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 (10+2) Science Stream or its equivalent examination

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

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	DS101T	Problem Solving and Python Programming	2	-	2	-	15	35	50
	DS102P	Lab Course on DS101T (Python Programming)	-	2	-	4	15	35	50
Subject-2	DS103T	Descriptive Statistics	2	-	2	-	15	35	50
	DS104P	Lab Course on DS103T (Descriptive Statistics)	-	2	-	4	15	35	50
Subject-3	DS105T	Computational Mathematics	2	-	2	-	15	35	50
	DS106P	Lab Course on DS105T (Computational Mathematics)	-	2	-	4	15	35	50
GE/OE	OE101DS	Office Automation/ Computer Fundamentals/ Introduction to Google Tools	2	-	2	-	15	35	50
SEC	SEC101DS	Computer Organization	2	-	2	-	15	35	50
IKS	DS101IKS	Indian Knowledge System (Generic)	2	-	2	-	15	35	50
AEC	AEC101 MAR/HIN/ ENG	MIL-I (Hindi) / MIL-I (Marathi)/ MIL-I (English)	2	-	2	-	15	35	50
VEC	VEC101 ENV	EVS-I	2	-	2	-	50	-	50
Total			**16**	**6**	**16**	**12**			**550**

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	DS151T	Advanced Python Programming	2	-	2	-	15	35	50
	DS152P	Lab Course on DS151T (Advanced Python Programming)	-	2	-	4	15	35	50
Subject-2	DS153T	Discrete Probability and Probability Distributions	2	-	2	-	15	35	50
	DS154P	Lab Course on DS153T (Discrete Probability and Probability Distributions)	-	2	-	4	15	35	50
Subject-3	DS155T	Graph Theory	2	-	2	-	15	35	50
	DS156P	Lab Course on DS155T (Graph Theory)	-	2	-	4	15	35	50
GE/OE	OE151DSP	Office Automation/ Computer Fundamentals/ Introduction to Google Tools	-	2	-	4	15	35	50
SEC	SEC151DS	Lab Course on Excel and Advanced Excel	-	2	-	4	15	35	50
AEC	AEC101 MAR/HIN/ ENG	MIL-I (Hindi) / MIL-I (Marathi)/ MIL-I (English)	2	-	2	-	15	35	50
VEC	VEC101 ENV	EVS-II	2	-	2	-	50	-	50
CC	CC151PE/ NSS/ NCC	From University Basket	2	-	2	-	50	-	50
Total			**12**	**10**	**12**	**20**			**550**

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	CE	EE	Total
Major Core	DS201MJ	Database Management System	2	-	2	-	15	35	50
	DS202MJ	Data Structure-I	2	-	2	-	15	35	50
	DS203MJP	Lab Course on DS201MJ and DS202MJ	-	2	-	4	15	35	50
VSC	DS221VSC	Foundations of Data Science	2	-	2	-	15	35	50
FP/ OJT/ CEP	DS231FP	Mini Project	-	2	-	4	15	35	50
Minor	DS241MN	Probability Distribution and Modelling	2	-	2	-	15	35	50
	DS242MNP	Lab Course on DS241MN	-	2	-	4	15	35	50
GE/OE	OE201DS	Ecommerce/ Web Design/ Digital Marketing	2	-	2	-	15	35	50
IKS	DS201IKS	Computing in Ancient India	2	-	2	-	15	35	50
AEC	AEC201ENG	Soft Skills-I	2	-	2	-	15	35	50
CC	CC201PE/ NSS/ NCC	From University Basket	-	2	-	4	50	-	50
		Total	**14**	**8**	**14**	**16**			**550**

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	CE	EE	Total
Major Core	DS251MJ	Relational Database Management System	2	-	2	-	15	35	50
	DS252MJ	Data Structure-II	2	-	2	-	15	35	50
	DS253MJP	Lab Course on DS251MJ and DS252MJ	-	2	-	4	15	35	50
VSC	DS221VSC	Data Analytics	-	2	-	4	15	35	50
FP/ OJT/ CEP	DS231FP	Mini Project	-	2	-	4	15	35	50
Minor	DS241MN	Testing of Hypothesis and Sampling Distributions	2	-	2	-	15	35	50
	DS242MNP	Lab Course on DS241MN	-	2	-	4	15	35	50
GE/OE	OE251DSP	Ecommerce/ Web Design/ Digital Marketing	-	2	-	4	15	35	50
SEC	SEC251DSP	Software Engineering	2	-	2	-	15	35	50
AEC	AEC201ENG	Soft Skills-II	2	-	2	-	15	35	50
CC	CC201PE/ NSS/ NCC	From University Basket	-	2	-	4	50	-	50
		Total	**10**	**12**	**10**	**24**			**550**

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	CE	EE	Total
Major Core	DS301MJ	NoSQL databases	4	-	4	-	30	70	100
	DS302MJ	R Programming	2	-	2	-	15	35	50
	DS303MJ	Foundations of Artificial Intelligence	2	-	2	-	15	35	50
	DS304MJP	Lab Course on DS301MJ (NoSQL databases)	-	2	-	4	15	35	50
	DS305MJP	Lab Course on DS302 (R Programming)	-	2	-	4	15	35	50
Major Elective	DS310MJ	Business Analytics	2	-	2	-	15	35	50
	DS311MJP	Lab Course	-	2	-	4	15	35	50
	OR					-			
	DS312MJ	Social Media Analytics	2	-	2	-	15	35	50
	DS313MJP	Lab Course	-	2	-	4	15	35	50
VSC	DS321VSCP	Lab Course on MATLAB	-	2	-	4	15	35	50
FP/CEP	DS331FP	Mini Project	-	2	-	4	15	35	50
Minor	DS341MN	Categorical 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		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS351MJ	Data Visualization and Modelling	4	-	4	-	30	70	100
	DS352MJ	Artificial Intelligence in Data Science	2	-	2	-	15	35	50
	DS353MJ	Data Security and Privacy	2	-	2	-	15	35	50
	DS354MJP	Lab on DS351MJT	-	2	-	4	15	35	50
	DS355MJP	Lab Course on DS352MJ	-	2	-	4	15	35	50
Major Elective	DS360MJ	HR Analytics	2	-	2	-	15	35	50
	DS361MJP	Lab Course	-	2	-	4	15	35	50
	OR								
	DS362MJ	Financial Analytics	2	-	2	-	15	35	50
	DS363MJP	Lab Course	-	2	-	4	15	35	50
OJT	DS381OJT	On Job Training	-	4	-	8	30	70	100
Minor	DS391MN	Multivariate 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

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

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS401MJ	Machine Learning	4	-	4	-	30	70	100
	DS402MJ	Basics of Cloud Computing	2	-	2	-	15	35	50
	DS403MJP	Lab Course on DS401MJ	-	2	-	4	15	35	50
	DS404MJP	Lab Course on DS402MJ	-	2	-	4	15	35	50
Major Elective	DS410MJ	Supply Chain & Logistics Analytics	2	-	2	-	15	35	50
	DS411MJP	Lab Course	-	2	-	4	15	35	50
	OR								
	DS412MJ	Healthcare Analytics	2	-	2	-	15	35	50
	DS413MJP	Lab Course	-	2	-	4	15	35	50
RP	DS431RP	Research Project	-	4	-	8	30	70	100
RM	DS441RM	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		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS401MJ	Machine Learning	4	-	4	-	30	70	100
	DS402MJ	Basics of Cloud Computing	2	-	2	-	15	35	50
	DS403MJP	Lab Course on DS401MJ	-	2	-	4	15	35	50
	DS404MJP	Lab Course on DS402MJ	-	2	-	4	15	35	50
	DS405MJ	Big Data Analytics	4	-	4	-	30	70	100
Major Elective	DS410MJ	Supply Chain & Logistics Analytics	2	-	2	-	15	35	50
	DS411MJP	Lab Course	-	2	-	4	15	35	50
	OR								
	DS412MJ	Healthcare Analytics	2	-	2	-	15	35	50
	DS413MJP	Lab Course	-	2	-	4	15	35	50
RM	DS441RM	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		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS451MJ	Data Mining and Warehousing	4	-	4	-	30	70	100
	DS452MJ	Deep Learning	2	-	2	-	15	35	50
	DS453MJP	Lab Course on DS451MJ	-	2	-	4	15	35	50
	DS454MJP	Lab Course on DS452MJ	-	2	-	4	15	35	50
Major Elective	DS460MJ	Geospatial Technology	2	-	2	-	15	35	50
	DS461MJP	Lab Course	-	2	-	4	15	35	50
	OR								
	DS462MJ	E-Commerce	2	-	2	-	15	35	50
	DS463MJP	Lab Course	-	2	-	4	15	35	50
RP	DS481RP	Research Project	-	8	-	16	60	140	200
		Total	**8**	**14**	**8**	**28**			**550**

OR

(Level 6.0) SEMESTER VIII (Honors Degree)

Course Type	Course code	Course Name	Credits		Teaching Scheme Hrs/Week		Examination Scheme and Marks		
			TH	PR	TH	PR	CE	EE	Total
Major Core	DS451MJ	Data Mining and Warehousing	4	-	4	-	30	70	100
	DS452MJ	Deep Learning	2	-	2	-	15	35	50
	DS453MJP	Lab Course on DS451MJ	-	2	-	4	15	35	50
	DS454MJP	Lab Course on DS452MJ	-	2	-	4	15	35	50
	DS455MJ	Natural Language Processing	4	-	4	-	30	70	100
Major Elective	DS456MJ	Geospatial Technology	2	-	2	-	15	35	50
	DS457MJP	Lab Course	-	2	-	4	15	35	50
	OR								
	DS458MJ	E-Commerce	2	-	2	-	15	35	50
	DS459MJP	Lab Course	-	2	-	4	15	35	50
OJT	DS481OJT	On Job Training	-	4	-	8	30	70	100
		Total	**12**	**10**					**550**

SavitribaiPhulePuneUniversity
B.Sc. Data Science (Pattern 2024)
Semester-I

DS101T : 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.

1	**Introduction to Python**	**10**	**CO1, CO2, CO3, CO6**

1.1 History, feature of Python, setting up path, working with python Interpreter, basic syntax, variable and data types, operators
1.2 Conditional statements-If, If-Else, nested if-else, Examples.
1.3 Looping-For,While,Nested loops, Examples
1.4 Control Statements-Break, Continue, Pass.
1.5 String Manipulation-Accessing String, Basic Operations, String Slices, Function and Methods, Examples.
1.6 Lists-Introduction, accessing list, operations, working with lists, function & methods.
1.7 Tuple-Introduction, accessing tuples, operations working, function & methods, Examples.
1.8 Dictionaries-Introduction,Accessingvalues in dictionaries, working with dictionaries, properties, function, Examples.
1.9 Functions-Defining a function, Calling a function, types of function, function arguments, anonymous function, global & local variable, Examples

2	**Classes, Objects and Inheritance**	**8**	**CO4**

2.1 Classes and Objects
2.1.1 Classes as User Defined Data Type
2.1.2 Objects as Instances of Classes
2.1.3 Creating Class and Objects
2.1.4 Creating Objects By Passing Values
2.1.5 Variables & Methods in a Class
2.2 Inheritance
2.2.1 Single Inheritance
2.2.2 Multilevel
2.2.3 Multiple Inheritance
2.2.4 Hybrid Inheritance
2.2.5 Hierarchical Inheritance
2.2.6 IS-A Relationship and HAS-A Relationship

3	**Modules and Packages**	**7**	**CO5**

3.1Built in Modules
3.1.1 Importing modules in python program
3.1.2 Working with Random Modules.
3.1.3 E.g. - built-ins, time, date time, calendar, sys, etc
3.2 User Defined functions
3.2.1Structure of Python Modules
3.3 Packages
3.3.1 Predefined Packages
3.3.2 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

DS102P : Lab Course on DS101T (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.

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

DS103T : 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 4^{th} 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

DS104P : Lab Course on DS103T (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, 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

DS105T : 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

DS106P: Lab Course on DS105T (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

SEC101DS: 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 convertors: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 – Moris Mano, 3rd Edition, Pearson 4. Computer Systems Organization & Architecture- John D. Carinelli Pearson publication.			

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

DS151T : 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

DS152P : Lab Course on DS151T (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)

DS153T : 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$P(AUB) = P(A) + P(B) - P(A \cap B)$. Generalization $P(AUBUC), 0 \leq P(A) \leq 1, P(A) + P(A') = 1, P(\Phi) = 0$ 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 reallife.			
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 (*pmf*) and cumulative Distribution Function (*cdf*), $F(\cdot)$ of discrete random variable, properties of*cdf*. 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 (*mgf*) and Cumulative Generating Function (*cgf*). Properties of *mgf* and*cgf*. 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, *mgf* and *cgf*. 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. 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.

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

DS154P : Lab Course on DS153T (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

DS155T : 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
DS156P : Lab Course on DS155T (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

1. To introduce students about graph, graph models, types of graph, connectivity, applications of graph theory.
2. To know how to find shortest path for different Eulerian and Hamiltonian circuit.
3. To introduce students about Trees, applications of trees, binary tree, tree traversal , spanning trees.
4. To know how to find minimum spanning trees.
5. To make students familiar with the use of all these concepts as tools in other areas of the course curriculum.
6. 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
SEC151DS : Lab Course on Excel and Advanced Excel

No. of Credits:	**Teaching Scheme**	**Examination Scheme**
02	Practical: 4 Hrs/Week	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			