SANT GAUGE BABA AMRAVATI UNIVERSITY, AMRAVATI

Scheme of Teaching, Learning & Examination leading to the P. G. Diploma in Data Analysis (NEP)

(One Year Two Semesters P. G. Diploma Programme)

Curriculum Scheme for First Year PG Diploma in Data Analysis SEM I & II

Sr.No.	Course Category	Subject Code No.	Subject/ Course	Credit	Teaching and Learning Scheme			Hours /Week	Examination and Valuation Scheme									
									Theory					Practical				
Semester -1					Lectures	Tutorial	Practical		Duration of Paper (Hrs)	Max Marks Theory Papers	Max Marks Sessional	Total	Min Pass Marks	Duration of Exam	External Marks	Internal Marks	Total	Mm Pass Marks Grade
1	**Pre requisite Mandatory**	**0PGDDA**	**Introduction to Computers**	**0**	**2**		**-**	**2**	**2**	**40**	**10**	**50**	**25**	**-**	**-**	**-**	**-**	
2	Core/Major	1PGDDA1	Basics of Data Analysis	4	4	-	-	4	3	70	30	100	50	-	-	-	-	-
3	Core/Major	1PGDDA2	Business Analysis using MS-Excel	4	4		-	4	3	70	30	100	50	-	-	-	-	-
4	Core/Major	1PGDDA3	Python Language	4	4	-	-	4	3	70	30	100	50	-	-	-	-	-
5	Lab	1PGDDA4	Lab-1: based on 1PGDDA2 and 1PGDDA3	2	-	-	4	4	-	-	-	-	-	3	25	25	50	25
6	Skill	1PGDDA5	Project/Seminar-I	4	-	-	8	8	-	-	-	-	-	3	50	50	100	25
7	Skill	1PGDDA6	Internship/Field Visit *	2	-	-	4	2	-	-	-	-	-			50	50	25
			TOTAL	**20**	**14**	**-**	**16**	**28**		**250**	**100**	**350**			**75**	**125**	**200**	

Semester-2																			
1	Core/Major	2PGDDA1	RDBMS and PL/SQL	4	4	-	-	4	3	70	30	100	50	-	-	-	-	-
2	Core/Major	2PGDDA2	R Programming	4	4	-	-	4	3	70	30	100	50	-	-	-	-	-
3	Core/Major	2PGDDA3	Data Analytics using R and Python	4	4	-	-	4	3	70	30	100	50	-	-	-	-	-
4	Lab	2PGDDA4	Lab-2: based on 2PGDDA2 and 2PGDDA3	2	-	-	4	4	-	-	-	-	-	3	25	25	50	25
5	Skill	2PGDDA5	Project/Seminar-II	6	-	-	12	12	-	-	-	-	-	3	100	100	200	100
			Total	**20**	**12**		**16**	**28**		**210**	**90**	**300**	**-**	**-**	**125**	**125**	**250**	**-**
Students are required to undertake mandatory internship Apprenticeship/Field Work/Work Experience (During vacation of Semester I) for duration of 60 hours, certified by authentic responsibles during vacation of Semester I. This will carry 2 Credit for learning of 60 hours. Its credits and grades will be reflected in final semester II Credit grade report.																		

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analysis Semester: I

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
0PGDDA	**Introduction to Computers**	**30**

Course Outcomes: On completion of this course, students would be able to:

CO1: Fundamentals of Computers, Organization of Computer, working of Computers
CO2: Understanding concepts of Memories
CO3: Understanding concept of Hardware, Software, Programming Languages.
CO4: Understand different types of Input / output devices
CO5: Understand Basics of Internet and Networking

Unit	Contents	Hrs
Unit I	**Basics of a Computer**: Hardware, Software, Generations of computers. Hardware - functional units, Components of CPU, Memory ï hierarchy, types of memory, Input and output devices. Software ï systems software, application software, packages, frameworks, IDEs	08
Unit II	INTRODUCTION TO INPUT -OUTPUT AND STORAGE DEVICES Input Devices - Keyboard, Mouse, Scanner, BCR, OMR, MICR, Touch screen, Voice Input etc. Output Devices - Monitor, Printer, Speakers, Plotter Storage ï Hard Disk, CD/DVD Drives, Memory Cards, Pen Drive	07
Unit III	INTERNET BASICS :- What is Internet, ISP, Browser, URL, Web Server, Domain Name, Searching on Web, E-Mail Address, Sending and Receiving Mails.	08
Unit IV	COMPUTER NETWORKING :- Definition of Computer Network, Types of Network (LAN, WAN, MAN), Network Components (Hub, Switch, Router, Repeaters, NIC), IP Address (IPV4, IPV6),VPN	07

Text books:

1) ñFundamentals of Computersò by P. K. Sinha2: A comprehensive introduction to computer science concepts.
2) ñFundamentals of Computersò by Rajaraman V and Adabala

Reference Books: format (Title, Author, Publisher, Edition)

1. ñComputer Fundamentalsò by Goel1
2. "Computer Fundamentals: Pradeep K. Sinha & Priti Sinha"
3. "Computer Fundamentals: Architecture and Organization" by B. Ram

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analysis Semester: I

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
1PGDDA1	Basics of Data Analysis	60

Course Outcomes: On completion of this course, students would be able to:

CO1: Articulate meaningful lines of inquiry that might be explored through the collection,
organization, visualization, and analysis of data

CO2: Understand what data is, how they are collected, the role of metadata in understanding a given set of data, and how to assess the quality/reliability of data.

CO3: Proficiently acquire, understand and organize the data.

CO4: Understand the strategies of data collection and pre-processing.

CO5: Demonstrate a sophisticated understanding of the concepts and methods; know the exact scopes and possible limitations of each method; and show capability of using data analytics skills to provide constructive guidance in decision making.

Unit	Contents	Hrs
Unit I	**Introduction to Data analysis**: Data and Taxonomy of Data Analytics, Big Data Architectures, History on Methodologies for Data Analytics, KDD Process, CRISP-DM Methodology, Scale Types, Descriptive Univariate Analysis, Univariate Frequencies, Univariate Data Visualization, Univariate Statistics, Multivariate Frequencies, Multivariate Data Visualization, Multivariate Statistics. Infographics	12
Unit II	**Descriptive Analysis**: Descriptive Univariate Analysis, Univariate Frequencies, Univariate Data Visualization, Univariate Statistics, Descriptive Bivariate Analysis, Two Quantitative Attributes, Two Ordinal Attributes, Multivariate Frequencies, Multivariate Data Visualization, Multivariate Statistics, Infographics.	12
Unit III	**Data Quality**: Missing , Redundant and Inconsistent Data, Noisy Data and Outliers, Converting to a Different Scale and Type, Data Transformation, Dimensionality Reduction: Attribute Aggregation, Principal Component Analysis, Independent Component Analysis Attribute Selection Search Strategies	12
Unit IV	**Regression**: Predictive Performance Estimation, Generalization, Model Validation, Predictive Performance Measures for Regression, Linear Regression, The Bias-variance Trade-off , Shrinkage Methods - Ridge and Lasso Regression, Technique and Model Selection, **Classification**: Binary Classification, K-nearest Neighbor Algorithms, Logistic Regression Algorithm, Naive Bayes Algorithm	12
Unit V	**Clustering**: Distance Measures, Clustering Validation, Clustering Techniques, K-means, DBSCAN, Agglomerative Hierarchical Clustering, Linkage Criterion, Pattern Mining: Setting the min_sup Threshold, Apriori ï a Join-based Method, Eclat, FP-Growth, Maximal and Closed Frequent Itemsets, Association Rules, Simpsonôs Paradox, Sequential patterns, Frequent Sequence Mining, Closed and Maximal Sequences.	12

Text books:

1) A General Introduction to Data Analytics by Moreira, Carvalho, Horvath Publication: Wiley & Sons Inc.

2) Microsoft Excel 365 Bible (1st Edition) by by Michael Alexander , Dick Kusleika

Reference Books: format (Title, Author, Publisher, Edition)

4. Data Analytics Made Accessible by Anil K. Maheshwari
5. Big Data Fundamentals: Concepts, Drivers & Techniques by Thomas Erl, Wajid Khattak & Paul Buhler Publication: Prentice Hall, ISBN-13: 978-0-13-429107-9 ISBN-10: 0-13-429107-7

Links of MOOC[SWAYAM]:

- https://tinyurl.com/3y2nua7a
- https://tinyurl.com/4n4e4ad7
- https://courses.mooc.fi/org/uh-cs/courses/dap-22

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analysis Semester: I

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
1PGDDA2	Business Analysis with MS-Excel	60

Course Outcomes: On completion of this course, students would be able to:

CO1: Identify the different component of Excel workbook
CO2: Perform the Data Formatting using Excel features.
CO3: Enter text and formulas into an Excel spreadsheet
CO4: Create a spreadsheet to tabulate and record numeric values
CO5: Construct formulas to manipulate numeric data in an Excel Worksheet
CO6: Set up the chart function of Excel to represent numeric data in multiple formats.

Unit	Contents	Hrs
Unit I	**Introduction to Excel:** ● Introduction ● Opening Screen of Excel 2010 ● Concept of Workbook and Spreadsheet ● Ranges in Excel ● File Menu(Tab): New, Open, Info, Save, Save As, History, Print, Share, Export, Publish, Close. ● Quick Access Toolbar	12
Unit II	**Formatting in Excel:** ● Conditional Formatting ● Sorting and Filtering ● Find and Select options ● **Page Setup Options**: Margins, Orientation, Size, Print Area, Breaks, Background, Print Titles. ● Proofing in Excel ● Protect Options in Excel: Protect Sheet, Protect Workbook, Allow Edit Ranges, Share and Unshare Workbook.	12
Unit III	**Working with Formulas and Functions**: ● Categories of Excel Functions ● Inserting and using Formulas in Excel ● Parts of Excel Formula ● Using Constants in Excel Formula ● Using References in Excel Formula ● Using calculation operators in Excel formulas ● Using functions and nested functions in Excel formulas ● Define and use names in formulas ● Delete or remove a formula	12
Unit IV	**Working with Charts**: ● Why do we use charts in Excel? ● Chart-specific terminologies: Chart Vs. Graph, X-axis, Y-axis, Legend, Plot Area. ● How can I create a chart in Excel? ● Change chart type or location. ● More on Charts: Giving Title to graph, Reorder chart data, Adjust colour and style, Switch the data on each axis, Change the size of your chart's legend and axis labels, Change the Y axis measurement options ● Types of Charts.	12
Unit V	**Working with Data**: ● Getting External Data: From Access, From Web, From other Sources. ● Queries options: New Query, Show Queries, From Table, Recent Source. ● Data Tools: Text to Columns, Flash Fill, Remove Duplicates, Data Validation, Consolidating Data, Macros	12
Text books:	**format (Title, Author, Publisher, Edition)**	

1. Microsoft Excel 365 Bible (1st Edition)by Michael Alexander, Dick Kusleika
2. Excel Dashboards and Reportsby John Walkenbach , Michael Alexander
3. Excel 2019 for Dummies (All-in-one) by Greg Harvey

MOOC Links:

1. https://www.coursera.org/projects/introduction-microsoft-excel
2. https://www.coursera.org/projects/using-basic-formulas-functions-microsoft-excel

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analytics Semester: I

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
1PGDDA3	**Python Language**	**60**

Course Outcomes: On completion of this course, students would be able to:

CO1: Understand basic principles of computers
CO2: Understand basics of binary computation
CO3: Understand the programming basics (operations, control structures, data types, etc.)
CO4: Readily use the Python programming language and Apply various data types and control structure
CO5: Understand class inheritance and polymorphism and Understand the object-oriented program design and development
CO6: Understand and begin to implement code

Unit	Contents	Hrs
Unit I	**Introduction to Python**: Python variables, Python basic Operators, Understanding python blocks. Python Data Types, Declaring and using Numeric data types: int, float etc. Python Program Flow Control Conditional blocks: if, else and else if, Simple for loops in python, For loop using ranges, string, list and dictionaries. Use of while loops in python, Loop manipulation using pass, continue, break and else. Programming using Python conditional and loop blocks.	12
Unit II	**Python Complex data types**: Using string data type and string operations, Defining list and list slicing, Use of Tuple data type. String, List and Dictionary, Manipulations Building blocks of python programs, string manipulation methods, List manipulation. Dictionary manipulation, Programming using string, list and dictionary in-built functions. Python Functions, Organizing python codes using functions.	12
Unit III	**Python File Operations**: Reading files, Writing files in python, Understanding read functions, read(), readline(), readlines(). Understanding write functions, write() and writelines() Manipulating file pointer using seek Programming, using file operations. Exception Handling in Python: try-except, Raising exception and error processing	12
Unit IV	**Object oriented design**, Programming types , Object Oriented Programming, Basics of Object Oriented Programming, Creating Class and Object, Constructors in Python: Parameterized and Non-parameterized. Inheritance in Python,In built class methods and attributes, Multi-Level and Multiple Inheritance,Method Overriding and Data Abstraction, Encapsulation and Polymorphism	12
Unit V	**Database Programming**: Connecting to a database, Creating Tables,INSERT, UPDATE, DELETE and READ operations, Transaction Control, Disconnecting from a database, Exception Handling in Databases.	12

Text books: format (Title, Author, Publisher, Edition)

1)Wesley J. Chun, ñCore Python Applications Programmingò, 3rd Edition , Pearson Education, 2016

Reference Books: format (Title, Author, Publisher, Edition)

1) Mark Lutz, ñ Learning Pythonò4 th Edition, OõReilly,2009

2)Charles Dierbach, ñIntroduction to Computer Science using Pythonò, Wiley, 2015
3)Jeeva Jose &P.SojanLal, ñIntroduction to Computing and Problem Solving with PYTHONò, Khanna Publishers, New Delhi, 2016
4)Downey, A. et al., ñHow to think like a Computer Scientist: Learning with Pythonò, John Wiley, 2015
5)Mark Lutz, ñLearning Pythonò, 5th edition, Orelly Publication, 2013, ISBN 978- 1449355739
6) John Zelle, ñPython Programming: An Introduction to Computer Scienceò, Second edition, Course Technology Cengage Learning Publications, 2013, ISBN 978- 1590282410
7)Michel Dawson, ñPython Programming for Absolute Beginersò , Third Edition, Course Technology Cengage
Learning Publications, 2013, ISBN 978-1435455009
8)David Beazley, Brian Jones., ñPython Cookbookò, Third Edition, Orelly Publication, 2013, ISBN 978-
1449340377

Sant Gadge Baba Amravati University, Amravati

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analytics Semester: I

Code of the Course/Subject	Title of the Course/Subject	(No. of Periods/Week)
1PGDDA4	Lab1 : based on 1PGDDA2 and 1PGDDA3	**06 periods per Batch**

Course Outcomes (CO's);

1. Able to Work with the Workbook of MS-Excel.
2. Able to design the Charts and work with it efficiently.
3. Able to do Data Analysis using Data Processing options in Ms-Excel
4. Able to write program in Python
5. Able to work with data in Python
6. Able to work with OOP concept in Python

*** List of Practical/Laboratory Experiments/Activities etc.**

List of Practical on Excel:

1. Create excel spreadsheet and apply editing, saving and printing operations on excel spreadsheet.
2. Prepare excel spreadsheet for Marksheet of class subject.
3. Create an employee salary sheet in excel.
4. Prepare bar chart and pie chart for analysis of five year results of the institution.
5. Work on the following exercise on a Workbook:
   a. Copy an existing Sheet
   b. Rename the old Sheet
   c. Insert a new Sheet into an existing Workbook
   d. Delete the renamed Sheet.
6. Prepare an Attendance sheet of 10 students for any 6 subjects of your syllabus. Calculate their total attendance, total percentage of attendance of each student & average of attendance.
7. Create a worksheet on Students list of any 4 faculties and perform following database functions on it.
   a. Sort data by Name
   b. Filter data by Class
   c. Subtotal of no. of students by Class
8. Prepare timetables of your college and apply securing & protecting operations to an excel spreadsheets.

List of Practical on Python:

1. Programs in Python to demonstrate conditional statement if, if-lse
2. Programs in Python to demonstrate looping in Python.
3. Programs in Python to demonstrate string manipulation functions.
4. Programs in Python to illustrate the concept of list and dictionary in Python
5. Programs in Python to illustrate the concept of functions in Python.
6. Programs in Python to demonstrate file operations in Python.
7. Programs in Python to demonstrate exception handling in Python.
8. Programs in Python to illustrate the concept of class and objects.
9. Programs in Python to illustrate the concept of constructors in Python.
10. Programs in Python demonstrate the concept of Inheritance.

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analytics Semester: II

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
2PGDDA1	**RDBMS & PL/SQL**	**60**

Course Outcomes: On completion of this course, students would be able to:

CO1: Demonstrate the basic elements of a relational database management system.
CO2: Identify the data models for relevant problems. R
CO3: Design entity relationship and convert entity relationship diagrams into RDBMS and
formulate SQL queries on the respect data into RDBMS and formulate SQL queries on the
data.
CO4: Demonstrate their understanding of key notions of query evaluation and optimization
techniques.
CO5: Extend normalization for the development of application softwareós.
CO6: Demonstrate their understanding of Transaction Control.

Unit	Contents	Hrs
Unit I	**BASIC CONCEPTS:** Database Management System - File based system - Advantages of DBMS over file based system - Database Approach - Logical DBMS Architecture - Three level architecture of DBMS or logical DBMS architecture - Need for three level architecture - Physical DBMS Architecture - Database Administrator (DBA) Functions & Role - Data files indices and Data Dictionary- Types of Database. Relational and ER Models Data Models - Relational Model Domains - Tuple andRelation - Super keys - Candidate keys - Primary keys and foreign key for the Relations - RelationalConstraints - Domain Constraint - Key Constraint - Integrity Constraint - Update Operations and Dealingwith Constraint Violations - Relational Operations - Entity Relationship (ER) Model Entities ,Attributes ,Relationships - More about Entities and Relationships - Defining Relationship for CollegeDatabase - E-R Diagram - Conversion of E-R Diagram to Relational Database.	12
Unit II	**DATABASE INTEGRITY AND NORMALISATION:** Relational Database Integrity - TheKeys - Referential Integrity - Entity Integrity - Redundancy and Associated Problems Single Valued Dependencies Normalisation - Rules of Data Normalisation - The First Normal Form -The Second Normal Form - The Third Normal Form - Boyce Codd Normal Form - Attribute Preservation ï Lossless join Decomposition - Dependency Preservation. File Organisation : Physical Database Design Issues -Storage of Database on Hard Disks - File Organisation and Its Types - Heap files (Unordered files) -Sequential File Organisation - Indexed (Indexed Sequential) File Organisation - Hashed File Organisation- Types of Indexes - Index and Tree Structure - Multi-key File Organisation - Need for Multiple AccessPaths - Multi-list File Organisation - Inverted File Organisation.	12
Unit III	**STRUCTURES QUERY LANGUAGE (**SQL) SQL ï Features of SQL ï Data types in SQL , Domains in SQL, CREATE TABLE command, Constraints ï NULL, DEFAULT,CHECK, PRIMARY KEY, UNIQUE, , referential Integrity ï DROP TABLE Command , ALTER TABLE , CREATE INDEX, CREATE UNIQUE INDEX, DROP INDEX SELECT statements with WHERE, ORDER BY, UPDATE Statement, INSERT, DELETE,	12

	nested queries, tuples and multi set comparison , Correlated nested queries ï EXISTS and UNIQUE functions in SQL IN Clause, Explicit sets and renaming of attributes in SQL, Joining tables ï aggregate functions, Grouping - GROUP BY, HAVING clauses ï Views ï Query optimization	
Unit IV	**TRANSACTIONS AND CONCURRENCY MANAGEMENT:** Transactions - ConcurrentTransactions - Locking Protocol - Serialisable Schedules - Locks Two Phase Locking (2PL) - Deadlockand its Prevention - Optimistic Concurrency Control. Database Recovery and Security: DatabaseRecovery meaning - Kinds of failures - Failure controlling methods - Database errors - Backup &Recovery Techniques - Security & Integrity - Database Security - Authorization.	12
Unit V	**DISTRIBUTED AND CLIENT SERVER DATABASES:** Need for Distributed DatabaseSystems - Structure of Distributed Database - Advantages and Disadvantages of DDBMS - Advantages ofData Distribution - Disadvantages of Data Distribution - Data Replication - Data Fragmentation. ClientServer Databases: Emergence of Client Server Architecture - Need for Client Server Computing -Structure of Client Server Systems & its advantages.	12

Text books: format (Title, Author, Publisher, Edition)

1.Database Systems: R.Elmasri & S.B. Navathe, Pearson.
2.Introduction to Database Management System: ISRD Group, McGraw Hill.
3Database Management System: R.Ramakrishnan & J.Gehrke, McGraw Hill.

Reference Books: format (Title, Author, Publisher, Edition)

1.Modern Database Management: J.A.Hoffer,V.Rames &H.Topi, Pearson.
2.Database System Concepts: Silberschatz, Korth & Sudarshan, McGraw Hill.
3.Simplified Approach to DBMS: Parteek Bhaia, Kalyani Publishers.
4.Database Management System: Nirupma Pathak, Himalaya.
5.Database Management Systems: Pannerselvam, PHI.
6.Relational Database Management System: Srivastava & Srivastava, New Age
7.PHPMySQL Spoken Tutorials by IIT Bombay.
8.Oracle Database: A Beginnerôs Guide: I.Abramson, McGraw Hill.

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

\Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analytics Semester: II

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
2PGDDA2	**R Programming**	**60**

Course Outcomes: On completion of this course, students would be able to:

CO1: Develop an R script and execute it
CO2: Install, load and deploy the required packages, and build new packages for sharing and reusability
CO3: Extract data from different sources using API and use it for data analysis
CO4: Visualize and summarize the data
CO5: Design application with database connectivity for data analysis

Unit	Contents	Hrs
Unit I	**Introduction to R?** Why is R Programming Important? Why Learn R? History of Python Features of R Applications of R Comparison between R and Python Which is Better to Choose Pros and Cons of R Companies using R R Packages Downloading and Installing R What is CRAN? Setting R Environment: o Search Packages in R Environment o Search Packages in Machine with inbuilt function and manual searching o Attach Packages to R Environment o Install Add-on Packages from CRAN	12
Unit II	**Programming in R**: Installing R Studio, **Variable Assignment**: Displaying Variables, Deleting Variables. **Comments**: Single Line, Multi Line Comments. Data Types in R, Operators in R, **R Decision Making**: if statement, if ï else statement, ifï else if statement, switch statement, **R Loops**: repeat loop, while loop, for loop ï Loop control statement: break statement, next statement.	12
Unit III	**Data Structures in R:** **Vector**: Vector Creation, Vector Manipulation. **List**: Creating, naming, accessing and manipulating List Elements, Converting List to Vector. **Matrix**: Creating Matrix, Matrix Manipulations **Data frame**: Creating, Vector to Data Frame, Extract Data from Data Frame **Array**: Create Array with Multiple Dimensions, Naming Columns and Rows, Manipulating Array Elements, **Factors**: Factors in Data Frame, Changing the Order of Levels, Generating Factor Levels, Deleting Factor Levels	12
Unit IV	**Loading and handling Data in R**: **Getting and Setting the Working Directory** :getwd(),setwd(), dir() **R-CSV Files**: Input as a CSV file, Reading a CSV File, **Analysing the CSV File**: summary(), min(), max(), range(), mean(), median(), apply() ,Writing into a CSV File, **R -Excel File**: Reading the Excel file.	12
Unit V	**Descriptive Statistics**: Data Range, Frequencies, Mode, Mean and Median: Mean Applying Trim Option, Applying NA Option, Median - Mode - Standard Deviation ï Correlation - Spotting Problems in Data with Visualization: Visually Checking Distributions for a single Variable R ï Pie Charts: Pie Chart title and Colors ï Slice Percentages and Chart Legend, 3D Pie Chart ï R Histograms ï Density Plot - R ï Bar Charts: Bar Chart Labels, Title and Colours.	12

Text books: format (Title, Author, Publisher, Edition)

1. Sandip Rakshit, R Programming for Beginners, McGraw Hill Education (India), 2017, ISBN : 978-93-5260-455-5.

Reference Books: format (Title, Author, Publisher, Edition)

1. Seema Acharya, Data Analytics using R, McGrawHill Education (India), 2018, ISBN: 978-93-5260-524-8.
2. Tutorials Point (I) simply easy learning, Online Tutorial Library (2018), R Programming, Retrieved from https://www.tutorialspoint.com/r/r_tutorial.pdf.
3. Andrie de Vries, Joris Meys, R for Dummies A Wiley Brand, 2nd Edition, John Wiley and Sons, Inc, 2015, ISBN: 978-1-119-05580-8

Links of MOOC[SWAYAM]:
https://onlinecourses.nptel.ac.in/noc19_ma33/preview
https://onlinecourses.swayam2.ac.in/aic20_sp35/preview

Sant Gadge Baba Amravati University, Amravati
Format and Template for Courses (Theory) of PG Programmes

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analysis Semester: II

Code of the Course/Subject	Title of the Course/Subject	(Total Number of Periods)
2PGDDA3	Data Analysis using R and Python	60

Course Outcomes: On completion of this course, students would be able to:

CO1: Establish an efficient scientific computing environment
CO2: Identify and use available R packages and associated Open-Source software to meet given scientific objectives
CO3: Analyse the data by using NumPy and Pandas
CO4: Do Data analysis in deep for
CO5: Process the Data using Python to facilitate Data Analysis.
CO6: Explore the techniques of Python to do the Statistical Analysis and Visualization of Data

Unit	Contents	Hrs
Unit I	Data Visualization using R**: Reading and getting data into R (External Data): Using CSV files, XML files, Web Data, JSON files, Databases, Excel files. Working with R Charts and Graphs: Histograms, Boxplots, Bar Charts, Line Graphs, Scatterplots, Pie Charts** Statistics with R: **Random Forest, Decision Tree, Normal and Binomial distributions, Time Series Analysis, Linear and Multiple Regression, Logistic Regression, Survival Analysis**	**12**
Unit II	**Data Extraction from DATABASES**: RMySQL Package, Connecting to MySQL, Creating Tables in MySQL, Dropping Tables in MySQL, Inserting Data into the Table, Querying the Tables, Query with Filter Clause, Updating Rows in the Tables, Using dplyr and tidyr package **Descriptive Statistics**: Data Range, Frequencies, Mode, Mean and Median: Mean Applying Trim Option, Applying NA Option, Median - Mode - Standard Deviation ï Correlation visually Checking Distributions for a single Variable	**12**
Unit III	Analysing Numerical Data with NumPy: **Arrays in NumPy, Operations on Numpy Arrays, NumPy Array Indexing, Python NumPy Array Indexing, NumPy Array Slicing, NumPy Array Broadcasting.** Analysing Data Using Pandas: **Series, Python Pandas Creating Series, Python Pandas Creating Dataframe, Creating Dataframe from CSV, Filtering DataFrame, Sorting DataFrame, Pandas GroupBy, Pandas Aggregation. SciPy libraries to work with different datasets.**	**12**
Unit IV	**Import and Export of Data:** Installing, loading and using packages for importing and exporting data in Python. **Data Preprocessing and Transformation:** Pre-processing Data in Python, Handling of missing data, Data cleaning and transformation, Data Formatting in Python, Data Normalization in Python, Binning in Python Turning categorical variables into quantitative variables in Python.	**12**
Unit V	**Data Exploration:** Exploring data using statistical methods: mean, median, mode1 , quantiles. Building contingency table 2. Basics of grouping data and Correlation. **GUI Programming:** Tkinter introduction, Tkinter and Python Programming, Tk Widgets, Tkinter examples. Python programming with IDE. **Data Visualization:** Scatter Plot, line graph, histogram, boxplot, line plots regression, word clouds2 , exporting plots as images.	**12**

Text books: format (Title, Author, Publisher, Edition)

1) Wesley J. Chun, ñCore Python Applications Programmingò, 3rd Edition , Pearson Education, 2016
2) An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics. W. N. Venables, D.M. Smith and the R Development Core Team. Version 3.0.1 (2013-05-16).

Reference Books: format (Title, Author, Publisher, Edition)

1) Grus, J. (2016). Data Science from scratch. First edition, Oâreilly (SPD).

2) VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. Second edition,
Oûreilly (SPD).
3) Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
4) Mckinney, W. (2017). Python for Data Analysis. Second edition, Oûreilly (SPD).
5) Jared P Lander, R for everyone: advanced analytics and graphics, Pearson Education, 2013

Links of MOOC[SWAYAM]:

Video Links:
https://www.geeksforgeeks.org/contingency-table-in-python/
https://www.tutorialspoint.com/contingency-table-in-python
https://onlinecourses.swayam2.ac.in/aic20_sp35/preview

Sant Gadge Baba Amravati University, Amravati

Syllabus Prescribed for One Year - Two Semesters Diploma Programme- NEP
PG Diploma Programme in Data Analytics Semester: I

Code of the Course/Subject	Title of the Course/Subject	(No. of Periods/Week)
2PGDDA4	Lab-2: based on 2PGDDA2 and 2PGDDA3	06 periods per Batch

COs

1. Able to Install and use R for simple programming tasks
2. Able to implement the programming techniques in R.
3. Able to do the programming for Data Structures in R.
4. Able to create a vector, import data, saving output and graphics using R
5. Able to to represent data diagrammatically and graphically in R

*** List of Practical/Laboratory Experiments/Activities etc.**

List of Practical on R:

1. WAP in R to print a simple message ñHello Indiaò.
2. WAP in R to demonstrate taking input from users.
3. Programs in R to demonstrate Control Structures in R.
4. Programs in R to demonstrate the use of Functions.
5. Programs in R to demonstrate Data Structures in R.
6. Programs in R to demonstrate the concepts of importing data.

1. Commands and functions in R.
2. Creating a data frame.
3. Diagrams using R.
4. Histogram for raw data.
5. Descriptive statistics.
6. Correlation and Regression.
7. Computing probabilities and drawing random samples for distributions.
8. Fitting of Binomial and Poisson distributions.
9. Test for proportions and means.
10. Chi square test for independence

List of Practical on Data Analytics using R and Python:

1. Practical on Reading and getting data into R from:
    a. CSV files
    b. XML files
    c. web
    d. JSON files
    e. Database
    f. Excel Files
2. Practicals on Random Forest in R.
3. Practicals in R for Statistical Analysis:
    a. Normal and binomial Distributions
    b. Time Series
    c. Regressions
    d. Survival Analysis
4. Practical on creating tables in MySQL and inserting data into tables using R.
5. Practicals on querying the tables in MySQL using R.
6. Practicals on updating and filtering the data of tables in MySQL using R.
7. Practicals to demonstrate dplyr and tidyr packages in R.
8. Python programs to demonstrate the use of the NumPy package.
9. Python programs to demonstrate the use of the Pandas package.
10. Python programs to demonstrate the use of the Tkinter package.
11. Python programs to demonstrate the use of the matplotLib package.