B. K. Birla College of Arts, Science and Commerce, Kalyan

(Empowered Autonomous Status)

(Affiliated to University of Mumbai)

Syllabus for M.Sc. Program as per

National Education Policy (NEP) 2020 framework:

M.Sc. Data Science and Big Data Analytics

In association with

Program Code: Information Technology

(Credit Based Semester and Grading System
Academic year 2024–2025)

Preamble:

In today's world there is data available in abundance from variety of sources like web server logs, social media, and large databases and from diverse domains like Ecommerce, Medical, Scientific etc. Big data analytics is the process of examining these data to uncover hidden patterns, unknown correlations and other useful information that can be used to make better decisions. Engineers, Business personnel, Doctors, Scientists etc. can use this to improve their services.

This course is designed to give students a comprehensive understanding of data science principles and techniques and the ability to analyze and interpret large volumes of complex data. In this era of digital transformation, organizations across industries are seeking skilled professionals who can leverage the power of data to drive insights, make informed decisions, and gain a competitive edge.

Throughout this program, students will delve into the world of data science and big data analytics, exploring various methodologies, tools, and technologies used to extract valuable knowledge from massive datasets. The curriculum is designed to equip students with both theoretical knowledge and practical skills, enabling them to navigate the entire data science lifecycle – from data collection and preprocessing to analysis, modeling, and visualization.

The program emphasizes a hands-on approach, allowing students to gain practical experience through real-world case studies, industry projects, and internships. They will have the opportunity to work with state-of-the-art tools and technologies, such as Python, R, Hadoop, Spark, and machine learning algorithms, to solve complex data challenges and uncover meaningful insights.

In addition to technical expertise, the program also focuses on developing critical thinking, problem-solving, and communication skills. Students will learn to effectively communicate their findings and insights to diverse stakeholders, making data-driven recommendations that drive business success.

Upon successful completion of this program, students will be equipped with the knowledge and skills to pursue various career paths in data science and big data analytics. They will be prepared to work in industries such as finance, healthcare, marketing, e-commerce, and more, where the ability to harness the power of data is crucial for making strategic decisions and driving innovation.

We are excited to embark on this data science journey with you, and we look forward to equipping you with the skills and knowledge needed to thrive in the dynamic world of data analytics. Get ready to explore the vast landscape of data, uncover hidden insights, and make a meaningful impact with your data-driven expertise.

Eligibility:

To secure entry into the MSc DSBDA program in any college, one needs to meet the following criteria:
- ▪ Applicants for the MSc DSBDA program must have earned a bachelor's degree.
- ▪ Candidates must have passed their Bachelor's degree in B.Sc. (I.T. / C.S. / Mathematics / Statistics) / B.E. (I.T. /C.S.) / B.Tech. (I.T. /C.S.) / BCA and any other equivalent degree
- ▪ M.Sc. DSBDA admission will be based on merit score and Personal Interview.

Duration:

Qualification Title	Credit Requirement		Semester	Year
	Minimum	**Maximum**		
PG Diploma	40	44	2	1
PG Degree	40	44	4	2

Program Outcome

- ● **Proficiency in Data Analysis and Interpretation:** Graduates will possess advanced skills in analyzing and interpreting large and complex datasets. They will be proficient in applying statistical methods, data mining techniques, and machine learning algorithms to extract meaningful insights and patterns from data.
- ● **Expertise in Data Visualization:** Graduates will demonstrate expertise in visualizing data effectively to communicate insights and findings. They will be skilled in using data visualization tools and techniques to present complex information in a clear and visually appealing manner.
- ● **Proficiency in Machine Learning and Predictive Analytics:** Graduates will have a solid foundation in machine learning algorithms and predictive analytics techniques. They will be able to apply these methods to build models that can make accurate predictions and support data-driven decision-making.
- ● **Knowledge of Big Data Technologies and Tools:** Graduates will be familiar with the tools and technologies used for handling big data. They will have practical experience working with distributed computing frameworks (such as Hadoop and Spark), NoSQL databases, and cloud-based data platforms.
- ● **Skills in Data Preprocessing and Feature Engineering:** Graduates will possess skills in data preprocessing and feature engineering. They will be able to clean and transform raw data, handle missing values, and engineer relevant features for machine learning models.
- ● **Proficiency in Programming and Data Manipulation:** Graduates will be proficient in programming languages commonly used in data science, such as Python or R. They will have expertise in data manipulation, including data cleaning, data integration, and data transformation.
- ● **Understanding of Data Privacy and Ethical Considerations:** Graduates will have a sound understanding of data privacy regulations and ethical considerations in data science. They will be aware of the importance of handling data responsibly, ensuring

data privacy and security, and adhering to ethical guidelines in data analysis.

- **Effective Communication and Collaboration:** Graduates will possess effective communication and collaboration skills. They will effectively communicate their findings and insights to technical and non-technical stakeholders. They will also be adept at working in interdisciplinary teams and collaborating with professionals from different domains.
- **Problem-Solving and Critical Thinking:** Graduates will demonstrate strong problem-solving and critical thinking skills in the context of data science and big data analytics. They will be able to analyze complex problems, identify appropriate methodologies, and propose effective solutions based on data-driven insights.
- **Lifelong Learning and Adaptability:** Graduates will recognize the importance of lifelong learning and adaptability in a rapidly evolving field. They will be equipped with the skills and mindset to stay updated with emerging trends, technologies, and methodologies in data science and big data analytics.

Syllabus as per NEP 2020

Sem	Major (Credits-14)	Electives (Credits- 4)	Minor (Credits – 4)	OJT (Credits – 4)	Total
I	**Course- I** Applied Statistics with Excel **Course -II** Data Visualization using Tableau and Power BI **Course -III** Data on Cloud	SPARK Technologies OR Retail Marketing Analytics	Research Methodology		22
II	**Course -I** Practical Approach to Data Mining & Analytics **Course -II** Machine Learning **Course -III** Optimization and Simulation for Data Science	Advance Data Structures and algorithm OR Sentiment, Web and Text Analytics	-	Internship with Project	22

Sem	Major (Credits-14)	Electives (Credits- 4)	Minor (Credits – 4)	OJT (Credits – 4)	Total
III	**Course- I** Deep Learning and Neural Networks **Course -II** Big Data Analytics-Advanced **Course -III** Ethics and Governance issues in Big Data	Data Science for Agriculture OR Econometrics & Finance OR Internet of Everything	-	Research and Academic Paper Writing	22
IV	**Course -I** Applied Artificial Intelligence **Course -II** Predictive Analytics	Blockchain Technologies for Data Science OR Financial Technologies OR Social Media Analytics	-	Research based Project	22

Course Code	Applied Statistics with Excel			L	T	P	C
				4	-	2	6
Pre-requisites	Basic Mathematics & Statistics concept		**Semester**			**I**	

Course Objectives:

1. Understand the fundamental concepts of descriptive and inferential statistics
2. Gain hands-on experience in using Excel to perform statistical analysis
3. Master the concepts of probability, random variables, and probability distributions
4. To apply inferential statistical methods such as hypothesis testing, confidence interval estimation, and sampling techniques
5. Understand and apply advanced statistical techniques

Course Outcomes (CO):	
CO 1	Understand the scope and necessity of Statistics.
CO 2	Tabulate and represent the data in diagrams and graphs.
CO 3	Apply the formula and calculate descriptive measures of statistics.
CO 4	Analyze the nature of data and interpret the measures
CO 5	Analyze the data and predict the future values using curve fitting.

UNIT I	
Introduction to Data Science and Applied Statistics	

Overview of Data Science: Definition and scope of data science, The data science process, Roles and responsibilities of a data scientist

Fundamentals of Statistics: Descriptive statistics: measures of central tendency, dispersion, and shape, Probability theory and distributions, Inferential statistics: hypothesis testing, confidence intervals

Descriptive Statistics Introduction to statistics and data, Ways of classifying data, levels of measurement , Critical thinking skills, Charts and graphs: Frequency distributions, bar charts, stem and leaf plots, Describing a distribution: shape, centre, spread, Measures of centre: mean, median, mode, midrange, Measures of spread: range, variation, variance, standard deviation, interquartile range,Measures of shape: Empirical rule, Chebyshev's rule, skewness, kurtosis ,Measures of relative position: quartiles, percentiles, midquartile

UNIT II

Probability Fundamentals of probability, Probabilities from frequency tables, Unions and intersections, Addition rule for "or" and multiplication rule for "and", Tree diagrams, Conditional probabilities, counting techniques, Random variables, Mean, variance, and standard deviation of a discrete random variable, Binomial and multinomial distributions, Mean, variance, and standard deviation for binomial distribution, Standard normal distribution. Finding areas from z-scores and z-scores from areas, Applications of the normal distribution. Converting from and to raw scores

UNIT III

Inferential Statistics Randomization testing, Types of sampling and sampling errors, Sampling distributions and the Central Limit Theorem, Student's t distribution, Sampling distributions for proportions, Confidence intervals for the population mean and population proportion, Hypothesis testing fundamentals, Probability value approach comparing p-value to significance level, Confidence interval approach comparing claimed value to confidence interval, Types of errors, significance level, p-value, Decisions vs conclusions, Testing a claim about a single population mean and proportion, Testing a claim about two population means and proportion, Paired samples t-test

UNIT IV

Advanced Inferential Statistics Linear correlation, Hypothesis test for correlation, Regression analysis, Linear & Logistic regression, finding regression equation from summary statistics and correlation coefficient, Explained, unexplained, and total deviations, Coefficient of determination, Table of coefficients and Analysis of Variance. F distribution., Multiple regression, adjusted R squared, Chi-square distributions, Chi-square goodness of fit test (multinomial experiments), Chi-square test for independence, test for homogeneity, One-Way Analysis of Variance • Two-Way Analysis of Variance, Interpreting ANOVA results.

Hypothesis Testing Basics of Hypothesis, Testing Type of test and Rejection Region, Type 0 errors -Type 1 Errors, Type 2 Errors, P value method, Z score Method

Predictive Modelling Introduction to Predictive Modelling using Decision trees, Formulate the model and estimate the parameters, check prediction accuracy

References:
1. Applied Multivariate Statistical Analysis – Johnson & Wichern, Pearson Publications, 6th Edition
2. Applied Business Statistics – Ken Black, Wiley Publication, 7th Edition
3. Probability for Engineers & Scientists by Walpole, Pearson Publications, 9th Edition
4. Introduction to Probability & Statistics – Mendenhall, Beaver
5. Predictive Analytics by Eric Siegel, Wiley Publication, Revised Edition

Course Code	Applied Statistics with Excel practical		L	T	P	C
			4	-	2	4
Pre-requisites	Basic Mathematics & Statistics concept	**Semester**		I		

Course Objectives:

1. Gain proficiency in using essential Excel functions for data manipulation, organization, and basic statistical analysis.
2. Utilize Excel to compute and interpret key descriptive statistics
3. Apply Excel to model data with various probability distributions and use the normal distribution to calculate probabilities and make data-driven estimates.
4. Understand and apply the Central Limit Theorem
5. To perform regression analysis, enabling the prediction of dependent variables based on independent variables, and validating these models through practical application.

Course Outcomes (CO):

CO 1	To use Excel for organizing, analyzing, and interpreting datasets through various functions and formulas.
CO 2	To compute, analyze, and present descriptive statistics
CO 3	To model data using probability distributions and apply the normal distribution for probability estimation and decision-making.
CO 4	To apply the Central Limit Theorem, construct confidence intervals, and perform hypothesis tests
CO 5	To develop and interpret regression models using Excel,

Practical List –

1. Basic Excel Functions
2. Understand data using descriptive statistics in Excel, including frequency distributions and measures of central tendency & variability
3. Model data with probability distributions, and use the normal distribution to calculate probabilities and make value estimates using Excel formulas
4. Introduce the Central Limit Theorem, which leverages the normal distribution to make inferences on populations with any distribution
5. Make estimates with confidence intervals, which use sample statistics to define a range where an unknown population parameter likely lies
6. Draw conclusions with hypothesis tests, which let you evaluate assumptions about population parameters using sample statistics
7. Make predictions with regression analysis, and estimate the values of a dependent variable via its relationship with independent variables
8. A mini-Project based on the study

Course Code	Data Visualization using Tableau and Power BI	L	T	P	C
		4	-	2	6
Pre-requisites	Basic knowledge of programming fundamentals	**Semester**		I	

Course Objectives:
1. Develop a foundational understanding of the importance of data visualization in business analytics and its impact on decision-making processes.
2. Gain proficiency in using Tableau to connect to various data sources, create and customize visualizations, and build interactive dashboards to effectively represent data insights.
3. Learn to clean, transform, and model data using Tableau and Power BI, preparing it for accurate and meaningful analysis.
4. Apply advanced Tableau features such as calculated fields, table calculations, and parameter-driven what-if analyses to enhance the depth and accuracy of data interpretations.
5. Acquire the skills to use Power BI for data import, transformation, and visualization

Course Outcomes (CO):	
CO 1	Skill building for identifying and working with business-oriented data sets.
CO 2	Importing and preparing for loading data models.
CO 3	Identifying data visualization types and core purpose
CO 4	Creating functional Power BI dashboards and reports.

UNIT I			

Introduction to Data Visualization Visualization an aspect of business analytics, importance of data visualization

Working with Tableau Tableau Architecture, Tableau repository, Connecting to Data & Introduction to data source concepts, Understanding the Tableau workspace, Dimensions and Measures, Data Types & Default Properties, Tour of Shelves & Marks Card, building basic views, Saving and Sharing your work-overview

UNIT II			

Building Views (Reports) – Date Aggregations and Date parts, Cross tab & Tabular charts, Totals & Subtotals, Bar Charts & Stacked Bars, Line Graphs with Date & Without Date, Tree maps, Scatter Plots, Individual Axes, Blended Axes, Dual Axes & Combination chart, Edit axis, Parts of Views, Sorting, Trend lines, Reference Lines, Forecasting, Filters, Context filters, Sets - In/Out Sets & Combined Sets, Grouping, Bins/Histograms, Drilling up/down – drill through Hierarchies, View data & Actions (across sheets), latitude and longitude, Default location/Edit locations, Symbol Map & Filled Map.

Calculated Fields Working with aggregate versus disaggregate data, Explain - #Number of Rows, Basic Functions (String, Date, Numbers etc), Usage of Logical conditions

	Table calculations Explain scope and direction, Percent of Total, Running / Cumulative calculations		
	Parameters Create What-If analysis, Using Parameters in Calculated fields, Bins, Reference Lines, Filters/Sets, Display Options (Dynamic Dimension/Measure Selection)		
UNIT III			

Building Interactive Dashboards- (Building & Customizing) Combining multiple visualizations into a dashboard (overview), Making your worksheet interactive by using actions Filter URL, Highlight, Options in Formatting your Visualization, Working with Labels and Annotations, Effective Use of Titles and Captions.

Working with Data Multiple Table Join, Data Blending, Difference between joining and blending data, and when we should do each, working with the Data Engine / Extracts, Working with Custom SQL, Toggle between to Direct Connection and Extracts

Working with Tableau Server – Accessing reports through web, Publishing to Tableau Server – Overview of publishing, Server Administration - Managing Users, Projects & Object level and Data Security as per Users, User Filters

UNIT IV			

Introduction to Power BI: Overview of Power BI, Power BI components: Power BI Desktop, Power BI Service, Power BI Mobile, Installation and setup of Power BI Desktop, Understanding the Power BI interface

Data Sources and Data Loading: Connecting to various data sources (Excel, CSV, databases, web, etc.), Importing data into Power BI, DirectQuery vs. Import mode, Data transformation basics using Power Query Editor

Data Cleaning and Transformation: Using Power Query Editor for data cleaning, handling missing data, removing duplicates, Data transformation techniques: merging, appending, pivoting, unpivoting, Creating and managing relationships between tables]

Data Modeling and DAX: Introduction to data modeling concepts, Creating calculated columns and measures, Introduction to DAX (Data Analysis Expressions), Common DAX functions for data analysis (SUM, AVERAGE, COUNT, etc.), Using DAX for complex calculations and aggregations

Data Visualization: Introduction to Power BI visualizations, Creating and customizing basic visualizations (charts, tables, maps), Using slicers and filters, Custom visuals in Power BI, Best practices for effective data visualization, Publishing reports to Power BI Service, Creating and managing dashboards, Sharing reports and dashboards with others.

Reference Books

1. Practical Tableau by Ryan Sleeper, OReilly Media, 2018

2. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018

3. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

4. Learning Tableau 10 – Second Edition, by Joshua Milligan
5. Practical Tableau by Ryan Sleeper

Course Code	Data Visualization using Tableau and Power BI Practical	L	T	P	C

			4	-	2	4
Pre-requisites		**Semester**			**I**	

Course Objectives:

1. Develop a strong understanding of Tableau and Power BI interfaces
2. Gain hands-on experience in creating diverse visualizations
3. Learn to import, clean, and manage data using Tableau and Power BI, mastering data joining, relationships, and transformations to prepare data for analysis.
4. Acquire the skills to create and customize interactive dashboards in Tableau and Power BI
5. Learn to craft compelling data stories by combining multiple visualizations and creating comprehensive reports, dashboards, and data stories using Tableau and Power BI.

Course Outcomes (CO):

CO 1	To navigate and effectively use the interfaces of Tableau and Power BI for data analysis and visualization tasks.
CO 2	To implement a variety of visual encodings, including charts, maps, and graphs, to represent data accurately and intuitively.
CO 3	To exhibit proficiency in importing, cleaning, and transforming data, ensuring that it is correctly structured and ready for detailed analysis.
CO 4	To design and develop interactive dashboards in Tableau and Power BI, enabling dynamic data exploration and enhancing the user experience.
CO 5	To create comprehensive data stories and reports, effectively communicating insights and supporting decision-making processes through well-structured visualizations and dashboards.

Practical List –
1. Getting acquainted with Tableau interface
2. Implement Visual Encodings, Bar charts, Pie chart, Line chart, Multiple chart and distribution
3. Implement Highlight tables, Scatterplot, Trendline
4. Import and manage data (join, relationship, replace)
5. Master Data Visualization
6. Implement Heatmap, Geographic mapping, impressive bar chart, bullet graph
7. Implement Gantt chart, data calendar, circle view, general operation
8. Create master dashboard
9. Adding filters and quick filters to dashboards
10. Create a data story in Tableau
11. Power BI - Importing data, Cleaning and Transformation
12. Data Modelling and implementation of DAX
13. Create reports using Power BI.
14. Perform data visualization using Power BI.
15. A mini project on any use case.

Course Code	**Data on Cloud**	**L**	**T**	**P**	**C**

			2	-		2
Pre-requisites		**Semester**		**I**		

Course Objectives:	

- Understand Cloud Computing Concepts
- Explore Cloud-based Data Storage Solutions
- Master Data Processing on the Cloud
- Ensure Data Security and Privacy on the Cloud
- Implement Cloud-based Data Integration and Migration

Course Outcomes (CO):

CO 1	Recall the key concepts and principles of cloud computing and its application in data management.
CO 2	Understand the distinct types of cloud-based data storage and processing services and their functionalities. Comprehend the security and privacy challenges associated with data on the cloud.
CO 3	Apply cloud-based data storage solutions to effectively store and retrieve large volumes of data. Utilize cloud-based data processing tools and technologies to analyze and transform data.
CO 4	Analyze the performance and scalability of cloud-based data solutions. Evaluate data security measures and propose strategies to protect data on the cloud.
CO 5	Design data integration and migration strategies for seamless transfer of data to and from cloud platforms.
CO 6	Create innovative data solutions using cloud resources that meet specific business requirements

UNIT-1	**Introduction to Cloud Computing and Data Management**		

Overview of cloud computing and its relevance in data science and big data analytics
Cloud service models (IaaS, PaaS, SaaS) and deployment models (public, private, hybrid)
Cloud-based data storage options and architectures, Cloud-based data management tools and technologies, Data security and privacy considerations in cloud environments
Object storage systems and services (e.g., Amazon S3, Google Cloud Storage), Cloud-based file systems (e.g., Hadoop Distributed File System, Azure Blob Storage), Cloud-based databases (e.g., Amazon RDS, Google Cloud SQL, Microsoft Azure Cosmos DB)
, Data migration and synchronization techniques between on-premises and cloud storage

UNIT-2	**Cloud-based Data Storage and Retrieval**		

Distributed computing frameworks (e.g., Apache Hadoop, Apache Spark) for big data processing, Cloud-based data processing services (e.g., Amazon EMR, Google Cloud Dataproc), Data preprocessing and feature engineering techniques on the cloud, Introduction to distributed data processing languages (e.g., Apache Pig, Apache Hive)
Virtualization system-specific attacks Guest hopping, attacks on the VM (delete the VM, attack on the control of the VM, Code or file injection into the virtualized file structure), VM migration attack, hyper jacking., Technologies for virtualization-based security enhancement IBM security virtual server protection, virtualization-based sandboxing. Storage Security- HIDPS, log management, Data Loss Prevention, Location of the Perimeter.

Books:

1.	"Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Viktor Mayer-Schönberger and Kenneth Cukier
2.	"Cloud Computing: Concepts, Technology, and Architecture" by Thomas Erl, Ricardo Puttini, and Zaigham Mahmood
3.	"Cloud Computing: Principles and Paradigms" by Rajkumar Buyya, James Broberg, and Andrzej Goscinski
4.	"Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett
5.	Data Analysis in the Cloud by Fabrizio Marozzo, Paolo Trunfio, Domenico Talia
6.	"Big Data Analytics: Turning Big Data into Big Money" by Frank J. Ohlhorst

Elective 1

Course Code	SPARK Technologies		L	T	P	C
			2	-	2	4
Pre-requisites		Semester		I		

Course Objectives:

1. Gain a comprehensive understanding of the components and features of the Spark
2. To create, load, and save RDDs, and perform transformations and actions
3. Acquire practical experience in creating and deploying Spark applications
4. Delve into advanced Spark functionalities such as Spark Streaming, MLlib for machine learning, Spark SQL for structured data processing, and GraphX for graph computations, broadening the scope of data analytics.
5. Learn about Spark performance tuning, including RDD persistence and distributed persistence, to optimize Spark applications and ensure efficient data processing and resource utilization.

Course Outcomes (CO):

CO 1	Understand the concept of SPARK Technologies and its implementation
CO 2	Understand the concept of RDD
CO 3	Understand the implementation of SPARK SQL, GraphX, Performance Tuning.
CO 4	Use Sparks Resilient Distributed Datasets to process and analyze large data sets across many CPUs
CO 5	Understand how the GraphX library helps with network analysis problems
CO 6	Understand how Spark SQL lets you work with structured data

UNIT I			

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

UNIT II			

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.
Reference Books: 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

Course Code	SPARK Technologies Practical		L	T	P	C
			2	-	2	4
Pre-requisites	Basic understanding of programming language (python) and Data mining techniques	**Semester**		I		

Course Objectives:
1. Develop practical skills in using Spark's Scala and Python shells 2. Learn to perform various transformations and actions on RDDs 3. Acquire the ability to integrate Spark with Hadoop's HDFS, handle sequence files, and manage RDD persistence to optimize data storage and processing across distributed environments. 4. Gain hands-on experience with Spark's advanced libraries

Course Outcomes (CO):	
CO 1	Understand the concept of SPARK Technologies and its implementation
CO 2	Understand the concept of RDD
CO 3	Understand the implementation of SPARK SQL, GraphX, Performance Tuning

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 a Data Frame
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)

Course Code	Retail Marketing Analytics	L	T	P	C

			2	-	2	4
Pre-requisites	Knowledge of statistics and mathematical concepts	**Semester**			I	

Course Objectives:

1. Develop a foundational understanding of how marketing analytics can enhance decision-making processes, optimize marketing strategies, and improve overall business outcomes.
2. Acquire practical skills in cleaning, tabulating, summarizing, and visualizing marketing data
3. Gain proficiency in designing and conducting marketing experiments
4. Learn to assess the outcomes of marketing experiments using hypothesis testing

Course Outcomes (CO):

CO 1	Understand the importance of marketing analytics for forward looking and systematic allocation of marketing resources
CO 2	Know how to use marketing analytics to develop predictive marketing dashboard for organization
CO 3	Analyze data and develop insights from it to address strategic marketing challenges

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

UNIT II

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

Reference Books:
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

Course Code	Retail Marketing Analytics Practical		L	T	P	C
			2	-	2	4
Pre-requisites	Knowledge of statistics and mathematical concepts	**Semester**			I	

Course Objectives:
1. Learn to effectively tabulate, summarize, and visualize marketing data
2. Acquire the skills to design and conduct experiments, perform hypothesis testing, and apply power calculations
3. Gain expertise in calculating and predicting CLV using various predictive modeling techniques, and apply cohort analysis to segment customers based on behavior for targeted marketing.
4. Develop and apply methods for customer segmentation

Course Outcomes (CO):

CO 1	To Learn working and analyzing with marketing data
CO 2	To develop predictive marketing dashboard for organization
CO 3	Understand the concept of hypothesis testing and its role in assessing experiment outcomes

Note: Being able to approach data using statistical software is one of the essential goals of this class. You are required to use R for all assignments and projects throughout this course. Completing homework and quizzes using Excel or other programming languages is not accepted. Programming knowledge prior to the class is preferred, but not required. As we spend time in class to familiarize you with the RStudio interface and basic functions in the first few weeks of
the class, take this time to ask questions and adapt to R as soon as possible.

- Download R from http://cran.r-project.org/
- Download R Studio from http://www.rstudio.com/products/rstudio/download/

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 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 Beautiful Soup 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, clickthrough rates, or revenue.
    - Interpret the results and provide recommendations for optimizing marketing campaigns based on the findings.

Course Code			L	T	P	C
	Research Methodology		2		2	4
Pre-requisites		**Semester**	**I**			

Course Objectives:

1. To identify and discuss the role and importance of research in the social sciences.
2. To identify and discuss the issues and concepts salient to the research process.
3. To identify and discuss the complex issues inherent in selecting a research problem, selecting an appropriate research design, and implementing a research project.
4. To identify and discuss the concepts and procedures of sampling, data collection, analysis and reporting.

CO 1	Understand the fundamentals of research
CO 2	Review various literature sources
CO 3	Collect data for research
CO 4	Interpret results and write a research paper
CO 5	Apply various digital tools for conducting research

UNIT-1	**Introduction to Research in Computing**		

Introduction-The 6 Ps of Research, purpose and product of research, Research Domains from Information Systems and recent Computing disciplines, Study of research carried out in different disciplines through the Research Papers, Finding and choosing research topics, Evaluating the purpose and products of research, Participants and research ethics, Research Ethics, Practical Work

UNIT-2	**Research Process**		

Model of research process, Literature Review, Open-Source Literature Review, Search Tools to find open-source articles. Research Question and conceptual framework, Formulating a

Research Question / Problem Statement, Tools and Strategies for answering the Research, Questions (Research Design), Practical work	
UNIT-3	**Use of Data Analysis for Research**
Quantitative Data Analysis, Types of quantitative data, sources of quantitative data, Tools for data analysis, Qualitative Data Analysis, Analyzing textual data, Analyzing non-textual qualitative data, Grounded theory, Computer aided qualitative analysis., Evaluating qualitative data analysis, Practical Work	
UNIT-4	**Writing and Presentation of Research**
Importance of publishing the research, Types of Research Publications and writing up the Research, Different formats of writing research, Conference paper presentation, Posters and exhibitions, Software demonstrations, Presenting yourself. Facing Vivas, Evaluating presentations for written and oral Presentations, Creating Your profile as a researcher Practical Work	
UNIT-5	**Tools / techniques for Research**
Use of tools/techniques for Research: methods to search required information effectively, Reference Management Software like Zotero. Research Ethics Meaning, Approaches to Research Ethics, Ethical Issues in Research, Measures to make research more ethical, Legal Aspects.	

References:

1. Researching Information Systems and Computing by Briony J. Oastes Sage Publications India Pvt. Ltd., New delhi, ISBN 1-4129-0224-X (pbk)
2. Your research Project, A Step-by-step Guide for the first-time researcher by Nicholas Walliman, Vistaar Publications (A division of Sage Publications), New Delhi, ISBN 81-7829-540-7
3. Research Methods by William M K Trochim Cornel University, Biztantra, An imprint of Dreamtech Press. WILEY- Dreamtech India Pvt. Ltd., ISBN 81-7722-372-0

Practical List:

1. Develop a research question and hypothesis based on a current trend in Artificial Intelligence. Explain why you chose this question and outline your deductive reasoning process.
2. Design an experimental setup to test your hypothesis from Question 1. Identify the independent and dependent variables. Describe the control measures you would put in place.
3. Explain how you would collect qualitative and quantitative data for your research. Discuss the pros and cons of using interviews, surveys, and observations in your study.
4. Discuss the validity and reliability of the measurement scales you intend to use for your research. Would you use nominal, ordinal, interval, or ratio scales? Justify your choice.
5. Describe the sampling method you would employ for your research. Explain how you intend to determine the sample size and ensure the sample's representativeness.
6. Use a dataset of your choice to perform univariate and bivariate analyses. Present your findings in the form of frequency tables, bar charts, or pie charts, and interpret the results.
7. After conducting a Chi-square test on your data, explain how you would interpret the results. Discuss the implications of your findings for the research question and hypothesis.
8. Describe the ethical considerations you will take into account during your research. How would you ensure the privacy and confidentiality of your participants?
9. Demonstrate the use of Reference Management Software like Zotero or Mendeley to manage your citations. Show how you would integrate citations into your research paper.
10. Based on your research findings, propose future research directions, draft a research paper and publish in UGC-CARE/Scopus Indexed Journals.