1. STRUCTURE OF THE PROGRAMME (INSTRUCTION&EXAMINATION)

S.No.	SEM	Paper Title	Course type	Instruction periods per week	External Marks	Internal Marks	Total Marks	Credits
1	II	Introduction to Data Science and R Programming	Theory	3	60	40	100	3
		Introduction to Data Science and R Programming -(P)	Lab	2	30	20	50	1
2	III	Python Programming for Data Analysis	Theory	3	60	40	100	3
		Python Programming for Data Analysis –(P)	Lab	2	30	20	50	1
3	IV	Data visualization using Tableau	Theory	3	60	40	100	3
		Data visualization using Tableau-(P)	Lab	2	30	20	50	1
4	IV	Data visualization using python	Theory	3	60	40	100	3
		Data visualization using python-(P)	Lab	2	30	20	50	1
5	V	Supervised Machine Learning with Python	Theory	3	60	40	100	3
		Supervised Machine Learning with Python Practical Course	Lab	2	30	20	50	1
6	V	Unsupervised Machine Learning with Python	Theory	3	60	40	100	3
		UnSupervised Machine Learning with Python –(P)	Lab	2	30	20	50	1

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

SEMESTER-II

R23	Data Science minor-2	L	T	P	C
	Introduction To Data Science And R Programming	4		2	5
	Total Contact Hours:60				

COURSE OBJECTIVES:

Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extractknowledgeandinsight.Thiscoursewillintroducestudentstothecollection.Preparation,analysis, modelling and visualization of data, covering both conceptual and practical issues.

Examples and case studies from diverse fields will be presented, and hands- on use of statistical and data manipulation software will be included.

SYLLABUS

UNIT- I (12 hrs)

Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. History and Overview of R, Getting Started with R, R Nuts and Bolts

UNIT- II (12 hrs)

The Data Science Process: Overview of the Data Science Process-Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

UNIT- III (12 hrs)

Machine Learning: Understanding why data scientists use machine learning-What is machine learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning-Supervised and Unsupervised.

UNIT- IV (12 hrs)

Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets.

UNIT- V (12 hrs)

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation. Case studies on preliminary data analysis.

REFERENCE BOOKS:

1. Davy Cielen, Arno.D.B.Maysman, Mohamed Ali,"Introducing Data Science" Manning Publications, 2016.
2. Roger D.Peng,"RProgramming for Data Science" LeanPublishing,2015.

LEARNING OUTCOMES OF COURSE:

1) Recognize the various discipline that contribute to a successful data science effort.
2) Understand the processes of data science identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization.
3) Beaware of the challenges that arise in Data Sciences.
4) Be able to identify the application of the type of algorithm based on the type of the problem.
5) Be comfortable using commercial and open source tools such as the R/Python language and its associated libraries for data analytics and Visualization.

Lab/Practical/Experiments/Tutorials syllabus:

1.Installing R and R studio, with proper not eson version management, cosmetic settings and different libraries.

2.Basic operations in r with arithmetic and statistics.

3.Getting data into R, Basic data manipulation, Loading Data into R

4.Basic plotting

5.Loops and functions

6.Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.

7.Demonstratethevisualizationandgraphicsusingvisualizationpackageslikeggplot2.

8.Implement Loop functions with lapply(),sapply(),tapply(),apply(),mapply().

9.Explore data using SingleVariables:Unimodal,Bimodal,Histograms,DensityPlots,Barcharts

10. ExploredatausingtwoVariables:Lineplots,ScatterPlots,smoothingcures,Barcharts

11. Explore and implement commands using dplyr package

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

B.Sc. DATA SCIENCE – Minor

R23	SEMESTER - III	L	T	P	C
	COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS (w.e.f. 2023-24 Admitted Batch)	5	3	2	4 (3+1)
	Total Hours - 45				

AIM AND OBJECTIVES OF COURSE:

- To be able to Program in Python
- To know and understand the data Analysis phases.
- To know the usage of all libraries

UNIT I 9hrs

What is Data Analysis? Differences between Data Analysis and Analytics, what is Python, Why Python for Data Analysis? What is Library, Essential Python Libraries. Python Language basics, I Python and Jupyter Notebook. Python Language Basics.

UNIT II 9hrs

Built-in Data Structures, Functions, Files and Operating System. NumPy Basics: Arrays and Vectorized Computation, The Numpy ndarray, Universal Functions, Array-Oriented Programming with Arrays, File Input and Output with Arrays, Linear Algebra, Pseudorandom Number Generation.

UNIT III 9hrs

Getting Started with Pandas: Introduction to Pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics

Data Loading, Storage and File Formats: Reading and Writing Data in Text Format, Binary Data Formats, Interacting with Web APIs, Interacting with Databases.

UNIT IV 9hrs

Data Cleaning and Preparation: Handling Missing Data, Data Transformation, String Manipulation.

Data Wrangling: Join, Combine and Reshape: Hierarchical Indexing, Combining and Merging Datasets, Reshaping and Pivoting.

UNIT V 9hrs

Introduction to Modeling Libraries in Python: Interfacing between pandas and Model code, Creating model descriptions with Patsy, Introduction to stats models.

Plotting and Visualization: A brief matplotlib APT Primer, Plotting with Pandas and Seaborn, Other Python visualization tools.

LEARNING OUTCOMES OF COURSE:

- Understands and learn all basic concepts of
- Python Program Data Analysis methods in Python
- Get used to Python Programming environments.

TEXTBOOKS:

1. Wes McKinney "Python for Data Analysis" 0 'reilly Publications Second edition
2. Charles R Suverance "Python for Everybody" Exploring data using Python 3

REFERENCE BOOKS:

1. John Zelle Michael Smith Python Programming, second edition 2010

Web links:

www.w3schools.com
www.greeksforgreeks.com
www.javatpoint.com
www.programmiz.com

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

B.Sc. DATA SCIENCE - MAJOR

LIST OF PRACTICAL EXPERIMENTS

1. Use matplotlib and plot an inline in Jupyter.
2. Implement commands of Python Language basics
3. Create Tuples, Lists and illustrate slicing conventions.
4. Create built-in sequence functions.
5. Create a functional pattern to modify the strings in a high level.
6. Create an n array object and use operations on it.
7. Use arithmetic operations on Numpy Arrays
8. Create an image plot from a two-dimensional array of function values.
9. Implement some basic array statistical methods and sorting with sortmethod.
10. Implement numpy. random functions.
11. Plot the first 100 values on the values obtained from random walks.
12. Implement the methods of descriptive and summary statistics
13. Transform the data using function or mapping

Lab Outcomes:

- Students will connect, join, and blend heterogeneous data sources into a unified Tableau worksheet.
- Students will generate and refine univariate charts using filters, colors, and labels for clarity.
- Students will design charts encoding 2+ variables with color, shape, and size for trend analysis.
- Students will create interactive maps and overlay demographic/geographic data for spatial insights.
- Students will build formulas and table calculations to enhance datasets and dashboards.
- Students will generate filtered extracts and evaluate efficiency vs. live connections.
- Students will publish workbooks, set permissions, and embed visualizations externally
- Students will share charts/data in multiple formats for cross-platform use.

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

B.Sc. DATA SCIENCE - Minor

R23	**SEMESTER – IV**	**L**	**T**	**P**	**C**
	COURSE 9: **DATA VISUALIZATION USING TABLEAU** **(w.e.f. 2023-24 Admitted Batch)**	**5**	**3**	**2**	**4** **(3+1)**
	Total Contact Hours - 45				

AIM AND OBJECTIVES OF COURSE:

To know the importance of data Visualization in the world of Data Analytics and PredictionTo know the important libraries in Tableau

To get equipped with Tableau Tool

UNIT I **9hrs**

Creating Visual Analytics with tableau desktop, connecting to your data-How to Connect to your data, What are generated Values? Knowing when to use a direct connection, Joining tables with tableau, blending different data sources in a single worksheet.

UNIT II **9hrs**

Building your first Visualization- How Me works- Chart types, Text Tables, Maps, bar chart, Line charts, Area Fill charts and Pie charts, scatter plot, Bullet graph, Gantt charts, Sorting data in tableau,Enhancing Views with filters, sets groups and hierarchies.

UNIT III **9hrs**

Creating calculations to enhance your data- What is aggregation, what are calculated values and table calculations, Using the calculation dialog box to create, Building formulas using table calculations, Using table calculation functions

UNIT IV **9hrs**

Using maps to improve insights-Create a Standard Map View, Plotting your own locations on a map,Replace Tableau's standard maps, Shaping data to enable Point-to-Point mapping.

UNIT V **9hrs**

Developing an Adhoc analysis environment- generating new data with forecasts, providing self evidence adhoc analysis with parameters, Editing views in tableau Server.

LEARNING OUTCOMES OF COURSE:

Students should be able to visualize data through seven stages of data analysis processShould be able to do explanatory and hybrid types of data visualization

Should be able to understand various stages of visualizing data

TEXT BOOKS:

1. Tableau your data-Daniel G. Murray and the Inter works Bl team, Wiley Publications
2. Tableau Data Visualizaton Cookbook, AshutoshNandeshwar, PACKT publishing.
3. A Data Visualization Guide for BusinessProfessionals by Cole NussbaumerKnaflic (2014)
4. ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham (2009)

REFERENCE BOOKS:

1. Designing Data Visualizations: Representing Informational Relationships by Noah Iliinsky, JulieSteele (2011)
2. Alexandra C. Telea - "Data Visualization principles and practice" SecondEdition, CRC Publications
3. Joshua N. Millign-" Learning Tableau -2019" -Third Edition- Packt publications

WebLinks:

www.w3schools.com

www.greeksforgreeks.com

www.javatpoint.com

www.programmiz.com

B.Sc. DATA SCIENCE – Minor

DATA VISUALIZATION USING TABLEAU

LIST OF PRACTICAL EXPERIMENTS

1. Connect to data Sources

2. Create Univariate Charts

3. Create Bivariate and Multivariate charts

4. Create Maps

5. Calculate user-defined fields

6. Create a workbook data extract

7. Save a workbook on a Tableau server and web

8. Export images, data.

LAB OUTCOMES:

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

1. Use Python, R and Tableau for data visualization
2. Apply data visuals to convey trends in data over time using tableau
3. Construct effective data visualsto solve workplace problems
4. Explore and work with different plotting libraries
5. Learn and create effective visualizations

R23	SEMESTER – IV	L	T	P	C
	COURSE 9: **DATA VISUALIZATION USING PYTHON** **(w.e.f. 2023-24 Admitted Batch)**	**5**	**3**	**2**	**4** **(3+1)**
	Total Contact Hours - 45				

COURSE OBJECTIVE:

This course introduces students to data analysis and visualization in the field of exploratory data science using Python.

Unit I **9hrs**

Introduction: Introduction to Data Science, Exploratory Data Analysis and Data Science Process. Motivation for using Python for Data Analysis, Introduction of Python Jupyter Notebook. Essential Python Libraries: NurnPy, pandas, matplotlib, SciPy, scikit-learn, stats models, seaborn.

Unit 2 **9hrs**

Getting Started with Pandas: Arrays and vectorized computation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Storage and File Formats. Reading and Writing Data in Text Format

Unit 3 **9hrs**

Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools. Advanced categorical and numeric plots.

Unit 4 **9hrs**

Data Aggregation and Group operations: Group by Mechanics, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation, Time Series Data Analysis: Date and Time Data Types andTools, Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.

Unit 5 **9hrs**

Advanced Pandas: Categorical Data: cleaning data and visualization techniques, Advanced Group Bymethods, Use Techniques for Method Chaining.

COURSE LEARNING OUTCOMES:

On successful completion of the course, the students will be able to

1. Use data analysis tools.
2. The specialization tool pandas library.
3. Load, clean, transform, merge and reshape data.
4. Data cleaning and grouping operations mechanism.
5. Advanced pandas data frames utilization.

TEXT BOOK:

1.McKinney, W.(2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy andIPython. 2nd edition. O'Reilly Media.

REFERENCE:

1. O'Neil, C., & Schutt, R. (2013). Doing Data Science: Straight Talk from the Frontline

Web Links:

www.w3schools.com

www.greeksforgreeks.com

www.javatpoint.com

www.programmiz.com

LIST OF PRACTICAL EXPERIMENTS

1. Practical based on NumPy and array
2. Practical based on Pandas Data Structures
3. Practical based on Data Loading, Storage and File Formats
4. Practical based on Interacting with Web APis
5. Practical based on Data Cleaning and Preparation
6. Practical based on Data Wrangling
7. Practical based on Data Visualization using matplotlib
8. Practical based on Data Aggregation
9. Practical based on Time Series Data Analysis

Lab Outcomes:

At the end of the course the student will be able to:

CO 1. Demonstrate the use of IDLE or PyCharm IDE to create Python Applications
CO 2. Use Python programming constructs to develop programs for solving real-world problems CO 3. Use Matplotlib for drawing different Plots
CO 4. Demonstrate working with Seaborn, Bokeh for visualization.
CO 5. Use Plotly for drawing Time Series and Maps

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

Programme: B.Sc. Honors Data Science (Minor)

w.e.f. AY 2023-24

R23	**SEMESTER-V**	**L**	**T**	**P**	**C**
	COURSE 12: SUPERVISED ML WITH PYTHON (w.e.f. 2023-24 Admitted Batch)	**5**	**3**	**2**	**4 (3+1)**
	Total Hours - 45				

AIM AND OBJECTIVES OF COURSE:

- The purpose of this course is to serve as an introduction to Supervised machine learning with Python.
- We will explore several classifications, regression algorithms and see how they can help usperform a variety of Supervised machine learning tasks.

UNIT I **9hrs**

Machine Learning Basics: What is machine learning? Key terminology, Key tasks of machine learning, How to choose right algorithm, steps in developing a machine learning, why python?Getting started with Numpy library Classifying with k- Nearest Neighbors: The k-Nearest Neighbors classification algorithm, Parsing and importing data from a text file, Creating scatter plots with Matplotlib, Normalizing numeric values

UNIT II **9hrs**

Splitting datasets one feature at a time-Decision trees: Introducing decision trees, measuring consistency in a dataset, using recursion to construct a decision tree, plotting trees in Matplotlib

UNIT III **9hrs**

Classifying with probability theory-Naïve Bayes: Using probability distributions for classification, learning the naïve Bayes classifier, Parsing data from RSS feeds, using naïve Bayes to reveal regional attitudes

UNIT IV **9hrs**

Logistic regression: Classification with logistic regression and the sigmoid function, Usingoptimization to

find the best regression coefficients, the gradient descent optimization algorithm, Dealing with missing values in the our data

UNIT V **9hrs**

Support vector machines: Introducing support vector machines, using the SMO algorithm for optimization, using kernels to “transform” data, Comparing support vector machines with other classifiers

Learning outcomes of Course:

- Able to understand introduction to machine learning concepts.
- Able to Loading datasets, build models and model persistence.
- Understand Feature extraction from data sets.
- Able to do Regression & Classification.
- Able to compare SVM with other classifiers.

TEXT BOOK:

1. Machine learning in action, Peter Harrington by Manning publicationsSupervised ML with Python Lab

Web Links:

www.w3schools.com

www.greeksforgreeks.com

www.javatpoint.com

www.programmiz.com

MAHARAJAH'S COLLEGE (AUTONOMOUS) :: VIZIANAGARAM

B.Sc. DATA SCIENCE - MINOR

LIST OF PRACTICAL EXPERIMENTS

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a CSV file. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.
2. Write a program to demonstrate the working of the decision tree based ID3 algorithm.
3. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a CSV file.
4. Assuming a set of documents that need to be classified, use the naïve BayesianClassifier modelto perform this task. Built-in Java classes/API can be used towrite the program. Calculate the accuracy, precision, and recall for your dataset.

Lab Outcomes:

Learn to build and evaluate models for classification and regression tasks, using techniques like data preprocessing, model training, hyper parameter tuning, and performance evaluation metrics.

	SEMESTER - V	L	T	P	C
R23	COURSE 13: UNSUPERVISED ML WITH PYTHON (w.e.f. 2023-24 Admitted Batch)	5	3	2	4 (3+1)
	Total Contact Hours – 45				

Aim and objectives of Course (Unsupervised ML with Python):

- Unsupervised Machine Learning involves finding patterns in datasets.
- The core of this course involves study of Clustering, feature extractionand optimizationalgorithms.
- The purpose of this course is to serve as an introduction to machine learning with Python.

UNIT I **9hrs**

Unsupervised Learning: Clustering: k-means clustering algorithm, Improving clusterperformance with post processing, Bisecting k-means, Example: clustering points on a map

UNIT II **9hrs**

Association analysis : Apriori algorithm: Association analysis, The Apriori principle, Finding frequent item sets with the Apriori algorithm, Mining association rules from frequent item sets, uncovering patterns in congressional voting

UNIT III **9hrs**

Finding frequent item sets: FP-growth –FP trees, Build FP-tree, mining frequent from an FP- tree, finding co-occurring words in a Twitter feed, mining a click stream from a news site.

UNIT IV **9hrs**

Principal component analysis: Dimensionality reduction techniques, using PCA to reduce the dimensionality of semiconductor manufacturing data

UNIT V **9hrs**

Singular value decomposition: Applications of the SVD, Matrix factorization, SVD in Python, Collaborative filtering–based recommendation engines, a restaurant dishrecommendation engine

Learning outcomes of Course:

- Able to do Clustering, feature extraction and optimization.

- Students will be able to understand and implement in Python algorithms of Unsupervised
- Machine Learning and apply them to real-world datasets.

Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

TEXT BOOK:

1. Machine learning in action, Peter Harrington by Manning publicationsUnsupervised MLwith Python Lab

Web Links:

www.w3schools.com

www.greeksforgreeks.com

www.javatpoint.com

www.programmiz.com

SEMESTER-V

COURSE 13: UNSUPERVISED ML WITH PYTHON

LIST OF PRACTICAL EXPERIMENTS

1. Implementation of K-Means Clustering
2. Implement the bisecting k-means clustering algorithm
3. Implement Apriori algorithm
4. Implement Association rule-generation functions
5. Implement FP-tree creation
6. Write a function to find all paths ending with a given item.
7. Implement Code to access the Twitter Python library
8. Implement the PCA algorithm
9. Write a program to find Rating estimation by using the SVD
10. Implement Image-compression functions using SVD.

Lab Outcomes:

- Explore techniques like clustering (grouping similar data points) and dimensionality reduction (reducing data complexity) on unlabeled datasets, aiming to discover patterns and insights.