PGDBDA SYLLABUS

Syllabus

Programme Code: 1PDBDA

2024-2025

SENGAMALA THAYAAR EDUCATIONAL TRUST
WOMEN'SCOLLEGE
(AUTONOMOUS)
(Affiliated to Bharathidasan University)
(Accredited by NAAC|AnISO9001:2015 Certified Institution)
SUNDARAKKOTTAI, MANNARGUDI–614016, TAMILNADU, INDIA.

POST GRADUATE DIPLOMA IN BIGDATA ANALYTICS COURSE

LEARNINIG OUTCOMES BASED CURRICULUM FRAMEWORK (CBCS - LOCF)
(For the candidates admitted in the academic year 2024-2025)

ELIGIBILITY: Any Bachelor Degree From a Recognized University Or Equivalent With Minimum Of 50% Aggregate Marks. The Candidate is also required to Score Minimum 50% Marks in the Entrance Test.

Sem	Nature of the Course	Course Code	Course Title	Ins. Hrs /Week	L	T	P	O	Credit	Exam Hrs	Marks CIA	Marks ESE	Total Mark
I	Core Course(CC)–I	P24DCA11	Foundations of Computer Science	6	4	2	-	-	4	3	25	75	100
	Core Course (CC)–II	P24DCA12	Big Data Analytics and Machine Learning	6	4	2	-	-	4	3	25	75	100
	Core Course(CC)– III	P24DCA13	Python Programming	6	2	2	2		4	3	25	75	100
	Core Practical (CP)–I	P24DCA14P	Python programming Lab	6	2	-	2	2	3	3	25	75	100
	Core Practical(CP)– II	P24DCA15P	Big Data Analytics and Machine Learning lab	6	2	-	2	2	3	3	25	75	100
			Total	**30**	**14**	**6**	**6**	**4**	**18**	**-**	**-**	**-**	**500**
II	Core Course (CC)– IV	P24DCA26	Programming Using R	6	4	2		-	4	3	25	75	100
	Core Course(CC)–V	P24DCA27	Statistics	6	4	2		-	4	3	25	75	100
	Core Practical (CP)– III	P24DCA28P	R Programming Lab	6	2	-	-	4	3	3	25	75	100
	Core Practical (CP)– IV	P24DCA29P	Data Visualization Lab	6	2	-	-	4	3	3	25	75	100
	Project	P24DCAPW	Project	6	2	-	-	4	4	3	25	75	100
			Total	**30**	**14**	**4**	**-**	**12**	**18**	**-**	**-**	**-**	**500**
			Grand Total	**60**	**28**	**10**	**-**	**16**	**36**				**1000**

Subject	No. of Courses	Total Credits
Core Course	5	20
Core Practical	4	12
Project	1	4
Total	**10**	**36**

Note:

	CIA	ESE
1.Theory	25	75
2.Practical	25	75

3. Separate passing minimum is prescribed for Internal and External marks

FOR THEORY

1. The passing minimum for CIA shall be 40% out of 25mark s[i.e.10marks]
2. The passing minimum for ESE shall be 40% out of 75marks[i.e.30marks]
3. The passing minimum not less than50% in the aggregate

FOR PRACTICAL

1. The passing minimum for CIA shall be 40% out of 25 marks[i.e.16 marks]
2. The passing minimum for ESE shall be40% out of 75 marks[i.e.24marks]

The passing minimum not less than 50% in the aggregate

QUESTION PAPER PATTERN
POST GRADUATE DIPLOMA IN BIGDATA ANALYTICS
(For the candidates admitted in the academic year 2024–2025)

Max. Time: 3 Hours   **Max. Marks: 75**

SECTION – A (10 x 2 =10 Marks)
Answer ALL questions in Two Sentences Each

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

* Two questions from each unit

SECTION – B (5 x 5 = 25 Marks)
Answer ALL the Questions in Short

11. a.
     (OR)
     b
12. a.
     (OR)
     b.
13. a.
     (OR)
     b.
14. a.
     (OR)
     b.
15. a.
     (OR)
     b.

* One question from each unit, with Internal choice

SECTION – C (3 x 10 = 30 Marks)
Answer any THREE questions in Detail

16.
17.
18.
19.
20.

*One question from each unit

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: I-CC-I: Foundations of Computer Science

Ins. Hrs./Week: 6   **Course Credit:** 4   **Course Code:P24DCA11**

UNIT-I: Introduction to Information Technology   **(18   Hours)**
Basic concepts of IT, Data Processing- Data and Information. Introduction to Computers- History, Generation, Types of Computers, Block Diagram of the Computer System,

UNIT- II: Components of Computer   **(18   Hours)**
Hardware- CPU, Memory, Input and Output devices, Auxiliary storage devices. Software- System and Application Software, Utility packages, Translators, Number system and its conversion.

UNIT-III: Operating System and Office Automation Tools   **(18   Hours)**
Operating Systems overview, MS DOS, Windows, and UNIX, Programming Languages Office Automation Tools: MS Word, MS Excel, MS Power point, and MS Access

UNIT-IV: File Management and Database Management System   **(18 Hours)** File management system: Character, Field, Record and File, Types of Data Files: Master File, Transaction File, File Organization: Sequential file, Indexed Sequential file, Random files.
Database Management system: Introduction to database, DBMS architecture, Keys: Primary Key, Secondary key, Foreign key, E-R diagram and SQL.

UNIT V: Computer Networks   **(18 Hours)**
Introduction to Computer Networks, Types of Networks, Topologies, client-server model, peer-peer network, OSI Layer, TCP/IP Layer, and Network protocols, Internet and World Wide Web, Electronic Mail, HTML.

                                              Total Lecture Hours- 90

COURSE OUTCOMES
After the completion of this course Students will be able to
1. Understand the fundamental of computers
2. Learn the basic of computer components
3. Explore the operating system and office automation tools
4. Introduce about database management system
5. Acquire knowledge of computer network and its layers

TEXT BOOK(S)
1. Donald H Sanders. 1988. Computer Today. Third Edition, Mc-Graw Hill, New Delhi.
2. Suresh K. Basandra.1995. Computers Today.Galgotia Publications Pvt. Ltd. New Delhi.
3. Microsoft Web Publishing Step by Step, Active Education. Microsoft press.
4. P Rob, C Coronel.2008. Database system concept. Cengage Learning India, New Delhi.
5. Andrew S Tanenbaum, David J Wetherall. 2010. Computer Networks. Fifth Edition, PHI, New Delhi

REFERENCE BOOK(S)
1. Raja Raman V.2003. Fundamental of Computers. Fourth Edition, Prentice Hall of India, New Delhi.
2. Trainer T., et al. 1994.Computers.Fourth Edition, McGraw Hill, New Delhi.
3. Peter Norton. 2017. Introduction to Computers. Seventh Edition, McGraw Hill Education, New Delhi.
4. B Ram.2007. Computer Fundamentals. Fourth Edition, New Age International Pvt. Ltd., Chennai.
5. Jaiswal S. 2007. Fundamental of Computer & IT. Wiley dreamtech India.

E-RESOURCES
1. https://www.vskills.in/certification/tutorial/introduction-to-networking-concepts.
2. https://www.just.edu.jo/~mqais/CIS99/PDF/Ch.01_Introduction_%20to_computers.pdf
3. https://homepage.cs.uri.edu/faculty/wolfe/book/Readings/Reading01.htm(computerfun damentals)
4. https://www.tutorialspoint.com/computer_fundamentals/index.htm
5. https://vikaspedia.in/education/digital-litercy/it-literacy-courses-in-associating-with-msup/computer-fundamentals

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: I-CC-II: Big Data Analytics and Machine Learning

Ins. Hrs./Week: 6   **Course Credit:** 4   **Course Code:** P24DCA12

UNIT- I: Introduction to Database and SQL   (18 Hours)

Introduction – Purpose of Database Systems – View of Data-Data Model-Data Base Languages –Data Storage And Querying – Database Users and administrator –Relational Databases:Structure of Relational Model – Database Schema –Relational Query Languages – Relational Operations – SQL: SQL Data Definition–Basic Structure of SQL Query – Additional basic operations –set operations –Aggregate operations –Nested Sub Queries – modification of the database.

UNIT-II: Big Data   (18 Hours)

Nuances of big data – Value – Issues – Case for Big data – Big data options Team challenge – Big data sources – Acquisition – Nuts and Bolts of Big data. Features of Big Data - Security, Compliance, auditing and protection - Evolution of Big data – Best Practices for Big data Analytics - Big data characteristics - Volume, Veracity, Velocity, Variety – Data Appliance and Integration tools – Greenplum – Informatica

UNIT- III: Big Data and Evolution   (18 Hours)

Industry Examples of Big Data– Big Data Technology: Hadoop Parallel World, old vs New Approaches, Open source Technology for Big Data Analytics, The Cloud and Bigdata, Predictive Analytics, Software as service BI, Big Data technology terms. Information Management – Business Analytics– Data Privacy and Ethics.

UNIT-IV: Machine Learning   (18 Hours)

Towards Intelligent Machines, Well posed Problems, Example of Applications in diverse fields, Data Representation, Domain Knowledge for Productive use of Machine Learning, Diversity of Data: Structured / Unstructured, Forms of Learning, Machine Learning and Data Mining, Basic Linear Algebra in Machine Learning Techniques.

UNIT- V: Supervised Learning   (18 Hours)

Learning from observation, BIAS and Variance, Computational Learning Theory, OCCAM's Razor principal and overfitting avoidance, Heuristic search in inductive learning, Estimating generalization errors, metrics for assessing regression, metrics for assessing classification Learning with Support Vector Machines: Introduction, Perceptron Algorithm.

Total Lecture Hours- 90

COURSE OUTCOME
After the completion of this course Students will be able to

1. Demonstrate the basic elements of a database management system.
2. Work with big data platform and explore the big data analytics techniques business applications
3. Understand the key issues in big data management and its associated applications in intelligent business and scientific computing
4. Understand the fundamental issues and challenges of machine learning: data, model selection, model complexity.
5. Explain the concepts of training data, hypothesis space, and the difference Between training and testing datasets.

TEXT BOOK(S)
1. Henry F. Korth, and Abraham Silberschatz,, Sudarshan "Database system Concepts", McGraw Hill, 6th Edition, 2010(Unit I)
2. Frank J Ohlhorst.2012. Big Data Analytics: Turning Big Data into Big Money. Wiley and SAS Business Series.(Unit II)
3. Michael Minelli, Michehe Chambers. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business. First Edition, AmbigaDhiraj, Wiely CIO Series, 2013.(Unit III)
4. M Gopal. 2019. Applied Machine Learning. First Edition, McGraw Hill Education, USA.(Unit IV & V)

REFERENCE BOOK(S)
1. Tom White. 2012. Hadoop: The Definitive Guide" Third Edition, O'Reilly Media, USA.
2. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos. 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw Hill Publishing.
3. Bill Franks. 2012. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics. First Edition, Wiley and SAS Business Series.
4. VenkatAnkam. 2016. Big Data Analytics. Packt publishing, USA.
5. Rajkamal, PreetiSaxena. 2019. Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning. McGraw Hill Education, New Delhi.

E-RESOURCES
1. https://www.datamation.com/big-data/big-data-technologies.html
2. https://data-flair.training/blogs/apache-hive-tutorial/
3. https://hbase.apache.org/apache_hbase_reference_guide.pdf
4. https://www.analyticsvidhya.com/resources-big-data/
5. https://bigdataanalyticsnews.com/big-data-resources/

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: I-CC-III: Python Programming

Ins. Hrs./Week: 6	**Course Credit:** 4	**Course Code: P24DCA13**

UNIT- I: Basics of Python (18 Hours)

Introduction –parts of python programing language: identifiers, variables, various operators, expressions –data types– comments – read input and print output –control flow statements: if, while – Functions: Built-in functions, user-defined functions.

UNIT-II: Strings and Lists (18 Hours)

Creating and Storing Strings – Basic String Operations – Accessing Characters in String by Index Number – String Slicing and Joining –String Methods – Formatting Strings –Lists –Creating Lists – Basic List Operations –Indexing and Slicing in Lists –Built-in Functions Used on Lists – List Methods – The del Statement.

UNIT-III: Dictionaries, Tuples and Sets (18 Hours)

Creating Dictionary – Accessing and Modifying key: value Pairs in Dictionaries – Built-in Functions Used on Dictionaries – Dictionary Methods – The del Statement – Creating Tuples – Basic Tuple Operations – Indexing and Slicing in Tuples – Built-in Functions Used on Tuples – Relation between Tuples and Lists – Relation between Tuples and Dictionaries – Tuple Methods – Using zip() Function –Sets, Set Methods – Traversing of Sets – Frozen set.

UNIT- IV: Files and Regular Expressions (18 Hours)

Types of Files –Creating and Reading Text Data – File Methods to Read and Write Data – Reading and Writing Binary Files – The Pickle Module – Reading and Writing CSV Files – Python OS and OS– path Modules – Regular Expression Operations – Using Special Characters – Regular Expression Methods – Named Groups in Python Regular Expressions – Regular Expression with glob Module.

UNIT- V: Object-Oriented Programming in Python (18 Hours)

Classes and Objects – Creating Classes in Python – Creating Objects in Python – The Constructor Method –Classes with Multiple Objects –Class Attributes versus Data Attributes – Encapsulation – Inheritance – The Polymorphism.

Total Lecture Hours- 90

COURSE OUTCOME

After the completion of this course Students will be able to

1. Interpret the fundamental Python syntax and semantics and be fluent in the use of Python control flow statements
2. Express proficiency in the handling of strings and functions
3. Determine the methods to create and manipulate Python programs by utilizing the data structures like lists, dictionaries, tuples and sets
4. Identify the commonly used operations involving file systems and regular expressions
5. Articulate the Object-Oriented Programming concepts such as encapsulation, inheritance and polymorphism as used in Python

TEXT BOOK(S)

1. Gowrishankar S, VeenaA. 2018.Introduction to Python Programming. First Edition, CRC Press, Florida, USA.
2. Wesley J Chun. 2001. Core Python Programming. First Indian Reprint, Addison Wesley Longman publishing Co., USA.

REFERENCE BOOK(S)

1. Allen B Downey.2012. Think Python: How to Think like a Computer Scientist. Second edition, O'Reilly Media, Newton, USA.
2. Aurelien Geron. 2019. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Second Edition, O'Reilly Media, Newton, USA.
3. Jake VanderPlas. 2016. Python Data Science Handbook: Essential Tools for Working with Data. First Edition, O'Reilly Media, Newton, USA.
4. Miguel Grinberg. 2018. Flask Web Development: Developing Web Applications with Python, Second Edition, O'Reilly Media, Newton, USA.
5. Wesley J Chun. 2015. Core Python Applications Programming. Third Edition, Pearson Education, USA.

E-RESOURCES

1. https://courses.minia.edu.eg/Attach/160011- %20Gowrishankar%20S.,%20Veena%20A.%20-%20Introduction%20to%20Python%20Programming%20(2019,%20CRC).pdf
2. https://greenteapress.com/thinkpython2/thinkpython2.pdf
3. http://index-of.es/Python/Core.Python.Applications.Programming.3rd.Edition.pdf
4. https://files.meetup.com/18552511/Learn%20Python%20The%20Hard%20Way%203rd%20Edition%20V413HAV.pdf
5. http://index-of.es/Varios-2/Fluent%20Python%20Clear%20Concise%20and%20Effective%20Programming.pdf

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: I-CP-I : Python Programming Lab

Ins. Hrs./Week: 6 **Course Credit:** 3 **Course Code: P24DCA14P**

LIST OF PROGRAMS:

1. Write a program to demonstrate different number data types in Python.
2. Write a program to perform different Arithmetic Operations on numbers in Python.
3. Input an array of n numbers and find separately the sum of positive numbers and negative numbers.
4. Write a program to insert a number in a sorted array.
5. Write a program to create, append, and remove lists in python.
6. Write a program to demonstrate working with tuples in python.
7. Write a program to demonstrate working with dictionaries in python.
8. Write a Python Program to check whether the given string is palindrome or not using built in string manipulation methods.
9. Write a Python Program to read a word and prints the number of letters, vowels and percentage of vowels in the word using dictionary
10. Write a Python Program to check a given sentence is a pangram or not using function/Module.
11. Write a Python Event driven Program for file operations Press 1: to open file in read mode 2: open the file in write mode 3: current position of the file pointer #4: Reposition the pointer at the beginning 5:exit.
12. Write an Object oriented Python program to create two Time objects: current Time, which contains the current time; and bread Time, which contains the amount of time it takes for a bread maker to make bread. Then we'll use add Time to figure out when the bread will be done. Write the print Time function to display the time when the bread will be done by the bread maker.
13. Write a python program for accessing Hospital Information System Using MYSQL Database.

Total Lab Hours- 90

COURSE OUTCOME
After the completion of this course Students will be able to
1. Develop simple programs in Python
2. Develop programs using lists, dictionaries and files etc.,
3. Apply object oriented concepts in Python programs
4. Develop event driven programs in python
5. Develop database programs in python

REFERENCE BOOK(S)
1. Allen B Downey. 2012. Think Python: How to Think like a Computer Scientist. Second Edition, O'Reilly media, USA.
2. Chun Wesley J. 2007.Core Python Programming. Second Edition, Pearson education, London.
3. Charles Dierbach. 2013. Introduction to Computer Science using Python: A Computational Problem- Solving Focus. Wiley India Edition.
4. Timothy A. Budd. 2015. Exploring Python. Mc-Graw Hill Education (India) Private Ltd.
5. Gowrishankar S, Veena A. 2018. Introduction to Python Programming. First Edition, CRC Press, Florida, USA.

E-RESOURCES
1. https://downloads.mysql.com/docs/connector-python-en.pdf
2. http://index- of.es/Python/Core.Python.Programming.2nd.Edition.Wesley.Chun.2006.pdf
3. http://index-of.es/Python/Exploring%20Python.pdf
4. https://files.meetup.com/18552511/Learn%20Python%20The%20Hard%20Way%20 3r d%20Edition%20V413HAV.pdf
5. www.ideone.com

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE

(AUTONOMOUS)

SUNDARAKKOTTAI, MANNARGUDI- 614016.

(For the Candidates admitted in the academic year 2024– 2025)

DEPARTMENT OF COMPUTER APPLICATIONS

POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: I-CP-II: Big Data Analytics and Machine Learning Lab

Ins. Hrs./Week: 6   **Course Credit:** 3   **Course Code: P24DCA15P**

List of Programs Database:

1. Creating a Database, Viewing all Tables in a Database, Creating Tables (With and Without Constraints), Inserting/Updating/Deleting Records in a Table, Saving (Commit) and Undoing (rollback)
2. Perform the following: a. Altering a Table, Dropping/Truncating/Renaming Tables, Backing up / Restoring a Database.
3. For a given set of relation schemes, create tables and perform the following Simple Queries, Simple Queries with Aggregate functions, Queries with Aggregate functions (group by and having clause), Queries involving- Date Functions, String Functions , Math Functions Join Queries- Inner Join, Outer Join Subqueries- With IN clause, With EXISTS clause
4. For a given set of relation tables perform the following a. Creating Views (with and without check option), Dropping views, Selecting from a view

Data Analytics:
1. **Getting Started with Hadoop**
   - Introducing Hadoop
   - Installing Hadoop
   - Introducing the MapReduce Model
   - Running Hadoop Examples and Tests
   -
2. **Configure and Run Hadoop and HDFS**
   - The Design of HDFS
   - HDFS Concepts
   - Blocks Namenodes and Datanodes
   - HDFS High-Availability
   - The Command-Line Interface
   - Basic Filesystem Operations Working with files in HDFS

3. **Implement Word Count Program Using Mapreduce**
   - The building blocks of Hadoop
   - Setting up SSH for a Hadoop cluster
   - Running Hadoop
   - Web-based cluster UI

11

4. Hive
- Installing Hive
- The Hive Shell
- Running Hive
- Configuring Hive
- Hive Services
-

5. Programming with Pig
- Installing Pig
- Running Pig
- Set operations (join, union)
- Sorting with Pig
- Speaking Pig Latin
- Working with user-defined functions

Machine Learning
1. Extract the data from database using python
2. Implement k-nearest neighbors classification using python
3. Implement the finite words classification system using Back-propagation algorithm

Total Lab Hours- 90

COURSE OUTCOME

After the completion of this course Students will be able to
1. Install and configure Python, Hadoop and HDFC
2. Implement map reduce concept
3. Explore the knowledge in Hive and Pig
4. Gain knowledge in NoSQL, Zoo Keeper and Sqoop
5. Write programs using HBase

REFERENCE BOOK(S)
1. Frank J Ohlhorst.2012. Big Data Analytics: Turning Big Data into Big Money. Wiley and SAS Business Series.
2. Michael Minelli, Michehe Chambers. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business. First Edition, Ambiga Dhiraj, Wiely CIO Series, 2013.
3. Anand Rajaraman and Jeffrey David Ullman.2012. Mining of Massive Datasets. Cambridge University Press, 2012.
4. Tom White. 2012. Hadoop: The Definitive Guide" Third Edition, O'Reilly Media, USA.
5. VenkatAnkam. 2016. Big Data Analytics. Packt publishing, USA.

E-RESOURCES
1. https://www.datamation.com/big-data/big-data-technologies.html
2. https://data-flair.training/blogs/apache-hive-tutorial/
3. https://hbase.apache.org/apache_hbase_reference_guide.pdf
4. https://www.analyticsvidhya.com/resources-big-data/
5. https://bigdataanalyticsnews.com/big-data-resources/

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: II-CC-IV: Programming Using R

Ins. Hrs./Week: 6   **Course Credit:** 4   **Course Code:P24DCA26**

UNIT- I: Introduction (18 Hours)

Introduction and preliminaries – The R environment –Related software and documentation – R and statistics – R and the window system – Using R interactively – An introductory session –Getting help with functions and features – R commands – case sensitivity –Recall and correction of previous commands –Executing commands from or diverting output to a file – Data permanency and removing objects.

UNIT-II: R-Objects (18 Hours)

Data with R-Objects – Reading Data in a file – saving data – Generating Data – Manipulating Objects – Creating objects – Converting objects – Operators –Accessing the values of an object: the indexing system –Accessing the values of an object with names – The data editor – Arithmetic and simple functions – Matrix computation.

UNIT- III: Vectors (18 Hours)

Scalars, Vectors, Arrays, and Matrices – Common Vector Operations – Using all() and any() – vectorized Operations – NA and NULL Values – Filtering – A vectorized if-then-else: The if else() Function – Testing Vector Equality – Vector Element Names.

UNIT-IV: Control Statement and Recursion (18 Hours)

R Programming Structures – Control Statements –Arithmetic and Boolean Operators and Values – Default Values for Arguments – Return Values – Functions Are Objects – Environment and Scope Issues – No Pointers in R – Writing Upstairs – Recursion – Replacement Functions – Tools for Composing Function Code – Writing Your Own Binary Operations – Anonymous Functions –reading data from files.

UNIT- V: Graphics and Advanced Methods (18 Hours)

Graphics with R – Managing graphics – graphical function – high level plotting commands – low level plotting commands – Graphical parameters – A grid and lattice packages – Generalized linear models – principal components and factor analysis – advanced methods for missing data – advanced graphics.

Total Lecture Hours- 90

COURSE OUTCOME

After the completion of this course Students will be able to
1. Understand the basic of R programming
2. Read and manipulate the data
3. Understand the concepts of vectors
4. Explore the knowledge of programming constructs
5. Visualize the data for end-users

TEXT BOOK(S)
1. Norman Matloff, Nostarch. 2011. The Art Of R Programming-A Tour of Statistical Software Design. SanFrancisco.
2. Tilman Davies. The book of R: a first course in programming and statistics. No starch Press, SanFransisco, USA.
3. Venables W. N., Smith, D. M. 2018. An Introduction to R.R Core Team.

REFERENCE BOOK(S)
1. Andy Field, Jeremy Miles, Zoe Field. 2012. Discovering Statistics Using R. First Edition,SAGE, USA.
2. James (JD) Long Paul Teetor.2019. R Cook book. Second Edition, O'Reilly, USA.
3. Keith M. Reynolds, Milena Lakicevic, and Nicholas Povak, 2020. Introduction to R for Terrestrial Ecology: Basics of Numerical Analysis, Mapping, Statistical Tests and Advanced Application of R. Springer Nature, Switzer land.
4. Lander. 2014. R for Everyone: Advanced Analytics and Graphics.First Edition, Pearson Education India.
5. Pedro J. Aphalo. 2020.Learn R: As a Language. CRC Press, Florida, USA.

E-RESOURCES
1. https://intellipaat.com/blog/tutorial/r-programming/introduction/
2. https://www.dataquest.io/course/introduction-to-data-analysis-in-r/
3. https://www.listendata.com/p/r-programming-tutorials.html
4. https://www.statmethods.net/r-tutorial/index.html
5. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: II-CC-V: Statistics

Ins. Hrs./Week: 6   **Course Credit:** 4   **Course Code: P24DCA27**

UNIT- I: Introduction and Collection of Data   **(18 Hours)**
Introduction Nature and Scope of Statistics Uses and Limitations of Statistics Collection of data – Census vs Sampling: Advantages and Limitation – Primary and secondary data – Methods of collecting Primary Data – Requisites of good questionnaire – Classification and tabulation of data – sources of secondary data.

UNIT-II: Classification and Presentation of Data   **(18 Hours)**
Classification of data – Rules of Tabulation – Frequency distribution – Methods of constructing class intervals in a frequency distribution – Cumulative frequency distribution – Diagram types – Graphic representation of data – Histogram – frequency polygon, frequency – ogive curve.

UNIT-III: Univariate Analysis   **(18 Hours)**
Measure of central Tendency – Mean, Median, Mode, Geometric Mean and Harmonic mean – Relative merits and demerits – Requisites of a good average – Measures of dispersion – range, mean deviation, quartile deviation and standard deviation.

UNIT-IV: Bivariate Analysis   **(18 Hours)**
Correlation – Karl Pearson's correlation – Spearman's Rank Correlation, Simple Regression – Equations – curve fitting by methods of Latest Squares – distinction between correlation and Regression.

UNIT-V: Index Numbers, Time Series and Probability   **(18 Hours)**
Index numbers Simple aggregative method, Simple average of price relatives Fisher's ideal method – Analysis of time series – Components – straight line – semi average – moving average – Probability – Addition theorem, Multiplication theorem – Simple problems.

                                                 Total Lecture Hours- 90

    COURSE OUTCOME
After the completion of this course Students will be able to
1. Discuss key terminology,concepts tools and techniques used in Statistics.
2. Classify and tabulate the collected data and represent with suitable diagrams and graphs.
3. Evaluate and infer the facts underlying different sets of data with measures of central tendency and dispersion.
4. Apply correlations and regression techniques to evaluate the relationship between variables.
5. Measure different components 0f time series and predict the trend values.

TEXT BOOK

1. Gupta S.P.(2021)Statistical Methods, Sultan Chand Sons& amp; Co,New Delhi

REFERENCE BOOK(S)

1. H.C.Saxena,(2016), Elementary Statistics,S Chand and company, New Delhi
2. Arora P.N.,SumeetArora and AmitArora(2007),elements of Statistical Methods,sultan chand& Sons, New Delhi
3. Ken Black(2012), Applied Business Statistics,Wiley Publishers
4. Agarwal B.L.(2013), Basic Statistics, New age international Private Limited
5. Daryab Singh(latest Edition), Principles of Statistics, Atlantic Publishers limited

E-RESOURCES

1. www.mids.ac.in

2. www.usiu.libguides.com

3. https://epgp.inflibnet.ac.in/Home/

4. https://ocw.mit.edu/courses/economics/14-30-introduction-to-statistical-method in spring- 2006

5. https://en.wikipedia.org/wiki/Bivariate_analysis

SENGAMALA THAYAAR EDUCATIONAL TRUST WOMEN'S COLLEGE
(AUTONOMOUS)
SUNDARAKKOTTAI, MANNARGUDI- 614016.
(For the Candidates admitted in the academic year 2024– 2025)
DEPARTMENT OF COMPUTER APPLICATIONS
POST GRADUATE DIPLOMA IN BIG DATA ANALYTICS

Semester: II-CP-III: R Programming Lab

Ins. Hrs./Week: 6   **Course Credit:**3   **Course Code: P24DCA28P**

LIST OF PROGRAMS
1. Installing R and Packages in R
2. Program and Datatypes in R
3. Built-in function in R
4. Creating and manipulating a vector
5. Creating and Manipulating a Matrix
6. Creating and Operations on Factors
7. Operations on Data frames
8. Operations on List
9. Comparison of Matrix and Vector
10. Programs on If Else Statement
11. Programs on For Loop
12. Programs on While Loop
13. Programs on Operators
14. PLOT functions in R to customize graph
15. 3D PLOT

Total Lab Hours- 90

COURSE OUTCOME
After the completion of this course Students will be able to
1. Install packages in R
2. Write simple R programs using data types and built-in functions
3. Implement concepts of vectors, data frame and factors
4. Write programs using R constructs
5. Visualize data using graphical methods

REFERENCE BOOK(S)
1. Norman Matloff, nostarch. 2011. The Art Of R Programming-A Tour of Statistical Software Design. San Francisco.
2. Tilman Davies. The book of R: a first course in programming and statistics. No starch Press, San Fransisco, USA.
3. Venables W. N., Smith, D. M. 2018. An Introduction to R.R Core Team.
4. Andy Field, Jeremy Miles, Zoe Field. 2012. Discovering Statistics Using R. First Edition, SAGE, USA.
5. Lander. 2014. R for Everyone: Advanced Analytics and Graphics. First Edition, Pearson Education India.

E-RESOURCES
1. https://intellipaat.com/blog/tutorial/r-programming/introduction/
2. https://www.dataquest.io/course/introduction-to-data-analysis-in-r/
3. https://www.listendata.com/p/r-programming-tutorials.html
4. https://www.statmethods.net/r-tutorial/index.html
5. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Semester: II-CP-IV: Data Visualization Lab

Ins. Hrs./Week: 6   **Course Credit:** 3   **Course Code: P24DCA29P**

List of experiments

1. Introduction to various Data Visualization tools

2. Basic Visualization in Python

3. Basic Visualization in R

4. Introduction to Tableau and Installation

5. Connecting to Data and preparing data for visualization in Tableau

6. Data Aggregation and Statistical functions in Tableau

7. Data Visualizations in Tableau

8. Basic Dashboards in Tableau

                                               Total Lab Hours- 90

COURSE OUTCOME:
After the completion of this course Students will be able to

1. Handle data and data visualizations in a manner that demonstrates an understanding of ethical considerations surrounding data (including data storage, citation, and protection).
2. Understand the various analysis techniques. Design effective data visualizations in order to provide new insights into a research question.
3. Handle and prepare the co-relation and multivariate analysis and display.
4. Understand how to use data visualization in python environment.
5. Understand the various graphs and plots in python for attractive visualization.

TEXT BOOK(S):

1. Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press, 2009.
2. Beginning Python Visualization: Crafting Visual Transformation Scripts Copyright © 2009 by ShaiVaingast" Apress,ISBN-13 (electronic): 978-1-4302-1844-9
3. Data Analysis and Visualization Using Python, Dr. Ossama Embarak, ISBN: 1484241088

REFERENCE BOOKS:

1. Edward R. Tufte, "The visual display of quantitative information", Second Edition, Graphics Press, 2001.
2. Nathan Yau, "Data Points: Visualization that means something", Wiley, 2013.
3. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2008.
4. Gert H. N. Laursen and JesperThorlund, "Business Analytics for Managers: Taking business intelligence beyond reporting", Wiley, 2010.

E_RESOURCES:

1. https://www.mygreatlearning.com/blog/introduction-to-data-visualisation-why-is-it-important/
2. https://www.tableau.com/learn/articles/data-visualization
3. http://www2.cs.uh.edu/~gnawali/courses/cosc6397-f13/intro-visualization.pdf
4. https://www.geeksforgeeks.org/python-language-introduction.
5. https://beginnersbook.com/2018/01/introduction-to-python-programming/