

Department of Data Science & Artificial Intelligence

# **Himachal Pradesh University**

(NAAC Accredited “A” Grade University)

**Gyan Path, Summer Hill,**

**Shimla -171005**

## **Plan, Scheme, and Syllabus**

to start a programme

### **Diploma**

**in**

### **Data Science**

(Effective from the session 2021-22)

**Department of Data Science and Artificial Intelligence**

**Under the Faculty of Physical Sciences**

**Himachal Pradesh University**

**Shimla – 5**



# Contents

	Page
1. Introduction	5
2. Program Details:	5
a) Program	
b) Duration	
c) Eligibility	
d) Mode of Admission	
e) Fee Structure	
3. Examinations	5
4. Minimum Criteria to Award Diploma	5
5. Program Outcomes (POs)	7
6. Program Specific Outcomes (PSO)	8
7. Scheme of Syllabus	9
8. Detailed Syllabus	10



## 1. Introduction

Data Science is amongst the most recent field of the 21st century that will impact all segments of daily life by 2025. The recent development in Data Science is bringing significant social and economic benefits to the world. As our daily lives are seamlessly integrating more and more data-driven applications, the role of data analytics and artificial intelligence becomes increasingly important in transforming organizations, industries, and society in general. Using Data Science techniques, digital machines can analyse and learn from big datasets and discover more efficient ways to do complex tasks; thereby, making intelligent decisions with much higher accuracy and speed than human beings. As per World Economic Forum, Data Scientists and Analysts will become the number one emerging role in the world and thus will find job opportunities which are expected to rise appreciably in the years to come. So, Academic Institutions must take timely initiatives to offer academic programs to equip aspirants with requisite job-oriented skills and training and thereby contribute to meet out the global industrial demand of workforce and diversity in and Data Science.

Diploma in Data Science proposed in this report, and to be started by Himachal Pradesh University, shall be one of such academic platforms, which caters to impart most advanced knowledge, methods, and processes to exploit data science-based solutions to real-world problems. After the completion of this course, the students may have career opportunities with exceptional prospective fields of healthcare, business, e-Commerce, social networking companies, climatology, biotechnology, genetics, and other important areas.

## 2. Program Details:

- (a). Program: Diploma in Data Science
- (b). Duration: One(01) Year Program divided into two (02) Semesters
- (c). Eligibility: Graduation in any stream with 50% aggregate marks in qualifying exam (45% for SC/ST/PWD)
- (d). Mode of Admission: Direct Admission
- (e). Fee Structure: Rs. 10,000 / Per Semester

## 3. Examinations:

As the degree is spanned over one year and distributed into two semesters, the learning outcomes shall be assessed after every semester. The assessment of the students shall consist of the following components:

Sr. No.	Assessment Component
1	Semester End External Examinations(Theory)
2	Internal Assessment(Theory)
3	Semester End External Examination (Practical)
4	Internal Assessment (Practical)

## 4. Minimum Criteria to Award the Diploma

Minimum Criteria to Award the Diploma is as per Himachal Pradesh University Norms.

**Diploma**  
**in**  
**Data Science**

**Scheme**  
**&**  
**Syllabus**

**Credit Based System**

**Effective from**

**Academic Session 2021-2022**

**Department of Data Science & Artificial Intelligence**

<b>PROGRAMME OUTCOMES (POs)</b>	
<b>PO1</b>	<b>Computational Knowledge:</b> Apply knowledge of computing fundamentals, computing specialization, mathematics, and domain knowledge appropriate for the computing specialization to the abstraction and conceptualization of computing models from defined problems and requirements.
<b>PO2</b>	<b>Problem Analysis:</b> Identify, formulate, research literature, and solve computing problem searching substantiated conclusions using fundamental principles of mathematics, computing sciences, and relevant domain disciplines.
<b>PO3</b>	<b>Design/Development of Solutions:</b> Design and evaluate solutions for computing problems, and design and evaluate systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.
<b>PO4</b>	<b>Conduct investigations of Computing Problems:</b> User search-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of the information to provide valid conclusions
<b>PO5</b>	<b>Modern Tool Usage:</b> Create, select, adapt, and apply appropriate techniques, resources, and modern computing tools to computing activities, with an understanding of the limitations.
<b>PO6</b>	<b>Innovation and Entrepreneurship:</b> Identify a timely opportunity and using innovation to pursue that opportunity to create value and wealth for the betterment of the individual and society at large
<b>PO7</b>	<b>Life-long Learning:</b> Recognize the need, and have the ability, to engage in independent learning for continual development as a computing professional
<b>PO8</b>	<b>Project Management and Finance:</b> Demonstrate knowledge and understanding of the computing and management principles and apply these to one's own work, as a member and leader in a team to manage projects and in multidisciplinary environments
<b>PO9</b>	<b>Communication Efficacy:</b> Communicate effectively with the computing community, and with society at large, about computing activities by being able to comprehend and write effective reports, design documentation, make effective presentations, and give and understand clear instructions
<b>PO10</b>	<b>Societal and Environmental Concern:</b> Understand and assess societal, environmental, health, safety, legal, and cultural issues within local and global contexts, and the consequential responsibilities relevant to professional computing practices
<b>PO11</b>	<b>Individual and Teamwork:</b> Function effectively as an individual and as a member or leader in diverse teams and in multidisciplinary environments.

**PROGRAMME SPECIFIC OUTCOMES (PSOs)**

On completion of the program students will be able to:

<b>PSO1</b>	Apply the concepts and practical knowledge in analysis, design and development of computing systems and applications to multi-disciplinary problems.
<b>PSO2</b>	Provide a concrete foundation and enrich their abilities to qualify for Employment, Higher studies, and Research in Data science with ethical values
<b>PSO3</b>	Acquire ability to use current techniques, skills, and tools in Data Science.

**Abbreviations Used:**

L	LECTURES
P	PRACTICALS
C	CREDITS
H	HOURS
EE	EXTERNAL EXAMINATIONS
IA	INTERNAL ASSESSMENT



## Semester-I

Sr. No.	Course Code	Course Title	Contact Hrs/week			C	Semester End Marks		Total Marks
			L	P	H		EE	IA	
1.	DDS-101	Fundamentals of Computer and Data Science	4	0	4	4	75	25	100
2.	DDS-102	Fundamentals of Programming Using C	4	0	4	4	75	25	100
3.	DDS- 103	Data Structure	4	0	4	4	75	25	100
4.	DDS- 104	Database Management System	4	0	4	4	75	25	100
5.	DDS-151	Lab-I (Data Structure using C Language)	0	6	6	3	50	25	75
6.	DDS-152	Lab-II (Database Management System)	0	6	6	3	50	25	75
<b>Total</b>					<b>28</b>	<b>22</b>	<b>400</b>	<b>150</b>	<b>550</b>

## Semester-II

Sr. No.	Course Code	Course Title	Contact Hrs/week			C	Semester End Marks		Total Marks
			L	P	H		EE	IA	
1.	DDS-201	Statistics for Data Science	4	0	4	4	75	25	100
2.	DDS-202	Data Warehousing and Mining	4	0	4	4	75	25	100
3.	DDS-203	Python for Data Analytics	4	0	4	4	75	25	100
4.	DDS-204	R for Data Analytics	4	0	4	4	75	25	100
5.	DDS-251	Lab-III (Python for Data Analytics)	0	6	6	3	50	25	75
6.	DDS-252	Lab-IV(R for Data Analytics)	0	6	6	3	50	25	75
7.	DDS-253	Project Work				6	100	50	150
<b>Total</b>					<b>28</b>	<b>28</b>	<b>500</b>	<b>200</b>	<b>700</b>

**Total Credits of the Programme are 22+28 = 50.**

# **Semester-I**

## DDS 101: Fundamentals of Computers and Data Science

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40 (i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
--	--

### Course Outcomes:

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

- CO1:** Understand the Working of Computer.
- CO2:** Demonstrate the concept of input and output devices of Computers.
- CO3:** Acquire the basic knowledge of data and various data science approaches.
- CO4:** Acquire the basic knowledge of Machine learning.

### **PART–A**

**Introduction:** Characteristics of Computers, Evolution of computers, Capabilities and limitations of computers, Generations of computers, Types of computers (micro, mini, main frame, supercomputers), Block diagram of computer, Basic components of a computer system, Input unit, output unit, Arithmetic logic Unit, Control unit, central processing unit, Instruction set, registers, processor speed, type of processors.

### **PART–B**

**Memory:** main memory organization, main memory capacity, RAM, ROM, EPROM, PROM, cache memory, PCs specifications, Hard Disk Drive (HDD), Solid State Drive (SSD), External Hard Drives, USB Drives.

Input devices: Keyboard, Pointing Devices mouse, Touch Screens, Electronic pen, Scanning Devices- Optical Scanners, OCR, OMR, Bar Code Readers, MICR, Digitizer, Electronic card reader, Image Capturing Devices-Digital Cameras.

Output devices- Monitors, Printers- Inkjet, Laser, Plotters- Drum, Flatbed, Screen image projector.

### **PART–C**

Introduction to Data Science, Evolution of Data Science, Data Science Roles, Stages in a Data Science Project, Applications of Data Science in various fields, Data Security Issues.

### **PART–D**

Machine Learning, Supervised Machine Learning, Unsupervised Machine Learning, Semi-Supervised Machine Learning, Reinforcement Machine Learning, Applications of Machine Learning.

### Text book:

1. Pradeep K. Sinha, Priti Sinha, “Computer Fundamentals”. BPB Publications.

### Reference books:

1. Rajaraman, V., “Fundamental of Computers”. Prentice Hall India, New Delhi.
2. Jojo Moolayil, “Smarter Decisions: The Intersection of IoT and Data Science, PACKT, 2016.
3. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O’Reilly, 2015.

## DDS 102: Fundamentals of Programming Using C

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

### Course Outcomes:

At the end of the course, students will be able to

- CO1:** Recollect various programming constructs to develop C programs.
- CO2:** Understand the fundamentals of C programming.
- CO3:** Choose the right data representation formats based on the requirements of the problem.
- CO4:** Implement different Operations on arrays and functions.

### **PART-A**

Programming Tools: Problem analysis, Program constructs (sequential, decision, loops), Algorithm, Flowchart, Pseudo code, Decision table, Modular programming, Top Down and Bottom up approaches, Concept of High Level Languages, Low Level Languages, Assembly Languages, Compiler, Interpreter, Type of errors.

Overview of C: General structure of C Program.

### **PART-B**

Data types, Operators and expressions: Constants and Variables, Data types, Declaring Variables, Storage Classes, Different types of expressions and their Evaluation, Conditional Expression, Assignment statement, Enumerated data type, Redefining/ Creating data types, Library functions, Type casting.

Input/Output: Unformatted and formatted I/O Functions (Character and strings I/O, Scanf (), Printf ())

### **PART-C**

Control Statements: Decision making using if, if-else, elseif and switch statements, Looping using for, while and do-while statements, Transferring Program controlling break and continue statements, Programming examples to illustrate the use of these control statements.

### **PART-D**

Array & strings: Introduction to arrays, Declaring arrays, Initializing arrays, Processing arrays, Pointers to arrays, Passing arrays as arguments to functions, Introduction to strings, Pointers to strings, Passing strings and Arrays of strings as arguments to a function, Programming examples to illustrate the use of arrays and strings. Functions: Defining a function, Local variables, return statement, invoking a Function, specifying and passing arguments to a function, Functions returning non-Integer, External, static, and register variable, block structure, initialization and recursion. Structures: Declaring a structure type, Declaring Variables of structure type,

Initializing Structures, Accessing Elements of structures, arrays of structures, nested structures.

### Text Book:

1. Yashwant Kanetkar: Letus C: BPB Publications

### Reference Books:

1. Mullis Cooper: Spirit of C: Jacob Publications
2. Kernighan B.W. & Ritchie D.M.: The C Programming Language: PHI

**Department of Data Science & Artificial Intelligence**

3. YashwantKanetkar: PointersinC:BPB
4. GotterfiedB.: ProgramminginC:TataMcGrawHill

**DDS 103: Data Structure**

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Recognize the need of different data structures and understand its characteristics.
- CO2:** Apply different data structures for given problems.
- CO3:** Demonstrate the concepts of algorithms for searching and sorting.
- CO4:** Apply appropriate algorithms and data structures for solution of real-world problems.
- CO5:** Analyze the complexity of different algorithms.

**PART-A**

Preliminaries: Concept & notation, common operation on data structures, algorithm complexity, time-space trade off between algorithm, physical & logical representation of different data structures.

Arrays: Arrays defined, representing arrays in memory, Various operation (traversal, insertion, deletion), Multidimensional arrays, Sequential allocation, Address calculation, Sparse arrays.

**PART-B**

Linked List: Definition, type (linear, circular, doubly linked, inverted), representing linked lists in memory, advantages of using linked list over arrays, various operations on Linked list (traversal, insertion, deletion).

Stacks: Definition & concepts of stack structure, Implementation of stacks, Operation on stacks (push & pop), Application of stacks (converting arithmetic expression from infix notation to polish and their subsequent evaluation), quick sort technique to sort an array, recursion).

**PART-C**

Queue: Definition & concept of queues, implementation of queue, operation on queues (insert & delete), Type of queues (circular queue, priority queue).

Trees Structures: Tree, Binary Trees, Tree Traversal Algorithms (Pre-Order, In-Order, Post-Order), Threaded Trees, Trees in various Sorting & Searching Algorithms & their Complexity (Heap Sort, Binary Search Trees).

**PART-D**

Sorting & Searching: Selection sort, Bubble sort, Merge sort, Radix sort, Quick sort, Sequential search, Linear search and their complexity.

**Text Book:**

1. Seymour Lipschutz, Data Structure, Schaum's Outlines Series.

**Reference Books:**

1. JeanPaulTremblay&PaulG.Sorenson:AnIntroductiontoDataStructureswithApplications: TataMcGrawHill.
2. AaronM.Tenenbaum, YedidiahLangsam, MosheJ.Augenstein:DataStructuresusingC:PHI

## DDS 104: Database Management Systems

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40 (i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
--	--

### Course Outcomes:

At the end of the course, students will be able to

- CO1:** Define data independence, data models for database systems, database schema and database instances.
- CO2:** Understand and use data manipulation language to query and manage a database.
- CO3:** Analyze and design a real database application.
- CO4:** Apply normalization concepts for designing a good database with integrity Constraints.

### **PART–A**

Basic Concepts, Data Modeling for a Database, Records and Files, Abstraction and Data Integration, The Three-Level Architecture Proposal for DBMS, Components of a DBMS, Advantages and Disadvantages of a DBMS.

Data Models, Data Associations, Data Models Classification, Entity Relationship Model, Relational Data Model, Network Data Model, Hierarchical Model.

### **PART–B**

The Relational Model, Relational Database, Relational Algebra, Relational Calculus.

Relational Database Manipulation, SQL, Data Manipulation, Basic Data Retrieval, Condition Specification, Arithmetic and Aggregate Operators, SQL Join: Multiple Tables Queries, Set Manipulation, Categorization, Updates, Views: SQL, QUEL, Data Definition, Data Manipulation; QUEL, Condition Specification, Renaming, Arithmetic Operators, Multiple Variable Queries, Aggregation Operators in QUEL, Retrieve into Temporary Relation, Updates, Views.

### **PART–C**

Relational Database Design, Relational Scheme and Relational Design, Anomalies in a Database: A Consequence of Bad Design, Universal Relation, Functional Dependency, Relational Database Design, Normalization.

### **PART–D**

Concurrency Management, Serializability, Concurrency Control, Locking Scheme, Timestamp-Based Order, Optimistic Scheduling, Multiversion Techniques, Deadlock and Its Resolution.

Database Security, Integrity, and Control, Security and Integrity, Threats, Defense Mechanisms, Integrity.

### Text Book:

1. Desai, B., "An Introduction To Database Concepts." Galgotia Publications, New Delhi.

### Reference Books:

1. Date C.J., "An Introduction to Database Systems", Narosa Publishing House, New Delhi.
2. Elimsari & Navathe, "Fundamentals of Database Systems", Addison Wesley, New York.
3. Ullman, J.D., "Principals Of Database Systems", Galgotia Publications, New Delhi.

<b>DDS 151:Lab-I (Data Structure using C Language)</b>	
<b>Credits:</b> 3 <b>External Maximum Marks:</b> 50 <b>External Pass Marks:</b> 20 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%)	<b>Instructions for paper setter / candidates</b> Laboratory examination will consist of two parts: (i) Performing a practical exercise assigned by the examiner (40 marks). (ii) Viva-voce examination (10 marks)

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Recognize the need of different data structures and understand its characteristics.
- CO2:** Apply different data structures for given problems.
- CO3:** Apply appropriate algorithms and data structures for solution of real-world problems.
- CO4:** Analyze the complexity of different algorithms.

**LIST OF EXPERIMENTS**

1. Write a program to insert and delete an element at a specified location in an array.
2. Write a program to print array elements in row and column major order.
3. Write a program to search an element in an array using Linear Search.
4. Write programs to search an element in the array using Binary Search.
5. Write a menu driven program to perform various operations on strings (string length, reverse, concatenate, comparison) using user defined programs.
6. Write a program to implement stack using arrays.
7. Write a program to implement queue using arrays.
8. Write a menu driven program for matrices to do the following operation depending on whether the operation requires one or two matrices
  - Addition and Subtraction of two matrices
  - Finding upper and lower triangular matrices
  - Trace of a matrix, Transpose of a matrix, Check of matrix symmetry
9. Write a program to implement Binary search tree.
10. Write a program to perform insertion & deletion operation on Binary Search trees.
11. Write a program to create a linked list & display elements of a linked list.
12. Create a linked list and perform the following operation on it
  - a) Add a node b) Delete a node c) Count no. of nodes
13. Write a program to implement breadth first search on a graph.
14. Write a program to implement depth first search on a graph.
15. Sorting: Bubble sort, Merge sort, Insertion sort, Selection sort, Radix Sort, Quick Sort



**DDS 152: Lab-II (Database Management System)**

<b>Credits:</b> 3 <b>External Maximum Marks:</b> 50 <b>External Pass Marks:</b> 20 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%)	<b>Instructions for paper setter / candidates</b> Laboratory examination will consist of two parts: (i) Performing a practical exercise assigned by the examiner (40 marks). (ii) Viva-voce examination (10 marks)
---	---

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Create Databases for specific applications
- CO2:** Use data manipulation language to query and manage a database.
- CO3:** Apply normalization procedures in the database tables
- CO4:** Design and develop applications using database technology

**LIST OF EXPERIMENTS**

1. Familiarization with RDBMS (ORACLE/FOXPRO) using VISUAL BASIC as front end) & developing a small application.

2. Create a database and write the programs to carry out the following operation:

(i) Add a record in the database (ii) Delete a record in the database (iii) Modify the record in the database (iv) Generate queries(v) Generate the report (vi) List all the records of database in ascending order.

3. Creation, altering and dropping of tables and inserting rows into a table (use constraints while creating tables) examples using SELECT command.

Queries (along with sub Queries) using ANY, ALL, IN, EXISTS, NOTEXISTS, UNION, INTERSET, Constraints. Example:- Select the roll number and name of the student who secured fourth rank in the class.

Queries using Aggregate functions (COUNT, SUM, AVG, MAX and MIN)GROUP BY, HAVING and Creation and dropping of Views.

4. Develop a menu driven project management of database system:

(i) Library information system

- Engineering
- MCA

(ii) Inventory control system

- Computer Lab
- College Store

(iii) Student Information System

- Academic
- Finance

# Semester-II

**DDS 201: Statistics for Data Science**

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Acquire knowledge about various methods of collecting data and get familiar with some elementary methods of data viz. Measures of central tendency, skewness, and kurtosis and to interpret them.
- CO2:** Use techniques to correlate the relationship between various variables.
- CO3:** Apply probability theory to solve problems in other disciplines.
- CO4:** Acquire knowledge about continuous and Discrete random variables

**PART-A**

**Descriptive Statistics:** Sampling Techniques, Data Classification, Tabulation, Frequency and graphic Representation, Measures of Central Tendency, Measures of Variation, Quartiles and Percentiles, Moments, Skewness and Kurtosis.

**PART-B**

**Correlation and Regression:** Scatter Diagram, Karl Pearson’s Correlation Coefficient, Rank Correlation, Correlation Coefficient for Bivariate Frequency Distribution, Regression Coefficients, Fitting of Regression Lines.

**PART-C**

**Probability Theory:** Random Experiment, Sample Space, Events, Axiomatic Definition of Probability, Addition Theorem, Multiplication Theorem, Baye’s Theorem, Applications.

**PART-D**

**Distribution Function:** Continuous and Discrete Random Variables, Distribution Function of a Random Variable, Probability Mass Functions and Probability Density Functions, Characteristic Functions, Central Limit Theorems.

**Text Book:**

1. Gupta, S.C. & Kapoor, V.K., Fundamental of Mathematical statistics, Sultan Chand & Sons.

**Reference Book:**

1. Kapur, J.N. & Sarema, H.C., Mathematical Statistics, S. Chand & Company Ltd.
2. Hastie, Trevor, et al. “The elements of Statistical Learning”, Springer, 2009.
3. Ross, S.M., “Introduction to Probability and Statistics”, Academic Foundation, 2011.

**DDS 202: Data Warehousing and Data Mining**

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Identify the key processes of data mining, data warehousing and knowledge discovery process.
- CO2:** Understand the basic principles used in practical data mining with their strengths and weaknesses.
- CO3:** Analyze data and relate to real-world scenario
- CO4:** Apply data mining techniques to solve problems in other disciplines in a mathematical way.

**PART-A**

Introduction: DSS, Data warehouse Architecture, Data Staging & ETL, Multidimensional Model, Meta data, Accessing data warehouse, ROLAP, MOLAP, HOLAP, Data mart.

Data Sources: Inspecting and normalizing schemata, Integration problems, Integration phases, Mapping.

Conceptual Design: ER schema-based design, Relational schema-based design, XML schema based design, Mixed approach design.

**PART-B**

Logical Modeling & Design: MOLAP, HOLAP & ROLAP systems, Views, Temporal scenarios, Fact schemata to star schemata, View materialization, View Fragmentation, dimension tables, fact tables, Cleansing data.

Data Warehouse Components: Overall architecture, database, Sourcing, acquisition, cleanup and transformation tools, Metadata, Access tools, Administration and management, Info delivery System.

**PART-C**

Building a Data Warehouse: Considerations - business, design, technical & implementation, Integrated solutions, Benefits.

DBMS Schemas: Data layout for best access, Multidimensional data models, Star schema.

Data Tools and Metadata: Tool requirements, Access to legacy data, Transformation engines, Metadata - definition, interchange initiative, repository, trends, Reporting & Query Tools – categories.

**PART-D**

OLAP: Need, Multidimensional data model, guidelines, Multidimensional Vs multirelational OLAP, Categorization of OLAP tools.

Data Mining: Introduction, Data mining & business process, measuring effectiveness, Discovery Vs prediction, Overfitting, Comparing the technologies, Decision trees, Strengths and weaknesses.

Clustering: Introduction, uses, predictions and differences, Working.

**Text Books:**

1. Data Warehousing, Data Mining & OLAP, Alex Berson & Stephen J. Smith, Tata McGraw-Hill, 2009.
2. Data Warehouse Design: Modern Principles and Methodologies, Matteo Golfarelli, Stefand Rizzi, Tata McGraw-Hill, 2009.

**Reference Books:**

## **Department of Data Science & Artificial Intelligence**

1. Decision support and data warehouse systems, Efram Mallach, Tata McGraw-Hill, 2009.
2. The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems, John Wiley & Sons, 2008.

**DDS 203: Python for Data Analytics**

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Recognize and construct common programming idioms: variables, loop, branch, subroutine, strings and input/output.
- CO2:** Define and demonstrate the use of the built-in data structures 'list' and 'dictionary'.
- CO3:** Use database for various applications of data science.
- CO4:** Visualize the data in different form in data science applications

**PART-A**

**Python Concepts and Data Structures:** Interpreter, Program Execution, Statements, Expressions, Flow Controls, Functions, Numeric Data Types, Sequences, Strings, Tuples, Lists, Dictionaries.

**PART-B**

**OOP in Python:** Class Definition, Constructors, Object Creation, Inheritance, Overloading, Text Files and Binary Files, Reading and Writing.

**PART-C**

**Introduction to Pandas Data Structures:** Series, Data Frame, Essential Functionality: Dropping Entries Indexing, Selection, and Filtering- Function Application and Mapping- Sorting and Ranking.

**PART-D**

**Visualization in Python:** Plotting Graph, Controlling Graphs, Adding Text, More Graph Types, Getting and Setting Values, Patches.

**Text Book:**

1. Mark Lutz, "Programming Python", O'Reilly Media, 4th edition, 2010.
2. Joel Grus, "Data Science from scratch", O'Reilly, 2015.

**Reference Books:**

1. Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", 1st Edition, O'Reilly Media, 2016. ISBN-13: 978-1491912058
2. Aurelien Geron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems", 1st Edition, O'Reilly Media, 2017. ISBN – 13: 978-1491962299.
3. Wesley J Chun, "Core Python Applications Programming", 3rd Edition, Pearson Education India, 2015. ISBN-13: 978-9332555365
4. Miguel Grinberg, "Flask Web Development: Developing Web Applications with Python", 2nd Edition, O'Reilly Media, 2018. ISBN-13: 978-1491991732

**DDS 204: R for Data Analytics**

<b>Credits:</b> 4 <b>Examination Duration:</b> 3 Hours <b>External Maximum Marks:</b> 75 <b>External Pass Marks:</b> 30 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%) <b>Total Maximum Marks:</b> 100 <b>Total Pass Marks:</b> 40(i.e. 40%)	<b>Instruction for paper setters:</b> The question paper shall be of 3 hours duration and shall carry 75 marks. Total 9 questions of equal marks will be set, 2 questions from each part of the syllabus and Question no. 9 will be covering the entire syllabus and having minimum 5 subparts. <b>Instruction for Candidates:</b> Candidates are required to attempt five questions in all, one question from each part and question no. 9 will be compulsory.
---	--

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Understand the basics in R programming in terms of constructs, control statements, string functions
- CO2:** Write R codes to manipulate data
- CO3:** Apply the R programming for Data Visualization perspective
- CO4:** Design the solution for the real-life applications.

**PART-A**

Introduction: Introduction to R, Help Functions in R, Vectors, Vectorized Operations, Functions in R, Packages in R.

Matrices, Arrays and Lists: Matrix Operations, Adding and Deleting Rows and Columns, Higher Dimensional Arrays, Lists, General List Operations, Accessing List Components and Values, Applying functions to Lists.

**PART-B**

Data Frames: Creating Data Frames, Matrix-like Operations on a Data Frame, Merging Data Frames, Applying functions to Data Frames, Factors and Tables, Common Functions used with Factors, Working with Tables.

**PART-C**

OOP: S3 Classes, S4 Classes, Managing the Objects, Input/Output, Accessing Keyboard and Monitor, Reading and Writing Files, accessing the Internet, String Manipulation.

**PART-D**

Data Visualization: Introduction to GGPlot2, Factors, Aesthetics, Plotting with Layers, Overriding Aesthetics, Mapping vs Setting, Histograms, Density Charts, Statistical Transformation, Facets, Coordinates, Themes.

**Text Book:**

1. Norman Matloff, "The Art of R Programming: A Tour of Statistical Software Design", No Starch Press, 2011.

**Reference Books:**

1. Jared P. Lander, "R for Everyone: Advanced Analytics and Graphics", Addison-Wesley Data & Analytics Series, 2013.
2. Mark Gardener, "Beginning R – The Statistical Programming Language", Wiley, 2013
3. Robert Knell, "Introductory R: A Beginner's Guide to Data Visualisation, Statistical Analysis and Programming in R", Amazon Digital South Asia Services Inc, 2013.

**DDS 251: Lab-III (Python for Data Analytics)**

<b>Credits:</b> 2 <b>External Maximum Marks:</b> 50 <b>External Pass Marks:</b> 20 (i.e. 40%) <b>Internal Maximum Marks:</b> 25 <b>Internal Pass Marks:</b> 10 (i.e. 40%)	<b>Instructions for paper setter / candidates</b> Laboratory examination will consist of two parts: (i) Performing a practical exercise assigned by the examiner (40 marks). (ii) Viva-voce examination (10 marks)
---	---

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Write, test, and debug Python programs.
- CO2:** Implement Python programs with Objects and Classes.
- CO3:** Develop Python programs stepwise by defining functions and calling them.
- CO4:** Use database for various applications of data science.

**LIST OF EXPERIMENTS**

1. Write Program to print “HELLO PYTHON”.
2. Write a Program to get input from user and print it on screen.
3. Write a program to swap two numbers.
4. Write a Program to perform basic calculator operations.
5. Write a Program to check if string is palindrome or not.
6. Write a Program to illustrate common string operations in python.
7. Write a Program to print prime numbers.
8. Write a Program that uses ten different inbuilt Mathematical functions.
9. Write a Program to find factorial of given number.
10. Write a Program by using Tuple in Python.
11. Write a Program by using Set in Python.
12. Write a Program to explain different types of loop control statements.
13. Write a Program to show different types of functions in Python.
14. Write a Program by using string in Python .
15. Write a Program to show use of five dictionary functions.
16. Write a Program to show types of inheritance in Python.
17. Write a Program to explain method overloading and method overriding.
18. Write a Program to show Exception Handling in Python.
19. Write a program to explain User-Defined Exception.
20. Write a Program to sort the list entered by the user.
21. Write a Program to Combining and Merging Data Sets
22. Create Pandas Series and Data Frame from various inputs.



**DDS 252: Lab-IV (R for Data Analytics)**

<b>Credits: 2</b> <b>External Maximum Marks: 50</b> <b>External Pass Marks: 20</b> (i.e. 40%) <b>Internal Maximum Marks: 25</b> <b>Internal Pass Marks: 10</b> (i.e. 40%)	<b>Instructions for paper setter / candidates</b> Laboratory examination will consist of two parts: (i) Performing a practical exercise assigned by the examiner (40 marks). (ii) Viva-voce examination (10 marks)
---	---

**Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Understand the basics in R programming in terms of constructs, control statements, string functions
- CO2:** Write R codes to manipulate data
- CO3:** Apply the R programming for Data Visualization perspective
- CO4:** Design the solution for the real-life applications.

**LIST OF EXPERIMENTS**

1. Creating and displaying Data.
2. Matrix manipulations
3. Creating and manipulating a List and an Array
4. Creating a Data Frame and Matrix-like Operations on a Data Frame
5. Merging two Data Frames
6. Applying functions to Data Frames
7. Using Functions with Factors
8. Accessing the Internet
9. String Manipulations
10. Visualization Effects
11. Plotting with Layers
12. Overriding Aesthetics
13. Histograms and Density Charts
14. Simple Linear Regression – Fitting, Evaluation and Visualization
15. Multiple Linear Regression, Lasso and Ridge Regression

<b>DDS 253: Project Work</b>	
<b>Credits:</b> 8 <b>External Maximum Marks:</b> 100 <b>External Pass Marks:</b> 40 (i.e. 40%) <b>Internal Maximum Marks:</b> 100 <b>Internal Pass Marks:</b> 40 (i.e. 40%)	<b>Instructions for paper setter / candidates</b> <b>Project will be evaluated on the basis of following:</b> <ol style="list-style-type: none"><li>1. Problem Definition</li><li>2. Need &amp; Scope of the Study</li><li>3. Methodology &amp; Objectives</li><li>4. Data Analysis &amp; Findings</li><li>5. Testing &amp; Implementations</li><li>6. Suggestions and Conclusions</li><li>7. Overall Report Writing &amp; Layout</li></ol>

### **Course Outcomes:**

At the end of the course, students will be able to

- CO1:** Understand the structure and working of existing software systems.
- CO2:** Get an idea and confidence in analyzing, designing a software.
- CO3:** Acquire the necessary skills for developing a software system.
- CO4:** Prepare documents in team and enhance his written and oral communication presentations.
- CO5:** Apply the knowledge of latest trends in designing a software.
- CO6:** Handle various real life software development assignments.

### **Objective:**

The objective of the project work is to help the student develop ability to apply multi- disciplinary concepts, tools and techniques to analyze and logically approach the organizational problems.

### **Project Proposal Evaluation:**

The project proposal/Synopsis will be submitted within 2 weeks from start of semester and evaluated by the panel of teachers in the presence of student, who will give the presentation to the panel.

In case of non-approval of the Synopsis the comments/suggestions for reformulating the Synopsis will be communicated to the student. In such case, the revised Synopsis should be submitted within 7 days, which shall be evaluated on similar guidelines.

### **Project will be evaluated on the basis of following:**

After period of 1 Month of approval of Synopsis, Progress report 1 has been evaluated by Guide on the basis of following:

1. Problem Definition
2. Need & Scope of the Study
3. Methodology & Objectives
4. Data Analysis & Findings

Then after period of 2 Month Progress report 2 has been evaluated by Guide on the basis of following:

## **Department of Data Science & Artificial Intelligence**

1. Testing & Implementations
2. Suggestions and Conclusions
3. Overall Report Writing & Layout

The guidelines for project report are as follows:

The length of the report may be about 60 to 80 double spaced typed pages not exceeding approximately 18,000 words (excluding appendices and exhibits). However, rational variation on either side is permissible.

The Project Report may have the following:

- Cover Page – must have the Title of the Project, Name & logo of college / university, Name and University Roll No of the Student and the Name of the Guide, along with the designation and department.
- Detailed table of contents with page nos.
- All pages of the Project Report must be numbered as reflected in the table of contents.
- Project Proposal, properly bound in the project and not just stapled. Please note that project with stapled Proposal will not be accepted.
- Certificate of originality- duly signed by the student with dates.
- Introduction to the Project and Review of Literature along with brief details of the organization/s understudy.
- Rationale
- Statement of problem
- Objectives of the Project
- Scope of the study
- Research Methodology
  - Research Design
  - Nature and Source of data/information collected
  - Sample and Sampling method with rationale
- Details of the tools:
  - The Questionnaire and other methods used and their purpose
  - Reliability and Validity of the tools used
  - Administration of tools and techniques
  - Data collection
  - Data Handling, Statistical tools used for Data Analysis
- Data Interpretation and Findings
- Recommendations
- Summary and Conclusion
- Limitations of the Project

## Department of Data Science & Artificial Intelligence

- Direction for further research (optional)
- Reference/Bibliography
- Annexures/Appendices (Questionnaire used etc.)

Note: Research Methodology of the Project Report must have elaborate detail of all the components of the methodology.

### EVALUATION CRITERIA FOR PROJECT REPORT

Sr. No	Name	Roll No.	Date of Presentation	Problem Definition/ Need & Scope of the Study	Methodology & Objectives	Data Analysis & Findings	Testing and Implementation	Suggestions & Conclusion	Overall Report Writing & Layout	Total Earned Marks
				15	15	15	20	10	25	100