॥ मा विद्या या विमुक्तये ॥

स्वामी रामानंद तीर्थ मराठवाडा विद्यापीठ, नांदेड

'ज्ञानतीर्थ', विष्णुपुरी, नांदेड – ४३१ ६०६ (महाराष्ट्र राज्य) भारत

SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY, NANDED

'Dnyanteerth', Vishnupuri, Nanded - 431 606 (Maharashtra State) INDIA

Established on 17th September, 1994, Recognized By the UGC U/s 2(f) and 12(B). NAAC Re-accredited with'B++' grade

| Fax : (02462) 215572 | **Academic-1 (BOS) Section** | website: srtmun.ac. |
| Phone: (02462)215542 | | E-mail: bos@srtmun.ac. |

विज्ञान व तंत्रज्ञान विद्याशाखे अंतर्गत राष्ट्रीय शैक्षणिक धोरण २०२० नुसार पदव्युत्तर स्तरावरील अभ्यासक्रम (Syllabus) शैक्षणिक वर्ष २०२४–२५ पासून लागू करण्याबाबत.

प रि प त्र क

या परिपत्रकान्वये सर्व संबंधितांना कळविण्यात येते की, या विद्यापीठा अंतर्गत येणा–या सर्व संलग्नित महाविद्यालये, विद्यापीठ संचलित महाविद्यालय, विद्यापीठ परिसर संकुले व उपपरिसर संकुलामध्ये शैक्षणिक वर्ष २०२३–२४ पासून पदव्युत्तर स्तरावर राष्ट्रीय शैक्षणिक धोरण –२०२० लागू करण्यात आले आहे. त्यानुसार विज्ञान व तंत्रज्ञान विद्याशाखेतील खालील अभ्यासक्रम लागू करण्याच्या दृष्टीने मा. कुलगुरू महोदयांनी मा. विद्यापरिषदेच्या मान्यतेच्या अधीन राहून मान्यता प्रदान केली आहे. त्यानुसार खालील अभ्यासक्रम शैक्षणिक वर्ष २०२४–२५ पासून लागू करण्यात येत आहे.

1. M. Sc. I year Computer Application (Affiliated college)
2. M. Sc. I year Data Science (Affiliated college)

सदरील परिपत्रक व अभ्यासक्रम प्रस्तुत विद्यापीठाच्या **www.srtmun.ac.in** या संकेतस्थळावर उपलब्ध आहेत. तरी सदरील बाब ही सर्व संबंधितांच्या निदर्शनास आणून द्यावी, ही विनंती.

'ज्ञानतीर्थ' परिसर,
विष्णुपुरी, नांदेड – ४३१ ६०६.
जा.क्र.:शै–१ /एनइपी /युजीअभ्यासक्रम /२०२४–२५/२०६
दिनांक ०६.०८.२०२४

डॉ. सरिता लोसरवार
सहा.कुलसचिव
शैक्षणिक (१–अभ्यासमंडळ) विभाग

प्रत : १) मा. आधिष्ठाता, विज्ञान व तंत्रज्ञान विद्याशाखा, प्रस्तुत विद्यापीठ.
२) मा. संचालक, परीक्षा व मुल्यमापन मंडळ, प्रस्तुत विद्यापीठ.
३) मा. प्राचार्य, सर्व संबधित संलग्नित महाविद्यालये, प्रस्तुत विद्यापीठ.
४) मा. संचालक, सर्व संकुले परिसर व उपपरिसर, प्रस्तुत विद्यापीठ
५) मा. प्राचार्य, न्यू मॉडल डिग्री कॉलेज हिंगोली.
६) सिस्टीम एक्सपर्ट, शैक्षणिक विभाग, प्रस्तुत विद्यापीठ. याना देवून कळविण्यात येते की, सदर परिपत्रक संकेतस्थळावर प्रसिध्द करण्यात यावे.

SWAMI RAMANAND TEERTH
MARATHWADA UNIVERSITY, NANDED - 431 606

(Structure and Syllabus of Two Years PG Degree Program with Multiple Entry and Exit Option)

TWO YEAR MASTERS PROGRAMME IN SCIENCE
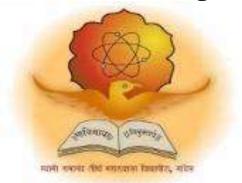
Subject M. Sc. Data Science

Under the Faculty of
Science and Technology

Effective from Academic year 2024 – 2025

(As per NEP-2020)

Swami Ramanand Teerth Marathwada University

Nanded

Affiliated Colleges

Faculty of Science and Technology

NEP-2020 Oriented Structure of Post Graduate Programs

(as per Govt of Maharashtra GR dated 16-05-2023)

M. Sc. Data Science (affiliated colleges) (2 years full time PG Programs)

Introduced from Academic Year 2024-2025

Swami Ramanand Teerth Marathwada University, Nanded

Faculty of Science and Technology
NEP-2020 oriented Structure of Two Years Post Graduate Program
Subject: M.Sc. Data Science (affiliated colleges)
(2 years' full time PG Programs in Affiliated Colleges)
Introduced from Academic Year 2024-2025 **(as per Govt. of Maha GR dated 16-05-2023)**

Program Year and Sem	Level	Semester	Major / Mandatory /		Faculty		Other courses			Total Sem. credits	Cumu. Credits
					Electives/		RM	OJT/FP/	RP		
			Theory	Practical	Theory	Practical					
			(04 credits)	(01 credits)	(04 credits) (03+01)		(03 credits)	(03 Credits)	(04 Credits)		
			SCMP		SCMP						
M.Sc. DS		First Semester	SDATSC-401 SDATSC-402 SDATSC-403	SDATSCP-401 SDATSCP-402 SDATSCP-403	SDATSE-401	---------	SVECR-401 Research Methodology Compulsory	---------	---------	22	22
M.Sc. DS	6.0	Second Semester	SDATSC-451 SDATSC-452 SDATSC-453	SDATSCP-451 SDATSCP-452 SDATSCP-453	SDATSE -451	---------	---------	SDATSOJ-451	----------	22	44
PG Diploma			24 credits + 06 Credits		06 credits +02 Credits		03 credits	03 credits	----------	44 credits	

Exit Option: After completion of First year as above with 44 credits, student will be awarded PG Diploma in Computer Science and Applications**
****(for students who have done 03 years UG program)**
****(available from AY 2024-2025)**

1. Abbreviations: **S- Science, DATS- Data Science,** Discipline **S**pecific **C**ore course (C- Core Course)
2. Abbreviations: **SDATSE- D**iscipline Supportive **E**lective **C**ourse (E- Elective Course)
3. Abbreviations: SVECR: Research Methodology course
4. Abbreviations: SDATSOJ: On Job Training, Internship/ Apprenticeship or Field Project
5. Abbreviations: SDATSR: Research Project

Syllabus First Semester

Core Courses Code	Title	Remarks Credits
SDATSC-401	Programming, Data Structures and Algorithms in Python	3
SDATSC-402	Relational Database Management System	3
SDATSC-403	Data Analysis with Excel and R	3
SDATSCP-401	Lab1-Python	2
SDATSCP-402	Lab2-RDBMS	2
SDATSCP-403	Lab3- Excel and R	2

SDATSE-401	**Chose any one**
	
	  A. Computational Statistics and Mathematics
	  B. Essentials of Business Etiquettes-Soft Skills
	  C. Data Structure
		03 Theory and 01 Lab

SVECR-401	Research Methodology	03

Syllabus Second Semester

Core Courses Code	Title	Remarks Credits
SDATSC-451	Visualization using Tableau and Power BI	04
SDATSC-452	Data Science with Python	04
SDATSC-453	Big Data Using Hadoop	04
SDATSCP-451	Lab1- Tableau and Power BI	01
SDATSCP-452	Lab2 Python	01
SDATSCP-453	Lab3-Hadoop	01

SDATSE-451	**Chose any one**
	
	  A. QA+ LR
	  B. Regression Analysis and Predictive Models
	  C. Advanced Web Technology
		03 Theory and 01 Lab

SDATSOJ-451	On Job Training , Internship/ Apprenticeship or Field Project	03

M. Sc. DS First Year, Semester I and II (*Level 6.0*): *Teaching Scheme*

	Course Code	Course Name	Credits Assigned per course			Teaching Scheme (Hrs/ week) per course	
			Theory	Practical	Total	Theory	Practical
Major	SDATSC-401 to SDATSC-403 and SDATSC-451 to SDATSC-453	All Core Course	04	--	**04**	**04**	**--**
Elective	SDATSE-401and SDATSE-451	All Elective Courses	03	--	**03**	**03**	**--**
Special Courses	SVECR-401 and SDATSOJ-451	Research Methodology and On Job Training	03	--	**03**	**03**	
Major Practical	SDATSCP-401 to SDATSCP-403 and SDATSCP-451 to SDATSCP-453	All Core labs	--	01	**01**	**--**	**02**
Elective Practical	SDATSEP-401 and SDATSEP-451	Elective lab	--	01	**01**	**--**	**02**
Total Credits per semester			**18**	**04**	**22**	**18**	**04**
Total credits per year			**36**	**08**	**44**	**36**	**08**

M. Sc. DS First Year, Semester I and II (Level 6.0): Examination Scheme

Course Code (2)	Course Name (3)	Theory				Practical		Total Col (6+7) / Col (8+9) (10)
		Continuous Assessment (CA)			ESA			
		Test I (4)	Test II (5)	Avg of (T1+T2)/2 (6)	Total (7)	CA (8)	ESA (9)	
SDATSC401 to SDATSC-403 and SDATSC-451 to SDATSC-453	All core courses	20	20	`20	80	--	--	100
SDATSE-401 and SDATSE-451	All elective courses	15	15	15	60	--	--	75
SVECR-401 and SDATSOJ-451	Research Methodology	15	15	15	60	--	--	75
SDATSCP-401 to SDATSCP-403 and SDATSCP-451 to SDATSCP-451	All Core Labs	--	--	--	--	05	20	25
SDATSEP-401 and SDATSEP-451	All Elective labs	--	--	--	--	05	20	25

Note: Teaching scheme and Examination scheme for Second year will be elaborated later, along with detailed syllabus of Second Year

Guidelines for Course Assessment:

   A. **Continuous Assessment (CA) (20% of the Maximum Marks):** This will form 20% of the Maximum Marks and will be carried out throughout the semester. It may be done by conducting **Two Tests** (Test I on 40% curriculum) and **Test II** (remaining 40% syllabus). Average of the marks scored by a student in these two tests of the theory paper will make his **CA** score (col. 6).

   B. **End Semester Assessment (80% of the Maximum Marks):** *(For illustration we have considered a paper of 04 credits, 100 marks and need to be modified depending upon credits of an individual paper)*
     1. **ESA Question paper will consists of 6 questions, each of 20 marks.**
     2. **Students are required to solve a total of 4 Questions.**
     3. **Question No.1 will be compulsory and shall be based on entire syllabus.**
     4. Students need to solve **ANY THREE** of the remaining Five Questions (Q.2 to Q.6) and shall be based on entire syllabus.

  C. **Question paper for campus PG and PG in affiliated colleges will be different**

Note: Number of lectures required to cover syllabus of a course depends on the number of credits assigned to a particular course. One credit of theory corresponds to 15 Hours lecturing and for practical course one credit corresponds to 30 Hours. For example, for a course of two credits 30 lectures of one hour duration are assigned, while that for a three credit course 45 lectures.

%%%%%

# M.Sc. Data Science
M.Sc.(DS) F. Y. (Semester I)

Course Code: SDATSC-401
Paper Title: Programming, Data Structures, and Algorithms Using Python

Course objective:
  i.   To teach the basic concepts of data structures and algorithms
  ii. To understand concepts about searching and sorting techniques
  iii. To understand basic concepts about stacks, queues, lists, trees and graphs
  iv. To understanding about writing algorithms and step by step approach in solving problems with the help of fundamental data structures

Course outcome:
  i.   Ability to analyze algorithms and algorithm correctness.
  ii. Ability to summarize searching and sorting techniques
  iii. Ability to describe stack, queue and linked list operation.
  iv. Ability to have knowledge of tree and graphs concepts

Unit I: Introduction to Python and Basic Programming Concepts
Introduction to Python: Installation, setting up the environment, and basic syntax, Variables and Data Types: Numbers, strings, lists, tuples, and dictionaries, Control Structures: Conditional statements (if-else) and loops (for, while), Functions: Defining and calling functions, scope, and recursion, Basic I/O Operations: Reading from and writing to files.

Unit II: Data Structures in Python
Lists and Tuples: Operations, slicing, and list comprehensions, Sets and Dictionaries: Operations, key-value pairs, and dictionary comprehensions, Stacks and Queues: implementation using lists and collections module, Linked Lists: Single and doubly linked lists, Trees: Basic tree concepts, binary trees, and binary search trees (BST).

Unit III: Algorithm Design and Analysis
Introduction to Algorithms: Definition, properties, and importance, Complexity Analysis: Time and space complexity, Big O notation, Searching Algorithms: Linear search, binary search. Sorting Algorithms: Bubble sort, selection sort, insertion sort, merge sort, and quicksort. Recursion and Divide-and-Conquer: Principles and examples.

Unit IV Advanced Data Structures
Heaps: Min-heap and max-heap, heap operations, Hash Tables: Hash functions, collision resolution techniques, Graphs: Representation (adjacency matrix and list), traversal algorithms (BFS, DFS), Advanced Trees: AVL trees, Red-Black trees, B-trees.

Unit V: Algorithmic Techniques and Dynamic Programming

Greedy Algorithms: Principles and examples (e.g., activity selection, Huffman coding), Dynamic Programming: Principles, memorization, and tabulation, Classic Dynamic Programming Problems: Fibonacci sequence, knapsack problem, longest common subsequence, and matrix chain multiplication. Backtracking: Principles and examples (e.g., N-queens problem, Sudoku solver).

Unit VI: Applications and Advanced Topics in Data Structures and Algorithms Using Python

Shortest Path Algorithms: Implementing Dijkstra's algorithm and Bellman-Ford algorithm in Python, Minimum Spanning Tree: Implementing Kruskal's algorithm and Prim's algorithm in Python, Network Flow: Implementing the Ford-Fulkerson method and understanding the max-flow min-cut theorem using Python.

TextBooks:

i. "Data Structures and Algorithms in Python" by Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser

ReferenceBooks:

i. "Python Algorithms: Mastering Basic Algorithms in the Python Language" by Magnus Lie Hetland
ii. "Problem Solving with Algorithms and Data Structures Using Python" by Bradley N. Miller and David L. Ranum

Course Code: SDATSCP-401

Course Title: Lab1: Programming, Data Structures, and Algorithms Using Python

1. At least 15 Practical Based on Subject.

<div align="center">

Course Code: SDATSC-402
Relational Database Management System

</div>

Learning Objectives:

   i.   To understand the features of Relational database.
  ii.   To use SQL- the standard language of relational databases for database operations.
 iii.   To understand the functional dependencies and design of the databases.

Course Outcomes:

After successful completion of this course, students should be able to:

   i.   To study the basic concepts of relational databases.
  ii.   Understand the use of Structured Query Language (SQL) and learn SQL syntax for writing queries.
 iii.   Apply normalization techniques to normalize the databases.

Unit I: Introduction and Basic Concepts of DBMS   10 Hrs.

Structure of DBMS, Advantages and Disadvantages of DBMS, Users of DBMS. Relational Database: Entities, Attributes and Domains, Tuples, Relations and their schemes.

Unit II: SQL Statements & Working with Tables   12 Hrs.

What is SQL?, Types of SQL Commands (DDL, DML, DQL, DCL, TCL), Data types in SQL, Creating Tables & Altering Table structure, Manipulation Table data, WHERE Clause tables, DISTINCT Clause, Column aliasing, Data Constraints

Unit III Operators & SQL Functions & Views   8 Hrs.

Arithmetic Operators, Relational Operators, Comparison Operators, LOGICAL Operators, SQL Functions: Single Row Functions, Multiple Row Functions, Views.

Unit IV Sorting & Grouping Data and Joining Tables & Sub queries in 8 Hrs.
ORACLE

What is sorting?, ORDER BY & GROUP BY & GROUP, BY HAVING Clauses, What is Join? Join Styles: Theta, ANSI, Using clause, Types of Joins: Equi Joins, Non-Equi Join, Outer Join: Left, Right, Full, Self-Join Cross Join, Joining three tables, Subqueries & its types

Unit V: Introduction to PL/SQL   10 Hrs.

a) PL/SQL Overview b) Declarations Section c) Executable Commands
Section d) Exception Handling Section

Unit VI Database Triggers & Cursors   12 Hrs.

a) What are Triggers? Triggers Syntax b) Types of triggers Row Level
Statement Level, Before , After Instead of Triggers c) Enabling and Disabling
Triggers Replacing and Dropping Triggers d) Working with Cursor % TYPE
Variable % ROWTYPE Variable

References:

1. "Oracle Database 10g PL/SQL Programming" by Scott Urman , Ron Hardman,
   MichaleMc Laughlin, Oracle Press, TMH, ISBN-0-07-059779-0.
2. "Oracle Database 10g The Complete Reference" By Kevin Loney, Bob Bryla Oracle Press
   (TATA McGraw Hill Edition) ISBN-13:978-0-07-059425-8, ISBN-10: 0-07- 059425-2
3. SQL, PL/SQL the programming language of ORACLE 4th Edition by Ivan Bayross
   ISBN81-7656964

Course Code: SDATSCP-402
Course Title: Lab2: Relational Database Management System

1. At least 15 Practical Based on Subject.

Course Code: SDATSC-403

Paper Title: Data Analysis with Excel and R

Course Objectives
   i.   Introduce students to the fundamentals of data analysis and familiarize them with basic Excel functionalities.
   ii. Teach students how to manipulate and analyze data using advanced Excel features and functions.
   iii. Teach students how to perform data analysis and visualization tasks using R.

Course Outcome
   i.   Understand the core principles of data analysis and its applications across various domains.
   ii. Develop a solid understanding of R programming language and RStudio environment.
   iii. Implement machine learning algorithms using R packages for predictive analysis.

Unit I: Introduction to Data Analysis and Excel Basics   **6**

Introduction to Data Analysis: Importance, process, and applications in various domains, Overview of Excel: Interface, workbook structure, and basic spreadsheet operations, Data Entry and Formatting: Entering data, applying formatting, and managing worksheets.

Unit II: Data Manipulation and Analysis in Excel   **8**

Formulas and Functions: Using built-in functions for mathematical calculations, logical operations, and text manipulation, Data Cleaning and Transformation: Sorting, filtering, removing duplicates, and handling missing values, PivotTables and PivotCharts: Creating PivotTables to summarize and analyze data dynamically.

Unit III: Statistical Analysis with Excel   **8**

Descriptive Statistics: Calculating measures of central tendency, dispersion, and distribution, Hypothesis Testing: Performing t-tests, chi-square tests, and ANOVA for inferential analysis, Regression Analysis: Using Excel's regression tools to analyze relationships between variables.

Unit IV: Introduction to R Programming   **8**

ntroduction to R: Installation, RStudio interface, and basic syntax,Data Structures: Vectors, matrices, data frames, and lists in R,Data Import and Export: Reading and writing data from/to external files (e.g., CSV, Excel).

Unit V: Data Analysis with R   **10**

Exploratory Data Analysis (EDA): Identifying patterns, trends, and outliers in data, Statistical Analysis: Calculating summary statistics, correlations, and distributions, Insights and Actionable Findings: Communicating insights and making data-driven decisions.

Unit VI: Machine Learning with R   **10**

Advanced Visualization Techniques: Animations, forecasting, and custom scripting, Integration with Other Tools: Embedding visualizations in web applications, integrating with R and Python for advanced

analytics, Best Practices and Future Trends: Optimization techniques, accessibility, and emerging trends in data visualization.

TextBooks:
1. "Data Analysis with Excel For Dummies" by Paul McFedries.
2. "R for Data Science" by Hadley Wickham and Garrett Grolemund.

ReferenceBooks:
1. "Excel 2019 Bible" by Michael Alexander, Richard Kusleika, and John Walkenbach
2. "Excel Dashboards and Reports For Dummies" by Michael Alexander and John Walkenbach
3. "Hands-On Programming with R" by Garrett Grolemund
4. "R for Everyone: Advanced Analytics and Graphics" by Jared P. Lander

Course Code: SDATSCP-403
Course Title: Lab2: Data Analysis with Excel and R

1. At least 15 Practical Based on Subject.

Course Code: SDATSE-401 A.
Computational Statistics and Mathematics

Learning Objectives:

i. To understand the features of Statistics & Mathematics.
ii. To be able to utilize technology, including computer algebra systems, to solve problems numerically, symbolically, and graphically.

Course Outcomes:

After successful completion of this course, students should be able to:

i. Students will formulate complete, concise, and correct mathematical proofs.
ii. Students will effectively use professional level **technology** tools to support the study of mathematics and statistics

Unit I: Set Theory   10 Hrs.

Meaning, Types of Set, Sub Set, Equity of Set, Operation on Set, Venn diagram, Problems on Set.

Unit II: Mathematical Logic   12 Hrs.

Proposition & Logical Operations, Truth Tables, Equivalence, Implications, Law of Logic, Predicates & Quantifier

Unit III Co-ordinate Geometry   8 Hrs.

Introduction, Co-ordinates of a point and quadrants, Distance between two points, Equation of straight line, Slope of line, Equation of circle.

Unit IV Frequency Distribution   8 Hrs.

Introduction of Statistics, Meaning of Data, Descript Variates, Continuous Variates, Formation of Frequency Distribution,

Unit V: : Measure of Central Tendency   10 Hrs.

Arithmetic Mean, Median, Mod-Definitions & Calculations, Quartile, Deciles& Percentile, Definitions & Calculations

Unit VI Probability   12 Hrs.

Definition, Random Experiment, Sample Space, Events, Definition of Probability, Examples on Probability

References:

1. Statistical Methods–S. P. Gupta 9th Edition, S. Chand Publication

2. Fundamental of Statics - S. C. Gupta, 6th Edition, Himalaya Publication.

3. Discrete Mathematical Structure- Y. N. Sing

Course Code: SDATSE-401 B.
Paper Title: Essentials of Business Etiquettes-Soft Skills

Learning Objectives:

The course aims to familiarize the students with the basics of Business Ethics and Soft Skills and its applications.

Course Outcome:

After successful completion of this course, students should be able to:

1. Study the basic concepts of Business Ethics and Soft Skills.
2. Develop a Business Ethics and Soft Skills for achieving the personal and professional goals. Acquire career skills required for getting jobs.

Unit I: Introduction- Business Ethics and Soft Skills   **10 Hrs**.
    Meaning, Definition and Importance and Nature of Business Ethics and Soft Skills
Unit II: Workplace Ethics   **10 Hrs**.
    Code of conduct for employees, Manners and Etiquettes, Integrity, Punctuality
Unit III: Soft Skills   **10 Hrs.**
    Empathy, Negotiation Skills, Interpersonal Skills, and Time Management
Unit IV: Communication Skills   **10 Hrs.**
    Meaning, Elements, Cycle, Types and Barriers to Communication
Unit V: Interview Skills   **10 Hrs.**
    Planning and Preparation of Job Interviews, Type of question for job interviews
Unit VI: Art of Writing E-mails   **10 Hrs**.
    Email Etiquettes, Email Writing Dos and Don'ts and Cover Letters

References:

1. Business Ethics by CSV Murthy
2. Soft Skills by K Alex
3. Developing Communication Skills by Krishna Mohan and Meera Banarji
4. Effective Communication Skills by Keith Koleman

Course Code: SDATSE-401 C.
Paper Title: Data Structures

Course Objectives:

i. To introduce data structures and algorithms with a focus on implementation in Python.

ii. To learn problem-solving skills through the use of algorithms and data structures, implemented in Python.

iii. Implementing and analyzing algorithms in Python to improve problem-solving efficiency.

Course Outcomes:
i. Strong foundational knowledge of data structures and algorithms.
ii. Analyze the performance of algorithms and optimize code for efficiency.
iii. Comprehensive understanding of Python programming for various applications using data structure

Unit 1 Introduction   **8Hrs.**
Introduction , Basic terminology, elementary data organization , Data structure , Data structure operation , Algorithm complexity

Unit 2 Array, Records and Pointers   **10Hrs**
Linear array, Representation of linear array in memory , Traversing linear array , Inserting and Deleting , Searching methods (Binary and linear search), Sorting Method (selection sort, bubble sort and Insertion sort)

Unit 3 Linked List   **8Hrs**
Linked list , Representation of Linked list in memory , Traversing a linked list ,Searching a linked list , Memory allocation, Garbage collection , Insertion into Linked List , Deletion from Linked List , Two way Linked List.

Unit 4 Stack   **10Hrs.**
Introduction, stack, Array Representation of stack, Linked Representation of stack, Push & pop operation, Arithmetic expression: Polish Notation, Infix, postfix & prefix notations ,Evaluation of postfix expression , Recursion :factorial, Fibonacci

Unit 5 Queue   **7Hrs.**
Introduction, Queues, Linked Representation of Queue, Insertion & Deletion on Queue. , Dqueue, Priority Queue.

Unit 6 Tree & graph   **7Hrs**
Binary Trees, Representation of Binary Tree in Memory, Types of Binary tree , Traversing of binary tree(pre-order, post-order, in-order) , Header Nodes:Threads , Graph Theory Terminology, Sequential Representation of graph , Operations on graph.

Reference Books

1. Data Structure Seymour Lipschutz Mcgraw Hill 2 An Introduction to Data Structure with Application jeanpaul, tremblay paul, g. sorenson tata Mcgraw Hill

Course Code: SDATSE-401
Course Title: Elective Lab

At least 15 Practical Based on Elective Subject.

# M.Sc. Data Science
M.Sc.(DS) F. Y. (Semester II)
Course Code: SDATSC-451

Paper Title: Visualization using Tableau and Power BI

Course Objectives
    i. Fundamental principles of data visualization and familiarize students with Tableau's basic features.

    ii. Equip students with advanced Tableau skills for creating complex visualizations and interactive dashboards.

    iii. Teach the basics of Power BI for data connection, modeling, and basic visualization creation.

Course Outcome
    i. Students will be able to create and customize basic visualizations in Tableau using various data sources.

    ii. Students will be able to connect to various data sources, perform basic data transformations, and create simple visualizations in Power BI.

Unit I: Introduction to Data Visualization and Tool Overview   6
Introduction to Data Visualization: Importance, principles, and types of visualizations, Overview of Tableau: Features, interface, and capabilities for data exploration and visualization, Overview of Power BI: Features, components, and advantages for creating interactive dashboards and reports.

Unit II: Data Connection and Preparation   8
Connecting to Data Sources: Importing data from Excel, CSV, databases, and cloud services, Data Preparation: Cleaning, transforming, and shaping data for analysis and visualization, Data Joins and Blending: Combining data from multiple sources using joins and blending techniques.

Unit III: Basic Visualization Techniques   8
Chart Types: Bar charts, line charts, scatter plots, pie charts, and heatmaps, Customizing Visualizations: Formatting axes, colors, labels, and annotations, Adding Interactivity: Filters, parameters, and actions to enhance user interaction.

Unit IV: Advanced Visualization Techniques   8
Advanced Charts and Graphs: Treemaps, box plots, histograms, and dual-axis charts, Dashboard Design Principles: Layout, composition, and storytelling techniques, Dashboard Interactivity: Drill-downs, tooltips, and parameter actions for interactive dashboards.

Unit V: Data Analysis and Insights   **10**

      Exploratory Data Analysis (EDA): Identifying patterns, trends, and outliers in data, Statistical Analysis: Calculating summary statistics, correlations, and distributions, Insights and Actionable Findings: Communicating insights and making data-driven decisions.

Unit VI: Advanced Topics and Integration   **10**

      Advanced Visualization Techniques: Animations, forecasting, and custom scripting, Integration with Other Tools: Embedding visualizations in web applications, integrating with R and Python for advanced analytics, Best Practices and Future Trends: Optimization techniques, accessibility, and emerging trends in data visualization.

Text Books:

1. "Tableau 10 Complete Reference" by Joshua N. Milligan
2. "Introducing Microsoft Power BI" by Alberto Ferrari and Marco Russo

Reference Books:

1. "Learning Tableau" by Joshua N. Milligan.
2. "The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios" by Steve Wexler, Jeffrey Shaffer, and Andy Cotgreave.
3. "Collect, Combine, and Transform Data Using Power Query in Excel and Power BI" by Gil Raviv.
4. "The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel" by Marco Russo and Alberto Ferrari.

<div align="center">

Course Code: SDATSCP-451
Course Title: Lab 1: Visualization using Tableau and Power BI
</div>

1. At least 15 Practical Based on Subject.

Course Code: SDATSC-452
Paper Title: Data Science with Python

Course Objectives:
  i.   Introduce fundamental data science concepts and Python programming.
  ii. To know about data cleaning, transformation, and exploratory data analysis techniques.
  iii. Introduce machine learning concepts and techniques using Python.

Course Outcomes:
  i.   Students will understand the basics of data science and be able to write and run
       Python scripts for basic data manipulation.
  ii. Students will be able to clean, transform, and visualize data using pandas, Matplotlib,
       and Seaborn.
  iii. Students will be able to implement and evaluate basic machine learning models using
       scikit-learn.

Unit I: Introduction to Data Science and Python Basics   **8 Hrs.**
  Introduction to Data Science, Definition, scope, and applications, Python Basics: Variables, data
  types, control flow, and functions, Introduction to NumPy and Pandas for data manipulation.

Unit II: Data Wrangling and Visualization
  Data Cleaning and Preprocessing: Handling missing values, outliers, and duplicates, Data
  Transformation, Feature engineering, scaling, and encoding, Data Visualization: Introduction to
  Matplotlib and Seaborn for data visualization.

Unit III: Applied Machine Learning   **10 Hrs.**
  Supervised Learning: Regression, classification, and model evaluation, Unsupervised Learning:
  Supervised Learning: Regression: linear regression, evaluation metrics (MSE, R²), Classification: logistic
  regression, decision trees, evaluation metrics (accuracy, precision, recall, F1 score).Clustering and
  dimensionality reduction, Clustering: k-means clustering, hierarchical clustering, Dimensionality
  reduction: PCA (Principal Component Analysis). Introduction to scikit-learn for machine learning
  tasks.

Unit IV: Advanced Machine Learning Techniques   **7 Hrs.**
  Ensemble Learning: Bagging, boosting, and stacking,Support Vector Machines (SVM) for
  classification, Introduction to neural networks and deep learning with TensorFlow/Keras.

Unit V: End-to-End Data Science Process   **10 Hrs.**
  Problem Formulation: Understanding the business problem or research question and defining
  clear objectives, Data Collection: Gathering relevant data from various sources, including
  databases, APIs, or web scraping, Data Exploration: Exploring and understanding the dataset,
  identifying patterns, trends, and potential issues, Modeling: Selecting appropriate machine

learning algorithms or statistical techniques to build predictive or descriptive models. Evaluation: Assessing the performance of models using appropriate evaluation metrics and fine-tuning them if necessary. Deployment: Implementing the model into production or making recommendations based on the analysis.

Unit VI: Big Data and Distributed Computing **8 Hrs.**
Introduction to tools like Apache Spark for large-scale data analysis, Resilient Distributed Datasets (RDDs), Transformations and Actions, Spark SQL, Spark Streaming, Natural Language Processing (NLP) and Text Analytics, Tokenization, Stemming and Lemmatization, Sentiment Analysis, Named Entity Recognition (NER), Time Series Analysis and Forecasting.

Text Books:
1. "Data Science from Scratch: First Principles with Python" by Joel Grus.

Reference Books:
1. "Python for Data Analysis" by Wes McKinney.
2. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
3. "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython" by Wes McKinney

Course Code: SDATSCP-452
Course Title: Lab 2: Data Science with Python

1. At least 15 Practical Based on Subject.

Course Objectives:
  i.   Introduce fundamental Big Data concepts and the Hadoop ecosystem.
  ii. Understand data storage mechanisms and processing frameworks within
       the Hadoop ecosystem.
  iii. Explore advanced Hadoop topics and Big Data analytics techniques

Course Outcomes:
  i.   Students will understand the characteristics and challenges of Big Data, and be able to
       explain the core components and architecture of the Hadoop ecosystem.
  ii. Students will be able to perform data storage operations using HDFS, write and
       execute MapReduce programs, and utilize Pig and Hive for data processing.
  iii. Students will be able to set up and manage Hadoop clusters, Students will be
       able to use Sqoop and Flume for data ingestion

Unit I: Introduction to Big Data and Hadoop Ecosystem   10

Definition and characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value). Big Data applications and challenges, Hadoop Fundamentals, History and evolution of Hadoop.Hadoop Distributed File System (HDFS): architecture, design, and functioning. Introduction to MapReduce programming model, Hadoop Ecosystem Overview Key components: Hadoop YARN, HDFS, MapReduce., Introduction to Hadoop ecosystem tools: Hive, Pig, Sqoop, Flume, HBase, Zookeeper.

Unit II: Data Storage with HDFS and HBase   08

Data Storage with HDFS: HDFS file operations: reading, writing, and managing data. Data replication and fault tolerance in HDFS, Best practices for HDFS file management. NoSQL Databases with HBase, Introduction to HBase: Hadoop database for structured data. HBase architecture and data model, CRUD operations in HBase, Integration of HBase with Hadoop.

Unit III: Data Processing with MapReduce and YARN   08

Processing Data with MapReduce: Detailed MapReduce programming: Mappers, Reducers, and data flow. Advanced MapReduce concepts: combiners, partitioners, and custom input/output formats. Resource Management with YARN, Introduction to YARN (Yet Another Resource Negotiator). YARN architecture and components. Resource allocation and management in YARN.
.
Unit IV: Data Ingestion and Integration Tools   08

Data Ingestion with Apache Sqoop and Flume: Introduction to Sqoop: transferring data between Hadoop and relational databases, Introduction to Flume: collecting and transferring log data to

HDFS, Real-Time Data Processing with Apache Kafka and Storm, Introduction to Kafka: distributed streaming platform, Introduction to Storm: real-time computation system. Integration of Kafka with Storm for real-time data processing.

Unit V: Advanced Analytics with Apache Spark   08

Introduction to Apache Spark: Overview of Spark architecture and components. Spark Core: RDDs (Resilient Distributed Datasets), transformations, and actions, Advanced Spark Components, Spark SQL for structured data processing, Spark Streaming for real-time data processing, Machine learning with Spark MLlib.

Unit VI: Hadoop Administration and Security   08

Hadoop Cluster Administration: Setting up and configuring a multi-node Hadoop cluster. Monitoring and managing Hadoop clusters with Apache Ambari and Cloudera Manager, Troubleshooting and optimizing Hadoop performance, Security in Hadoop, Security challenges in Hadoop. Implementing authentication and authorization in Hadoop. Data encryption and secure communication.

TextBooks:
1.   "Big Data Analytics with Hadoop 3" by Sridhar Alla and Venkat Ankam.

ReferenceBooks:
1.   "Hadoop: The Definitive Guide" by Tom White.
2.   "Hadoop in Practice" by Alex Holmes.
3.   "Hadoop Application Architectures: Designing Real-World Big Data Applications" by Mark Grover, Ted Malaska, Jonathan Seidman, and Gwen Shapira

Course Code: SDATSCP-453
Course Title: Lab 3: Big Data Using Hadoop

At least 15 Practical Based on Subject.

Paper Title: Numerical Ability and Logical Reasoning

Course Objectives:

1. Practicing Basics of mathematics
2. Use of Numbers
3. Finding Percentage and Profit or Loss, Average
4. Finding Time, Speed, Distance
5. Use of permutation and combination and Probability

Course Outcomes:

1. Develops problem solving skills of student
2. Improves Basic and advanced calculations used in day to day life.
3. Improves Mental Alertness
4. Analytical Thinking

Unit I: Introduction of Number system and HCF and LCM   8Hrs.

Types of numbers, Divisibility tests of numbers, Arithmetic progression, Geometric progression, Relationship between Arithmetic progression and Geometric progression, Methods of calculating highest common factor and greatest common divisor, Division method, Finding HCF and LCM more than two numbers, LCM and HCF of fractions and decimal numbers

Unit II: Percentage   8Hrs.

Concept of percentage, Application of percentage, Results on populations, Result on depreciations, Theoretical problem on percentage

Unit III: Average   7Hrs.

Definition of average, Formulae, theoretical problem on average

Unit IV: Series, Analogy   7Hrs.

Types of series, Alphabet series, Alpha numeric, Completing the Analogous Pair.

Unit V: Direction Sense Test Lectures   9Hrs.

Problems based on angular changes in direction, Problems on Shadows, General Problems based on Pythagoras Theorem.

Unit VI: Coding-Decoding Lecturers   11Hrs.

Letter coding, Direct Letter Coding, Number/Symbol Coding.

Reference Books:

Sr. no.	Name of the book	Author	Publication
1.	Quantitative Aptitude	Dr.R.S Aggarwal	S.Chand and Company
2.	Quantitative Aptitude	AbijitGuha	Tata McGraw Hill Education
3.	Modern Approach to Verbal & NonVerbal Reasoning	Dr.R.S Aggarwal	S.Chand and Company
4.	Test of Reasoning	Edgar Thorpe	McGraw Hill Education

Course Code: SDATSE-451 B

Paper Title: Regression Analysis & Predictive Models

Course Objectives:

    i.   Understand the fundamentals of regression analysis, including simple linear regression.
    ii.   Learn techniques to check the adequacy of regression models.

Course Outcome

    i.   Grasp the basics of regression analysis, including the purpose, scope, and types of regression models.
    ii.   Execute and interpret simple linear regression models, including estimating parameters, hypothesis testing, and constructing confidence intervals.
    iii.   Execute and interpret simple linear regression models, including estimating parameters, hypothesis testing, and constructing confidence intervals.

Unit I: Introduction to Regression Analysis   **8 Hrs.**

Introduction to Regression Modeling, The Simple Linear Regression Model, Least Squares Estimation of the Parameters, Hypothesis Testing in Simple Linear Regression, Confidence Intervals and Prediction Intervals

Unit II: Multiple Linear Regression   **8 Hrs.**

Multiple Linear Regression Model, Least Squares Estimation in Multiple Linear Regression, Hypothesis Testing in Multiple Linear Regression, Confidence Intervals and Prediction Intervals in Multiple Regression

Unit III: Model Adequacy Checking   **6 Hrs.**

Residual Analysis, Influential Observations, Transformations to Achieve Linearity, Weighted Least Squares

Unit IV: Polynomial Regression and Interaction Models   **8 Hrs.**

Polynomial Regression Models, Interaction Between Variables, Centering and Scaling Predictors, Regression Models with Categorical Predictors

Unit V: Variable Selection and Model Building   **10 Hrs.**

Criteria for Model Selection, Stepwise Regression, All Possible Regressions, Model Validation Techniques

Unit VI: Advanced Topics in Regression Analysis   **10 Hrs.**

Generalized Linear Models (GLMs), Logistic Regression, Ridge and Lasso Regression, Random Effects Models

Text Books:
1. Introduction to Linear Regression Analysis (6th Edition) by Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining

Reference Books:
1. Introduction to Linear Regression Analysis (6th Edition) by Douglas C. Montgomery, Elizabeth A.Peck, and G.Geoffrey Vining.
2. Applied Regression Analysis (4th Edition) by Norman R. Draper and Harry Smith

Course Code: SDATSE-451 C

Paper Title: Advanced Web Technology

Course Objectives
  i.   To aware the Students with advanced web technology
  ii.   To develop a skill to write applications using PHP and Java Script
      \
Course Outcome
   Students will be able to develop a dynamic webpage by the use of PHP and java script. On completion of this course, a student will be able to develop a web application using PHP and java script.

Unit-I: Introduction   **10**
   Web Technology & XML Internet – current state, hardware and software requirement, ISP, an internet account, web home page, URL, browser, security on web, searching tools, search engines, FTP, Gopher, Telnet, emails, TFTP Web browser architecture, web page and multimedia, static dynamic and active web page, simple mail transfer protocol, simple network management protocol, hypertext transfer protocol

Unit-II: Basics of PHP   **6**
   Introduction to PHP, what does PHP do?, History of PHP, language basics, datatypes, variables, expressions and operators, flow control statements, including code, embedding PHP in web pages.

Unit-III: Functions & Strings   **6**
   Calling a function, defining a function, variable scope, function parameters, return values, variable functions, and anonymous functions. Strings: Accessing individual characters, cleaning strings, encoding and escaping, comparing strings, manipulating and searching strings, regular expressions.

Unit-IV: Arrays & Objects: **10**
   Indexed vs. associative arrays, identifying elements of an array, storing data in arrays, multidimensional arrays, extracting multiple values, converting between arrays and variables, traversing arrays, sorting. Objects: Creating an object, accessing properties and methods, declaring a class, introspection.

Unit-V: Database Structure   **10**
   Overview Introduction, connecting to and disconnecting from the server, Entering queries, Creating and using a database, Creating and selecting a database, creating a table, loading data into a table, Retrieving information from a table, selecting all data, selecting particular rows, selecting particular columns, sorting rows, date calculations, working with NULL values, pattern matching, counting rows, using more than one tables.

Unit-VI: MySQL Database   **8**
   MySQL databases in PHP: Introduction, connecting to a MySQL database, querying the database, Retrieving and displaying the results, modifying data, deleting data.

Course Code: SDATSE-451
Course Title: Elective Lab

At least 15 Practical Based on Elective Subject.

Course Code: SDATSOJ-451

Paper Title:

On Job Training/ Internship/ Apprenticeship or Field Project

Students have to complete On Job Training/Internship/ Apprenticeship or Field Project
And the report of this to be submitted