Sheth NKTT College of Commerce and Sheth JTT College of Arts, Thane
(Autonomous)
Syllabus (w.e.f. 2025-26)
M.Sc. (Data Science) Part I

Semester I

Category	Code	Course	Credit
Mandatory	MDE101	Essential Technologies for Data Science	**4**
	MDEP102	Essential Technologies for Data Science Practical	**2**
	MDD103	Data Analysis and Visualization	**4**
	MDDP104	Data Analysis and Visualization Practical	**2**
	MDS105	Statistical Methods for Data Science	**2**
Elecive (Any One)	MDT106	SPARK Technologies	**2 + 2**
	MDTP107	SPARK Technologies Practical	
	MDR108	Retail Marketing Analytics	
	MDRP109	Retail Marketing Analytics Practical	
	MDS110	Sports Data Analytics	
	MDSP111	Sports Data Analytics Practical	
	MDRM112	Research Methodology	**4**
		Total	**22**

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M. Sc. (Data Science)**	**Semester: I**
Course Category/Vertical: Major Mandatory	
Name of the Dept: **Science and Technology**	
Course Title: **Essential Technologies for Data science**	
Course Code: MDE101	Course Level:6.0
Type: Theory	
Course Credit: 4 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 100 Marks	

Course Objectives:

1. Learn Python basics including programming elements, data structures, control flow, and functions for data handling.
2. Understand core statistics concepts like probability, distributions, hypothesis testing, and apply them using NumPy.
3. Perform data wrangling and analysis using Pandas for cleaning, transforming, and summarizing datasets.
4. Gain skills in data visualization using Matplotlib and R (ggplot2) and understand different types of data and analysis methods.

Course Outcomes:

CO1. Students should be able to understand the core concepts of programming before starting to write new programs. Students should be able to develop logic for Problem Solving.
CO2. To manage data by relying on data structures such as strings, arrays, files, lists, and dictionaries.
CO3. Exemplify the various levels of decision making on a program and implement a mix of loops, functions, and control flow to extract information from a data structure.
CO4. Student should able to learn different programming techniques and tools related to data science.

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
	Unit 1: Introduction to Data science and Python a) Introduction to Data Science, data science life cycle, Applications, and advantages of Python over other programming languages b) What is Python? Why Should I learn Python? Installing Python How to execute Python program Writing your first program. c) Basic programming elements of Python-variables and constants, identifiers, Typecasting or Type Conversion in Python, indentation, comments, rules of writing identifiers, primitive data types, writing command line programs in python d) Operators in Python: Arithmetic operators, relational operators, Logical operators, Membership operators, Taking user input. **Unit 2: Data structures and control flow** a) Collection data structures in Python- List, tuples, dictionary, sets and strings b) Control flow- Sequential, Branching or Conditional, Iteration or Repetition, Modular or Subroutines Conditional and iteration statements: if else statements, loops, for loop and while loops. c) User defined functions in Python- No Value Pass and No Return, Value Pass and No Return, Value Pass and Return, Function With default arguments, Function with variable arguments, Higher order functions, list comprehension	15
II	**Unit 3: Statistics for Data Analysts** a) Permutations and combinations, probability, Descriptive statistics (mean, median, mode), point estimation, quartiles and boxplot, methods of dispersion, random variables and probability distribution b) Measures of shape- skewness, kurtosis, outlier detection, transformation (log, square root)	15

	c) Inferential statistics- Sampling techniques, Hypothesis testing, Z-score normalization, correlation, ANOVA d) Introduction to NumPy, creating NumPy arrays, indexing and slicing, vectorization, Boolean indexing, transformation, inferential statistics using NumPy **Unit 4: Data wrangling using Pandas.** a) Introduction to data: NOIR (nominal, Ordinal, Interval and Ratio), continuous and discrete numeric data. Types of data analysis (descriptive, diagnostic, predictive and prescriptive analysis) b) Data wrangling using Pandas - Creating Series, Creating Data frame from dictionary, attributes, and method description of a data frame. Drop columns, add columns, add rows, iloc , loc, indexing and slicing data frames, selection with condition, group by summary operation, sorting operations c) Introduction to R IDE- components of R IDE, Basic data types in R, Data structures in R, data coercion, importing files, visualization using ggplot2. Basic visualization using matplotlib- Components of a chart, line chart, scatter chart, pie chart, sub plots.	
	Total Hours	60

References:

1. Data Analysis with Pandas and Python by Boris Paskhaver, Manning Publications. Available at: https://www.perlego.com/book/2881120/pandas-in-action-pd
2. Practical Statistics for Data Scientists: 50 Essential Concepts by Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 2017 ISBN-10: 1491952962 ISBN-13: 978-1491952962
3. Foundations of Statistics for Data Scientists With R and Python By Alan Agresti, Maria Kateri, CRC Press Taylor and Francis group, 2022

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M. Sc. (Data Science)**	**Semester: I**
Course Category/Vertical: Major Mandatory	
Name of the Dept: **Science and Technology**	
Course Title: **Essential technologies for Data Science Practical**	
Course Code: MDEP102	Course Level:6.0
Type: Practical	
Course Credit: 2 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. Student will be able to develop hands-on skills in Python programming by practicing arithmetic, relational, logical operations, loops, and working with core data structures.
2. Student will be able to gain experience in data analysis by importing datasets, performing univariate, bivariate, and multivariate analysis, and visualizing data using Python, R, or Excel.
3. Student will be able to apply statistical techniques such as hypothesis testing, ANOVA, and correlation analysis with appropriate visualizations like heatmaps.
4. Student will be able to understand data wrangling and preprocessing through operations like data selection, filtering, updating, grouping, and sorting using tools like Pandas or spreadsheet software.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Should be able to write basic programming in Python.

CO2. Should be able to use Python data structures and able to use conditional and iterative control flow.

CO3. Should be able to demonstrate descriptive, diagnostic, and inferential statistics using, Python, R or Excel (use Data analysis tool pack in Excel or Data analyzer tool in Microsoft office 365)

CO4. Perform basic data wrangling using R or Pandas

CO5. Perform data visualization using R or Pandas

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Name of Practical	Hours
1	Write a Python program to accept inputs from users and perform arithmetic operations.	
2	Write a program to demonstrate relational and logical operators in Python.	
3	Write a Python program to demonstrate usage of loops. Use both for and while loops to distinguish between them. e.g., Reversing the digits of a number without converting to String.	
4	Demonstrate the use of data structures list, sets, dictionary.	
5	Import a dataset and perform univariate analysis on the numeric columns	
6	Demonstrate Hypothesis testing, and ANOVA using a dataset *use Python, R or Excel+	
7	Demonstrate correlation analysis. Use heatmap for visualization. Write inferences.	
8	Import a csv or Excel dataset and demonstrate data wrangling, view shape, dimension, column names of the dataset, ways to select data using column number, column names, simple and compound conditional selection, update and modify dataset.	
9	Demonstrate group by summary operations and sorting techniques.	
10	Perform univariate, bivariate and multivariate analysis using visualization techniques in Python, R or Excel	
		60

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M. Sc. (Data Science)**	**Semester: I**
Course Category/Vertical: Major Mandatory	
Name of the Dept: **Science and Technology**	
Course Title: **Data Analysis and Visualization**	
Course Code: MDD103	Course Level:6.0
Type: Theory	
Course Credit: 4 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 100 Marks	

Course Objectives:

1. Understand the basics of data analysis, including the data science process, types of data, sources, file formats, and tools used by data professionals.
2. Gain practical skills in data wrangling, such as importing data, cleaning, transforming, merging datasets, and handling missing or unstructured data.
3. Learn to visualize data effectively using tools like Tableau by creating various charts, dashboards, and adding interactivity for deeper insights.
4. Develop storytelling skills with data using Power BI by designing interactive reports, customizing visuals, and presenting data-driven insights clearly**.**

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Understand various data formats, sources and storage mechanisms.
CO2. Handle missing data and manage data wrangling and manipulation
CO3. Create data visualization and report making using various software tools
CO4. Demonstrate the visualizations and make interpretations and create a data story using various software tools.

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
I	**Unit 1: Introduction to Data Analysis** Data Analysis - Exploratory Data Analysis and Data Science Process- Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Sources of Data - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis	
II	**Unit2: Data Wrangling** Data Sources - How to gather and Import Data - Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting Tools for Data Wrangling; Data Cleaning and Preparation Handling Missing Data, Data Transformation, String Manipulation.	
III	**Unit 3: Data Visualization** Intro to data visualization - Introduction to Visualization and Dashboarding Software - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) – Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts.	
IV	**Unit 4: Story Telling** Introduction to Power BI - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles. Report Format Options - Introduction to data storytelling - Creating a data story	
	Total Hours	**60**

References :

1. Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython by McKinney, W., 2nd edition. O‟Reilly Media, 2017
2. Doing Data Science: Straight Talk from the Frontline by O‟Neil, C., & Schutt, R, O‟Reilly Media, 2013
3. The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017
4. Practical Tableau by Ryan Sleeper, O‟Reilly Media, 2018
5. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
6. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M. Sc. (Data Science)**	**Semester: I**
Course Category/Vertical: Major Mandatory	
Name of the Dept: **Science and Technology**	
Course Title: **Data Analysis and Visualization Practical**	
Course Code: MDDP104	Course Level:6.0
Type: Practical	
Course Credit: 2 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To learn data loading, storage, and file handling techniques, including reading data from various formats, web scraping, and interacting with Web APIs.
2. To apply data wrangling and cleaning methods such as handling missing values, transforming data, and performing string manipulation for analysis.
3. To develop skills in data visualization by creating charts, applying filters and sorts, and building interactive dashboards using tools like Tableau and Power BI.
4. To create and present data-driven stories through well-structured reports and visualizations, and publish them using Tableau Cloud or Power BI services.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Handle missing data and manage data wrangling and manipulation

CO2. Create data visualization and report making using various software tools

CO3. Demonstrate the visualizations and make interpretations

CO4. Create a data story using various software tools.

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Content	Hours
1	Implement Data Loading, Storage and File Formats. Read data and store them in text format.	
2	Implement the code to interact with Web APIs and to perform web scrapping.	
3	Demonstrate Data Cleaning and Preparation.	
4	Implement Data wrangling on a data set.	
5	Demonstrate the handling of missing data and string manipulation.	
6	Create common charts with title, labels and descriptions using Tableau.	
7	Perform sorting and filtering using tableau, create visualizations and publish it on Tableau Cloud.	
8	Perform data visualization using Power BI.	
9	Create reports using Power BI.	
10	Create a data story in Tableau or power BI.	
		60

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M. Sc. (Data Science)**	**Semester: I**
Course Category/Vertical: Major Mandatory	
Name of the Dept: **Science and Technology**	
Course Title: **Statistical Methods for Data Science**	
Course Code: MDS105	Course Level:6.0
Type: Theory	
Course Credit: 2	
Hours Allotted: 30 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To understand foundational concepts in statistics and inference, including significance testing, error types, confidence intervals, and power analysis.
2. To apply various statistical tests such as t-tests, ANOVA, correlation analysis, and nonparametric methods to analyze and interpret data.
3. To learn multivariate statistical techniques, including linear and logistic regression, MANOVA, discriminant analysis, and their applications in data science.
4. To explore advanced techniques like Principal Component Analysis, Factor Analysis, and Cluster Analysis

Course Outcomes:

CO1: Understand and apply fundamental statistical concepts and methods, including probability, sampling distributions, hypothesis testing, confidence intervals, regression analysis, ANOVA, chi-square tests, and non-parametric tests for data analysis in Data Science.

CO2: Utilize statistical software and programming tools such as Python or R to perform statistical analyses, visualize data, automate analytical processes, and solve data-driven problems using critical thinking and statistical reasoning.

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
I	**Unit 1: Introduction to Applied Statistics** The Nature of Statistics and Inference, what is "Big Data"? Statistical Modelling, Statistical Significance Testing and Error Rates, Simple Example of Inference Using a Coin, Statistics is for Messy Situations, Type I versus Type II Errors, Point Estimates and Confidence Intervals, Variable Types, Sample Size, Statistical Power, and Statistical Significance, The Verdict on Significance Testing, Training versus Test Data. Means, Correlations, Counts: Drawing Inferences: Computing z and Related Scores, Statistical Tests, Plotting Normal Distributions, Correlation Coefficients, Evaluating Pearson's r for Statistical Significance, Spearman's Rho: A Nonparametric Alternative to Pearson. Tests of Mean Differences: t-Tests for One Sample, Two Sample t- Test, Paired-Samples t- Test Categorical Data: Binomial Test, Categorical Data Having More Than Two Possibilities. Power Analysis and Sample Size Estimation: Power for t-Tests, Power for One-Way ANOVA, Power for Correlations. Analysis of Variance: Fixed Effects, Random Effects, Mixed Models, Introducing the Analysis of Variance (ANOVA), Performing the ANOVA, Random Effects ANOVA and Mixed Models, One-Way Random Effects ANOVA	15
II	**Unit 2: Multivariate Techniques** Simple and Multiple Linear Regression, Hierarchical Regression, How Forward Regression Works Logistic Regression and the Generalized Linear Model, Predicting Probabilities, Multiple Logistic Regression, Training Error Rate Versus Test Error Rate. Multivariate Analysis of Variance (MANOVA)and Discriminant Analysis: Multivariate Tests of Significance, Example of MANOVA, Outliers, Homogeneity of Covariance Matrices, Linear Discriminant Function Analysis, Theory of Discriminant Analysis, Predicting Group Membership, Visualizing Separation. Principal Component Analysis: Principal Component Analysis Versus Factor Analysis, Properties of Principal	15

	Components, Component Scores, How Many Components to Keep? Exploratory Factor Analysis, Common Factor Analysis Model, Factor Analysis Versus Principal Component Analysis on the Same, Initial Eigenvalues in Factor Analysis, Rotation in Exploratory Factor Analysis, Estimation in Factor Analysis. Cluster Analysis: k-Means Cluster Analysis, Minimizing Criteria, Example of k-Means Clustering Hierarchical Cluster Analysis, Why Clustering Is Inherently Subjective. Nonparametric Tests: Mann– Whitney U Test, Kruskal–Wallis Test, Nonparametric Test for Paired Comparisons and Repeated	
	Total Hours	30

References:

1. Gupta S. C., Kapoor V. K.: Fundamentals of Mathematical Statistics; Tenth Edition.Sultan Chand &Sons. (2000)
2. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, Prentice-Hall, NewJersey, 2002.
3. Draper, N. R. and Smith, H. (1998), Applied Regression Analysis (John Wiley), Third Edition.
4. Purohit, S. G. Gore, S.D. and Deshmukh, S.R. (2015). Statistics using R, second edition. Narosa Publishing House, New Delhi.
5. Daniel W. W.: Applied Non-Parametric Statistics, First edition Boston-Houghton Mifflin Company.

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Information Technology)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title: **SPARK Technologies**	
Course Code: MDT106	Course Level:6.0
Type: Theory	
Course Credit: 2 credits	
Hours Allotted: 30 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To explain the fundamentals of Apache Spark including its architecture, unified stack, ecosystem, features, and in-memory computing model for big data processing.
2. To apply Spark programming concepts such as RDD creation, transformations, actions, key-value pair RDDs, MapReduce operations, and partitioning using Scala and Python.
3. To design and develop Spark applications by working with Spark Context, Spark Shell, Hadoop-HDFS integration, data persistence, lineage, and performance tuning techniques.
4. To utilize advanced Spark components like Spark SQL, Spark Streaming, GraphX, and Spark ML libraries to perform large-scale data analytics, statistics, and machine learning tasks.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Understand the concept of SPARK Technologies and its implementation and the concept of RDD

CO2. Understand the implementation of SPARK SQL, GraphX, Performance Tuning and use Sparks Resilient Distributed Datasets to process and analyze large data sets across many CPUs

CO3. Understand how the GraphX library helps with network analysis problems and understand how Spark SQL lets you work with structured data

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
I	**Unit I: Basics of SPARK Technologies Unit** **Introduction to SPARK Technologies** Components of the Spark unified stack, Features of Spark, Spark Web UI, an introduction to RDDs - Resilient Distributed Datasets, Launching and using Spark's Scala and Python shell, Spark Context, Spark Ecosystem, In- Memory data – Spark, Creating, Loading and Saving RDD, Transformations in RDD, Actions in RDD, Key-Value Pair RDD, Map Reduce and Pair RDD operations RDD Partitions	
II	**Unit 2: Implementation of SPARK Technologies** Spark Applications vs. Spark Shell, Creating Spark Context, Building a Spark Application, Spark and Hadoop Integration-HDFS, Handling Sequence File, Spark RDD- RDD Lineage, RDD Persistence Overview, Distributed Persistence. Spark Streaming, ML library for Spark, Working with Statistics, SPARK SQL, GraphX, Performance Tuning.	
	Total Hours	30

References:

1. Learning Spark: Lightning-Fast Data Analytics 2nd Edition, by Jules S. Damji, Brooke Wenig , Tathagata Das, Denny Lee, O"Reilly , 2020
2. Apache Spark Machine Learning Blueprints 1st Edition, Kindle Edition by Alex Liu, Packt Publishing, 2016
3. Apache Spark 2.x Cookbook: Cloud-ready recipes for analytics and data science 2nd Edition, by Rishi Yadav, Packt Publishing, 2017

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title: **SPARK Technologies Practical**	
Course Code: MDTP107	Course Level:6.0
Type: Practical	
Course Credit: 2 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To install and configure Apache Spark and understand Spark basics, RDD interface, and core execution concepts.
2. To apply RDD operations and transformations such as filtering, flatMap(), and aggregation to solve real-world data processing problems.
3. To work with Spark SQL and DataFrames to execute SQL commands, perform analytical queries, and implement business use cases like customer spending analysis.
4. To develop advanced Spark applications using broadcast variables, Spark ML for recommendation systems, and Spark Structured Streaming for real-time data analysis.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Understand the concept of SPARK Technologies and its implementation

CO2. Understand the concept of RDD

CO3. Understand the implementation of SPARK SQL, GraphX, Performance Tuning.

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Content	Hours
1	Installation of Apache Spark	
2	Spark Basics and RDD interface	
3	Filtering RDDs, and the Minimum Temperature by Location Example	
4	Counting Word Occurrences using flatmap()	
5	Executing SQL commands and SQL-style functions on aDataFrame	
6	Implement Total Spent by Customer with DataFrames	
7	Use Broadcast Variables to Display Movie Names Instead of ID Numbers.	
8	Create Similar Movies from One Million Rating	
9	Using Spark ML to Produce Movie Recommendations	
10	Use Windows with Structured Streaming to Track Most-Viewed URLs (Spark Streaming)	
	Total Hours	60

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title:**Retail Marketing Analytics**	
Course Code: MDR108	Course Level:6.0
Type: Theory	
Course Credit: 2 credits	
Hours Allotted: 30 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To understand the role of marketing analytics in data-driven decision making by explaining marketing optimization frameworks and the importance of analytics in modern marketing.
2. To perform exploratory data analysis (EDA) using R by cleaning, summarizing, tabulating, and visualizing marketing data using plots such as histograms, boxplots, scatter plots, and ggplot().
3. To design and conduct marketing experiments by applying principles of experimental design, randomization, sample splitting, and hypothesis testing to evaluate campaign effectiveness.
4. To analyse customer behaviour and value by calculating and predicting Customer Lifetime Value (CLV) using regression techniques and applying cohort analysis for experiment-based insights.

Course Outcomes (OC):

Upon completing this course, the student will be able to:

CO1. Understand the importance of marketing analytics for forward looking and systematic allocation of marketing resources

CO2. Know how to use marketing analytics to develop predictive marketing dashboard for organization

CO3. Analyze data and develop insights from it to address strategic marketing challenges

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Content	Hours
I	**Unit 1: Introduction to Marketing Analytics and Exploratory Data analytics using R** **a) Course Introduction** • why marketing analytics? • course description and learning objectives **b) Marketing Analytics Overview** • how analytics can assist marketing decision-making • the framework of marketing optimization **c) Tabulate and Summarize data** • what cleaned data looks like • simple histogram plot • use histogram and boxplot to inform data distribution **d) Visualize data** • elements of data visualization • histogram, scatter plot, line plot, bar charts, line fits with the ggplot() function.	
II	**Unit 2: Marketing Campaigns - Experiment Design, Customer Lifetime Value (CLV) and Cohort Analysis** **a) Design and Conduct Experiments** design experiments, examples • randomization/sample splitting • conduct experiments **b) Assess Experiment Outcome Using Hypothesis Testing** • why hypothesis testing for experiment outcomes • terminologies for hypothesis testing • how does hypothesis testing work • power calculation • conduct hypothesis testing in R **c) Calculate and Predict CLV** • calculate CLV • typical frameworks in predicting CLV • using linear regression and logistic regression to predict CLV **d) CLV Analysis and Cohort Analysis Introduction to Experiment**	
	Total Hours	30

References:

1. Hands-on Data Science for Marketing by Yoon Hyup Hwang, Packt Publishing, 2019
2. Retail Analytics: The Secret Weapon by Emmett Cox,1st edition , Weily , 2011
3. Cutting Edge Marketing Analytics: Real World Cases and Data Sets for Hands on Learning by Venkatesan Rajkumar, Farris Paul and Ronald Wilcox, Pearson FT Press, 2014
4. Marketing Analytics: A Practical Guide to Real Marketing Science by Grigsby Mike,Kogan Page, 2015

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title: **Retail Marketing Analytics Practical**	
Course Code: MDRP109	Course Level:6.0
Type: Practical	
Course Credit: 2 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To understand and analyse retail marketing data using basic analytical techniques.
2. To develop simple dashboards and visualizations for marketing data interpretation and decision making.
3. To apply basic statistical concepts such as hypothesis testing to evaluate marketing experiments and outcomes.
4. To interpret customer and sales data to identify trends and support retail marketing strategies.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Analyse retail marketing datasets to identify customer behaviour and sales patterns.

CO2. Apply basic statistical techniques such as hypothesis testing to evaluate marketing experiments and develop simple dashboards and visualizations to present marketing insights.

CO3. Interpret analytical results to support retail marketing decision-making.

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Content	Hours
1	**Learn how to tabulate and summarize marketing data using R.** • Clean and preprocess the marketing data. • Generate a simple histogram plot to visualize data distribution. • Use tabulation and summary functions to gain insights from the data. • Interpret the findings and discuss the implications for marketing analysis.	
2	**Gain proficiency in visualizing marketing data using R.** • Understand the key elements of data visualization. • Create various visualizations such as histograms, scatter plots, line plots, and bar charts using the ggplot() function in R. • Apply appropriate visualization techniques to effectively communicate marketing insights.	
3	**Design and conduct experiments for marketing campaigns.** • Learn about experimental design and its application in marketing. • Design experiments using examples from marketing scenarios. • Implement randomization and sample splitting techniques. Conduct the experiments and collect relevant data for analysis.	
4	**Understand the concept of hypothesis testing and its role in assessing experiment outcomes.** • Explore the purpose of hypothesis testing in analyzing experiment results. • Familiarize with key terminologies related to hypothesis testing. • Learn the process of hypothesis testing and power calculation. • Conduct hypothesis testing using R to evaluate experiment outcomes.	
5	**Calculate and predict Customer Lifetime Value (CLV).** • Calculate CLV using different approaches and frameworks. • Explore predictive modeling techniques such as linear regression and logistic regression for CLV prediction. • Assess the accuracy and reliability of CLV predictions	

6	**Apply CLV analysis and cohort analysis in marketing analytics.** • Analyze CLV data and identify patterns and trends. • Perform cohort analysis to segment customers based on their behavior or characteristics. • Interpret the results of CLV analysis and cohort analysis to derive actionable insights for marketing strategies.	
7	**Extract data from social media platforms and perform analysis to gain insights into customer behavior and preferences.** • Utilize Python libraries like BeautifulSoup and requests to scrape data from social media platforms. • Clean and preprocess the scraped data. • Analyze the data to identify trends, sentiment analysis, or customer engagement metrics. • Visualize the findings using appropriate charts or graphs.	
8	**Analyze customer purchasing patterns and build a recommender system based on market basket analysis.** • Use transactional data to identify frequently occurring item sets using association rule mining algorithms. • Calculate support, confidence, and lift for the identified item sets. • Build a recommendation engine using collaborative filtering techniques. • Evaluate the performance of the recommender system and make recommendations based on customer preferences.	
9	**Segment customers based on their recency, frequency, and monetary value (RFM) to better target marketing efforts.** • Analyze customer transaction data to calculate RFM scores. • Segment customers into different groups using clustering algorithms such as k-means or hierarchical clustering. Perform descriptive analysis on each customer segment to understand their characteristics. • Develop targeted marketing strategies for each segment based on their RFM profiles.	
10	**Conduct A/B testing to evaluate the impact of different marketing strategies and make data-driven decisions.** • Design and implement A/B tests for marketing campaigns using randomized assignment. • Collect relevant data and perform statistical analysis to compare the performance of different strategies. • Calculate key metrics such as conversion rates, click-through rates, or revenue. • Interpret the results and provide recommendations for optimizing marketing campaigns based on the findings.	

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title: **Sports Data Analytics**	
Course Code: MDS110	Course Level:6.0
Type: Theory	
Course Credit: 2 credits	
Hours Allotted: 30 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To understand the role and importance of data analytics in sports.
2. To develop skills in collecting, cleaning, and managing sports data.
3. To apply statistical analysis and basic predictive techniques for analysing sports performance data.
4. To use data visualization tools to present sports data insights effectively.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1: Understand the role and significance of sports analytics and develop the ability to collect, clean, manage, analyze, and visualize sports data using appropriate statistical techniques.

CO2: Apply predictive modeling and machine learning techniques to sports data to generate insights, support decision-making, and improve performance analysis in sports.

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
I	**Unit 1: Fundamentals of Sports Data Analytics** **A. Introduction to Sports Data Analytics** • Overview of sports data analytics • Evolution of analytics in sports • Importance and applications of sports data analytics **B. Data Collection and Preprocessing** • Sources of sports data • Data collection methods • Data cleaning and preprocessing techniques • Database management for sports analytics **C. Data Visualization for Sports Analytics** • Principles of data visualization • Visualization tools and libraries • Creating effective visualizations for sports data • Interactive dashboards for sports analytics **D. Statistical Analysis in Sports** • Descriptive statistics for sports data • Hypothesis testing in sports analytics • Regression analysis in sports • Analysis of variance (ANOVA) in sports	
II	**Unit 2: Advanced Techniques in Sports Data Analytics** **E. Predictive Modeling in Sports Analytics** • Introduction to predictive modeling • Feature selection and engineering for sports data • Linear regression models for sports prediction • Classification models for sports outcomes **F. Machine Learning in Sports Analytics** • Overview of machine learning algorithms • Decision trees and random forests in sports analytics • Support vector machines for sports prediction Neural networks and deep learning in sports analytics **G. Advanced Topics in Sports Data Analytics** • Sports performance analysis • Player tracking and motion analytics • Sports marketing and fan engagement analytics • Sports injury prediction and prevention **H. Sports Business Analytics** • Revenue generation and marketing in sports • Fan engagement and customer analytics	

Reference Books:

1. Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers by Benjamin C. Alamar, Columbia university press, 2013
2. Sports Analytics and Data Science: Winning the Game with Methods and Models by Thomas Miller, 1st edition, Pearson FT Press, 2015
3. Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performanceby Daniel Memmert, Tim McGarry, and Tony Reilly, 2018
4. Cricket Analytics: Analytics and Data Science in Cricket by Tapan Bagchi and S. Raghunathan
5. Machine Learning using Python by Manaranjan Pradhan and U. Dinesh Kumar, Weily , 2020

URL for Online Study Material:

1. www.coursera.org
2. www.kaggle.com
3. www.datacamp.com
4. www.sportsanalyticsinstitute.com

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Elective	
Name of the Dept: **Science and Technology**	
Course Title: **Sports Data Analytics Practical**	
Course Code: MDSP111	Course Level:6.0
Type: Practical	
Course Credit: 2 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 50 Marks	

Course Objectives:

1. To develop a foundational understanding of Sports Data Analytics, including key concepts, metrics, and applications in player and team performance evaluation.
2. To build proficiency in data collection, data manipulation, and exploratory data analysis (EDA) using appropriate analytical tools and techniques.
3. To enable students to analyse player and team performance and develop predictive models for forecasting outcomes and improving strategic decision-making.
4. To equip students with skills in data visualization and analytical reporting, enabling effective communication of insights to coaches, managers, and stakeholders.

Course Outcomes:

Upon completing this course, the student will be able to:

CO1. Understanding of Sports Data Analytics and proficiency in Data Manipulation and Exploratory Data Analysis

CO2. Player Performance/ Team Performance Analysis and predictive Modeling, Data Visualization and Reporting

Syllabus: NEP 2020 w.e.f 2025-26

Sr. No.	Content	Hours
1	**Exploratory Data Analysis** • Perform exploratory data analysis on a cricket dataset, analyzing variables such as number of matches, runs, not outs, wickets, etc. • Visualize the distribution of player performance metrics using histograms, box plots, or scatter plots. • Investigate the relationship between player age and performance metrics using correlation analysis.	
2	**Batting Performance Analysis** • Analyze batting performance in a Cricket dataset, calculating metrics such as batting average, strike rate, and runs scored by players. • Identify top-performing batsmen based on performance metrics and compare their performance against different opponents or in specific conditions.	
3	**Bowling Performance Analysis** • Analyze bowling performance in the Cricket dataset, calculating metrics such as bowling average, economy rate, and wickets taken by players. • Identify top-performing bowlers based on performance metrics and analyze their performance against different teams or in various match situations	
4	**Performance Comparison** • Compare the scoring averages of top-performing batsman in different seasons. • Analyze the runs scoring (strike rate of Batting) of players from various teams in a specific league	
5	**Player Position Analysis** Calculate position-specific performance metrics and compare players within each position.	
6	**Injury Analysis** • Investigate the relationship between player injuries and their subsequent performance using historical injury and performance data. • Identify patterns and trends in the data to determine the impact of injuries on player performance and team success.	
7	**Team Analysis** • Analyze the impact of toss on a team's overall scoring and winning percentage. • Study the relationship between batting averages of players and their team's win-loss record	
8	**Sports Revenue Analysis** • Analyze revenue generation in sports organizations by examining factors such as ticket sales, merchandise	

	sales, and sponsorship deals. • Identify key drivers of revenue and provide recommendations for maximizing financial performance	
9	**Predictive Modeling** • Build a regression model to predict the number of runs scored by players based on their historical performance data. • Develop a classification model to predict the outcome of match based on team's statistics	
10	**Visualization and Reporting: (Mini-Project)** Prepare a comprehensive report summarizing the findings of the analysis and providing actionable insights for sports teams or organizations.	
	Total Hours	60

Sheth T. J. Education Society's
Sheth N.K.T.T College of Commerce and
Sheth J.T.T College of Arts, (Autonomous)
Thane (W)

Programme Name: **M.Sc (Data Science)**	**Semester: I**
Course Category/Vertical: Mandatory	
Name of the Dept: **Science and Technology**	
Course Title:**Research Methodology**	
Course Code: MDRM112	Course Level:6.0
Type: Theory	
Course Credit: 4 credits	
Hours Allotted: 60 Hours	
Marks Allotted: 100 Marks	

Course Objectives:

1. To develop the ability to formulate meaningful research problems and design appropriate data-driven research methodologies, including sampling strategies and data collection techniques relevant to Data Science.
2. To equip students with skills in data acquisition, preprocessing, management, and large-scale data analytics, ensuring data quality, security, and ethical handling.
3. To enable students to apply statistical methods, machine learning, deep learning, and big data tools to analyse datasets and extract actionable insights.
4. To prepare students for scholarly communication and responsible research practice, including research paper writing, presentation skills, manuscript submission, peer-review understanding, and adherence to ethical and data privacy regulations.

Course Outcomes (OC):

Upon completing this course, the student will be able to:

CO1: Formulate research problems, design appropriate research methodologies, and select suitable sampling and data collection techniques for data-driven research in Data Science.

CO2: Collect, preprocess, manage, and analyze datasets using statistical, machine learning, deep learning, and big data analytics techniques to extract meaningful insights.

CO3: Communicate research findings effectively through research papers, presentations, and scholarly publications while understanding the manuscript submission and peer-review process.

CO4: Apply ethical principles, data privacy regulations, and security practices to ensure responsible and secure conduct of research involving sensitive and large-scale datasets.

Syllabus: NEP 2020 w.e.f 2025-26

Unit No.	Content	Hours
I	**Unit 1: Introduction to Research Methodology for Data Sciences** Understanding Research in Data Sciences: Definition and significance of research in Data Sciences, Types of research: Basic research, applied research, and practical research, Research approaches: Quantitative vs. Qualitative, The role of research in Data Sciences. **Research Process and Ethics:** Steps in the research process in Data Sciences, Research ethics and responsible conduct, Formulating research questions and objectives, Identifying research problems in Data Sciences, Ethical considerations in research, including data privacy and security. **Research Design and Sampling for Data Sciences:** Principles of research design in Data Sciences, Types of research design, Sampling techniques in Data Sciences, Choosing the appropriate research design for data-driven research, Ethical considerations in sampling and data collection, including informed consent.	
II	**Unit 3: Data Analysis and Modeling for Data Sciences** Statistical Analysis for Data Sciences: Inferential statistics for hypothesis testing, Regression analysis and correlation, Machine learning concepts for predictive modeling. Machine Learning and Deep Learning: Introduction to machine learning and deep learning algorithms, Model selection and evaluation, **Big Data Analytics:** Understanding big data and its challenges, Tools and frameworks for big data processing (e.g., Hadoop, Spark), Analyzing and extracting insights from large datasets. **Unit 4: Research Communication, Publishing, and Data Privacy** Research Paper Writing and Presentation: Structure of research paper, Writing the abstract, introduction, literature review,	

	methodology, and results sections, Effective research paper presentations. **Publishing Research in Data Sciences:** Choosing the right conferences and journals, The peer-review process in Data Sciences, Preparing manuscripts for submission, Ethical considerations in publishing, including data privacy and security in publications. Ethical Considerations in Data Privacy and Security: Data privacy regulations and compliance, Ethical considerations in data anonymization and de- identification, Securing research data and protecting sensitive information.	

References:

1. Research Methodology: Methods and Techniques, C.R. Kothari, New Age Internatio

Evaluation Scheme

Theory courses of 4 credits: Total marks 100. Out of the total, 50% each for internal and external evaluation.

A. Internal Evaluation (30m + 10m + 10m = 50 Marks)

The internal assessment marks shall be awarded as follows:

1. 30 marks (Any one of the following):

a. Written Test of 30 Marks

b. SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or

c. NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or

d. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)

e. Certification marks of one completed exam shall be awarded to one course only. For four courses, the students will have to complete four certifications.

(Note: Only those certification/courses suggested by the department shall be deemed valid, Student cannot do any certification on their own)

2. 10 marks

10 marks from every course (Two 4 credits mandatory courses, one 2 credits mandatory course, one 4 credits elective course) coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved / Other Journal with plagiarism less than 15%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.

3. 10 marks

Open Book examination based on problem solving related to the respective subject.

Suggested format of Question paper of 20 marks for the written test.

Q1. Attempt any two of the following: 16 marks

a.

b.

c.

d.

Q2. Attempt any two of the following: 14 marks

a.

b.

c.

d.

B. External Examination: (50 marks) Duration : 2 hrs

All questions are compulsory

Q1 (Based on all units) Attempt any two of the following: 10 marks

a. Unit 1

b. Unit 2

c. Unit 3

d. Unit 4

Q2 (Based on Unit 1) Attempt any two of the following: 10 marks

Q3 (Based on Unit 2) Attempt any two of the following: 10 marks

Q4 (Based on Unit 3) Attempt any two of the following: 10 marks

Q5 (Based on Unit 4) Attempt any two of the following: 10 marks

Theory courses of 2 credits: Total marks 50. Out of the total, 50 % each for internal and external evaluation.

A. Internal Evaluation (25 Marks)

The internal assessment marks shall be awarded as follows:

1. 10 marks from every course (Two 4 credits mandatory courses, One 2 credits mandatory course, One 4 credits elective course) coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved / Other Journal with plagiarism less than 15%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.
2. 10 marks - Open Book examination based on problem solving related to the respective subject.
3. 5 marks - Assignment/Group discussion.

B. External Examination: (25 marks) Duration : 1 hr

All questions are compulsory

Q1 (Based on Unit 1) Attempt any two of the following: 13 marks

Q2 (Based on Unit 2) Attempt any two of the following: 12 marks

Practical courses of 2 credits: Total marks 50. Out of the total, 50 % each for internal and external evaluation.

A. Practical Evaluation Internal (25 marks)

1. Performance during all practical sessions 10
2. Problem solving with the acquired programming skills 10
3. Viva Voce 05

B. Practical Evaluation External (25 marks)

A Certified copy of hard-bound journal is essential to appear for the practical examination.

1. Practical Question 15
2. Journal 05
3. Viva Voce 05

Practical courses of 4 credits: Total marks 100. Out of the total, 50 % each for internal and external evaluation.

A. Practical Evaluation Internal (50 marks)

1. Performance during all practical sessions 20
2. Problem solving with the acquired programming skills 25
3. Viva Voce 05

B. Practical Evaluation External (50 marks)

A Certified copy of hard-bound journal is essential to appear for the practical examination.

1. Practical Question 20
2. Practical Question 20
3. Journal 05
4. Viva Voce 05