marks
Objectives - To apply data science concepts to real-world industrial problems. - To gain practical exposure to data collection, preprocessing, and analysis techniques. - To work with data science tools, programming languages, and visualization platforms. - To develop analytical thinking and problem-solving abilities using data-driven approaches. - To understand business intelligence and decision-making processes. - To improve teamwork, communication, and professional ethics. - To gain industry exposure in analytics, and dashboard development.		
Course Outcomes On Completion of this course, student will be able to - CO1: Understand real-world applications of data science in industry. CO2 : Perform data preprocessing, cleaning, and transformation using suitable tools. CO3 : Apply Python/R programming for data analysis and basic machine learning tasks. CO4 : Create interactive dashboards and reports using visualization tools CO5 : Analyze datasets and derive meaningful insights for decision-making CO6 : Prepare professional reports and presentations based on job training experience.		
Sr. No.	Guidelines For On Job Training (OJT)	
1	Student must undergo for OJT/Internship during semester-VI after teaching and practical hours	
2	Student are expected to complete the related work/project within 120 hours assigned by organization (company/ industry/ consultancy/ institution)	
3	The internship work may involve the Data Analytics or IT related assignment(s) OR the maintenance of existing project OR the design/development of new project OR equivalent work	
4	College should assign the mentors/guides for students to monitor the progress throughout the OJT	
5	Students have to submit the weekly progress report duly signed by the concern authorities of organization to the assigned mentor	
6	At the end of OJT, students should prepare the documentation and submit a report to the college in prescribed format given below	
7	After completion, the final presentation and documentation will be evaluated by the examination panel as per the University norms	
	OJT Report Documentation Format	
	- Front Page - College Certificate - Company Certificate - Index	

	1) Learning objectives Every OJT comes with learning objectives, something you want to gain and acquire from that company. Hence, listing down the knowledge and skills will allow the employers/directors to know if you've accomplished your learning objectives upon finishing the OJT. 2) Company information [Company name] was founded in [year] with the vision to [details of the vision]. The company's mission is to [mission statement]. [Company name] plays an important role in the [name of the industry] industry. I chose to OJT in this organization because I want to contribute to their mission which is aligned and relevant to my career goals. 3) OJT description My role at [Company name] was to [details of your job description]. I worked directly with the [department] and [other relevant departments (if any)] in ensuring [OJT responsibilities]. 4) Position and OJT responsibilities / Assignments done 5) Learning, challenges, and skills while doing OJT 6) Overview of OJT experience During my OJT with [Company name], I was able to experience working first-hand with [the tasks you've completed], and through that, I'm able to develop my [skills you've acquired]. I found out that my [experiences] are useful in solving issues related to [problems you've solved]. Although I am considered to be less experienced with [the tasks/issues you've faced], I'm able to solve the problems with the help of my team by [describing how you apply the skills you've acquired]. 7) Outgoing consideration While I've gained many experiences and skills at the [Company name], I feel that I'm still lacking the confidence to solve [tasks you wished you were better at doing]. If I'm allowed more time to learn and participate at the [the aspects of the job you wanted to join], I would be able to polish my skills more in solving the [tasks/issues you've faced]. 8) Observation & Findings - Front end tool used - Back-end tools used - Tools used other than curriculum - Code of conduct - Employment process - Appraisal process - Areas where you need improvement - What is your learning / outcome of OJT?
--	--