

MAHARSHI DAYANAND SARASWATI UNIVERSITY

AJMER



Syllabus

**Scheme of Examination and
Courses of Study**

Faculty of Science

MCA (Data Science)

Semester I and II w.e.f. 2026-27

Semester III and IV w.e.f. 2027-28

SCHEME OF TWO YEARS POST GRADUATE DEGREE PROGRAM

MCA (Data Science)

Program Code:

Program Faculty: Computer Science

Program Name: MCA (Data Science)

1. Introduction to the Program

The Master of Computer Applications in Data Science, MCA (Data Science) is an advanced, multi-disciplinary postgraduate program designed to bridge core computer science engineering with advanced statistical intelligence. The curriculum focuses on transforming high-volume, unstructured data into actionable strategic insights.

While conventional MCA tracks prioritize software development, this specialized program integrates deep computational data engineering, machine learning pipelines, cloud computing platforms, and business intelligence architectures. Students transition from building isolated applications to engineering data-driven ecosystems capable of automated pattern recognition, predictive forecasting, and massive scale optimization.

2. Objectives of the Program

Build Advanced Computational Competency: Develop an expert-level mastery of foundational programming frameworks, database engines, data warehousing schemas, and large-scale cloud operations.

Cultivate Mathematical Rigor: Instill a deep theoretical and practical understanding of statistical analysis, probability spaces, regression diagnostics, and optimization math required to configure machine learning networks.

Master the End-to-End Data Lifecycle: Train students to design and govern complete data pipelines, covering automated scraping, processing, feature engineering, and visual storytelling.

Deploy Scalable Intelligence: Enable students to implement parallel computing systems, distributed query architectures, and deep neural network models to process petabyte-scale data.

Instill Ethical Data Leadership: Prepare students to address issues of data privacy compliance, algorithmic governance, explainable AI, and mitigate algorithmic bias in global automated operations.

3. Employment and Entrepreneurial Scope

Employment Scope

Graduates of the MCA (Data Science) program are qualified for high-growth technical and strategic roles within global technology, healthcare, finance, retail, and security sectors, including:

Data Scientist / Machine Learning Engineer: Designing predictive models, fine-tuning deep learning networks, and embedding automated inference scripts into enterprise systems.

Big Data Architect / Data Engineer: Structuring cloud-native data lakes, orchestrating high-throughput ETL/ELT pipelines, and optimizing modern distributed lakehouses (Delta Lake/Iceberg).

Predictive Business Intelligence Developer: Formulating complex data models, writing advanced analytics syntax (DAX/M), and building intelligent executive reporting apps (Power BI/Tableau).

Cybersecurity Threat Analyst / Risk Modeler: Building unsupervised anomaly detection arrays to isolate zero-day attacks, insider threats, fraud rings, and structural credit risks.

Entrepreneurial Scope

The program builds a comprehensive toolkit for launching data-driven ventures and consulting practices:

AI Product Ventures: Founding software-as-a-service (SaaS) startups that deliver vertical AI solutions, real-time analytics engines, or automated text/vision analysis APIs.

Data Analytics Agencies: Establishing specialized B2B consulting firms that engineer customized data warehousing, modern lakehouse scaling, and dynamic dashboard infrastructures for enterprises.

Quantitative Product Operations: Launching platform-centric micro-ventures utilizing predictive revenue management, dynamic pricing engines, or hyper-localized urban and retail intelligence models.

4. Learning Outcomes of the Program (PO)

Upon successful completion of the MCA (Data Science) program, graduates will demonstrate the following Program Outcomes:

- (PO-1) Advanced Technical Knowledge: Apply foundational computing concepts across computer systems architecture, database management (DBMS), and open-source operating systems.
- (PO-2) Problem Analysis & Algorithmic Optimization: Analyze complex computation problems, evaluating the space and time complexities of recursive and non-recursive algorithms to optimize processing limits.
- (PO-3) Full-Stack Design & Development: Architect scalable, multi-tier web and enterprise systems utilizing structured design principles, clean data schemas, and object-oriented paradigms.
- (PO-4) Mobile Ecosystem Engineering: Build secure, data-driven mobile computing solutions that interact with external web services and manage platform background states seamlessly.
- (PO-5) Modern Tool & Infrastructure Usage: Configure and manipulate industry-standard toolchains including the Android SDK, Cisco switching environments, Git/Linux environments, and Figma layout grids.
- (PO-6) Intelligent Systems & Analytics Modeling: Extract knowledge patterns from high-density datasets by constructing data mining pipelines, machine learning matrices, and generative prompt spaces.
- (PO-7) Research Inquiry & Problem Scoping: Conduct systemic literature reviews, design test scripts, isolate driver variables, and execute an independent scientific dissertation.
- (PO-8) Communication & Documentation: Draft precise Software Requirements Specifications (SRS), build clear project wireframes, and deliver professional technical defenses of complex software assets.

SCHEME AND WEIGHTAGE OF COURSES

Semester I

S. No.	Course Category	Title of the Course	Credits			Credits earned	Evaluation			Remarks
			L	T	P		IA	ESA	Total	
SEMESTER – I										
1	mcads-101 DCC – 01	Mathematical foundations of Data Science	3	1	0	4	30	70	100	All Courses are compulsory Courses
2	mcads-102 DCC – 02	Database Technology	3	1	0	4	30	70	100	
3	mcads-103 DCC – 03	Object Oriented Programming using Java	3	1	0	4	30	70	100	
4	mcads-104 DCC – 04	Data Visualization using Tableau	3	1	0	4	30	70	100	
5	mcads-105 DCC – 05	Lab - Database Technology and Object-Oriented Programming using Java	0	0	4	4		100	100	
		Total				20			500	

Semester II

S. No.	Course Category	Title of the Course	Credits			Credits earned	Evaluation			Remarks
			L	T	P		IA	ESA	Total	
SEMESTER - II										
6	mcads-201 DCC – 06	Data-Driven Artificial Intelligence	3	1	0	4	30	70	100	
7	mcads-202 DCC – 07	Python for Data Science	3	1	0	4	30	70	100	
8	mcads-203 DCC – 08	Data Warehousing and Data Mining	3	1	0	4	30	70	100	
9	mcads-204a DSE/GEC– 01A	Data Science for Web Intelligence	3	1	0	4	30	70	100	Student may opt any one
	mcads-204b DSE/GEC – 01B	Data Analytics and Actionable Insights	3	1	0	4	30	70	100	
10	mcads-205 DCC – 09	Lab- Python for Data Science and Data Warehousing and Data Mining	0	0	4	4		100	100	
		Total				20			500	

Semester III

S. No.	Course Category	Title of the Course	Credits			Credits earned	Evaluation			Remarks
			L	T	P		IA	ESA	Total	
SEMESTER - III										
11	mcads-301 DCC – 10	Deep Learning	3	1	0	4	30	70	100	
12	mcads-302 DCC – 11	Data Science for Financial Technology (FinTech)	3	1	0	4	30	70	100	
13	mcads-303a DSE – 02A	Enterprise Cloud Architecture and Virtualization	3	1	0	4	30 30	70 70	100	Student may opt any one
	mcads-303b DSE – 02B	Data Science for Retail Intelligence								
14	mcads-304a DSE – 03A	Data Science for Smart Cities	3	1	0	4	30 30	70 70	100	Student may opt any one
	mcads-304b DSE – 03B	Data Science for Social media								
15	mcads-305 DCC – 12	Lab- Deep Learning and Data Science for Financial Technology (FinTech)	0	0	4	4		100	100	
		Total				20			500	

Semester IV

S. No.	Course Category	Title of the Course	Credits			Credits earned	Evaluation			Remarks
			L	T	P		IA	ESA	Total	
SEMESTER - IV										
16	mcads-401 DCC – 13	Data Science with R	3	1	0	4	30	70	100	
17	mcads-402 DCC – 14	Dissertation	3	1	0	4	30	70	100	
18	mcads-403a DSE – 04A	Data Science for Healthcare	3	1	0	4	30 30	70 70	100	Student may opt any one
	mcads-403b DSE – 04B	Data Science for Industry 4.0								
19	mcads-404a DSE – 05A	Data Science for Urban Analytics	3	1	0	4	30 30	70 70	100	Student may opt any one
	mcads-404b DSE – 05B	Revenue Management through Data Science								
20	mcads-405 DCC – 15	Lab Data Science with R	0	0	4	4		100	100	
		Total				20			500	

Note: The evaluation of Dissertation shall be done on the basis of Dissertation Report and Viva voce Examination conducted by an external examiner and an internal examiner instead of semester end external exam.

Scheme of Evaluation

(A) Internal Evaluation – 30 Marks

The internal assessment shall consist of the following components: -

- First Internal Test – 10 Marks
- Second Internal Test – 10 Marks
- Hand written Assignment – 05 Marks
(to be written in approximately 500 words)
- Attendance – 05 Marks
- Total – 30 Marks**

The detailed records and supporting documents relating to Internal Assessment shall be maintained by the respective Departments/Colleges in accordance with the Standard Operating Procedure issued by the University Examination Section.

(B) End Semester Examination – 70 Marks

There shall be a Terminal Examination at the end of each Semester carrying 70 marks. The question paper shall comprise three sections **Section A, Section B and Section C.**

Section A

Section A shall be of **20 marks** and consist of **10 questions** carrying **2 marks** each to be answered in approximately *20 words* each. All questions shall be compulsory. The examiner shall set the question paper choosing at least 3 questions from each unit.

Section B

Section B shall be of **30 marks** and consist of **9 questions** carrying **5 marks** each to be answered in approximately *250 words* each. A Candidate shall attempt any six questions. The examiner shall set the questions ensuring that at least three questions are drawn from each Unit.

Section C

Section C shall be of **20 marks** and consist of three questions carrying 10 marks each to be answered in *500 words* each. A Candidate shall attempt any two questions. The examiner shall set one question from each Unit.

Distribution of Marks Component: -

- **Internal Evaluation** – **30 marks**
- **End-Semester Examination** – **70 marks**

Note : –

- (i) The examiner shall ensure adequate coverage of the prescribed syllabus while setting the question paper and, maintain a balanced representation of all Units.
- (ii) The Teacher concerned, shall upload the marks online and submit awards in **3 Copies**, One will be retained by the Head / Principal and Two Copies of awards duly signed both by the Head/ Principal and the Teacher concerned shall be sent to the University Examination Section.
A Copy of the Proforma shall be sent every college/ department by the University Examination Section.
- (iii) The student has to qualify both the Internal Assessment and External Assessment separately. The minimum qualifying marks in both shall be 40 percent.
- (iv) The scheme will be applicable only to Major/ Multidisciplinary Subject (Not to AEC, SEC and VAC) opted by the student.

Scheme of Examination

MCA (Data Science)

Reg. 17 (a)

The MCA Data Science Program will consist of 4 semesters. The examination shall consist of (a) Theory papers (b) Laboratory / Practical work and Dissertation work as per the scheme provided supra. Candidates will be required to pursue a regular, full time course of study at the College / University department for a period of two academic years in order to be eligible for appearing in the examination.

1. Eligibility for MCA (Data Science) : BCA / BSC (CS) / BSC (IT) / Vocational / BSC (Maths/Bio) / (B.A. or B.Com with Mathematics as one of the subjects in graduation) with at least 50% marks in aggregate. Admission is strictly on the basis of merit.
2. Examination:
 - i. Scheme of Evaluation is provided supra.
 - ii. To pass a semester a candidate shall have to score 40% marks in each subject (theory and practical) separately and also 40% marks in aggregate of all the papers prescribed for the examination.
 - iii. Due paper(s) will be applicable if a candidate obtains 40% marks in aggregate and fails in not more than two (2) papers (theory). Due paper(s) of first semester will be held along with the third semester and the due paper(s) of second semester will be held along with the fourth semester. The third and fourth semester due paper(s) will be held in the first and second semester respectively of the next year. The chance of due paper(s) will be given only thrice in each semester.
 - iv. Wherever a candidate appears at for a due paper examination he/she will do so according to the syllabus in force.
 - v. A candidate not appearing at any examination/absent in any paper of term end examination shall be deemed as fail.
3. No Division will be awarded in the first, second and third semester examinations. Division shall be awarded at the end of the fourth semester Examination on the combined marks obtained at the first, second third and fourth semester taken together as noted below:

Passed with First Division	60% of the aggregate marks taken together of all the
	four semester examinations
Passed with second division	48%

All the rest will be declared to have passed the examination.
4. The grace marks shall be given up to 1% of the total aggregate marks of theory and practical of that semester in maximum one paper.
5. Candidates reappearing at an examination in a subsequent year shall be examined in accordance with the scheme and syllabi in force and shall be entitled to the award of the degree of year in which they clear the last failing/unclear paper.

mcads-101 Mathematical foundations of Data Science	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

Introduce core concepts, lifecycle stages, and foundational disciplines of data science.

Build proficiency in data cleaning, preprocessing, and exploratory data analysis (EDA).

Train students to communicate data insights effectively using visual storytelling tools.

Establish a firm understanding of descriptive and inferential statistical concepts.

Teach the theoretical foundations of probability theory and its application to data modelling.

Build mathematical competency in simple and multiple linear regression architectures.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Explain the fundamental principles and lifecycle of data science projects.

CO2: Apply data preprocessing techniques to clean and transform raw datasets for analysis.

CO3: Perform exploratory data analysis and statistical tests to derive initial insights from data.

CO4: Calculate and interpret descriptive statistics to summarize raw data properties.

CO5: Mathematically derive and construct simple and multiple linear regression models

CO6: Create impactful data visualizations and communicate data-driven findings to stakeholders.

3. Course Syllabus

mcads-101	Duration: 3 hours
Mathematical foundations of Data Science	Max Marks: 70

Unit 1

Introduction to Data Science: Definition, evolution, the data science lifecycle, and roles in the industry. **Types of Data:** Structured, unstructured, semi-structured, time-series, and spatial data. **Data Acquisition:** Introduction to relational databases, NoSQL systems, web scraping, and APIs. **Data Cleaning & Preprocessing:** Handling missing values, removing duplicates, outlier detection, data normalization, and feature scaling. **Exploratory Data Analysis (EDA):** Summary statistics (mean, median, variance), data distributions, and correlation analysis.

Unit 2

Descriptive Statistics: Measures of central tendency (mean, median, mode, trimmed mean), measures of dispersion (variance, standard deviation, range). **Data Visualization Theory:** Empirical distributions, frequency tables, histograms. **Probability Basics:** Sample spaces, events, mutually exclusive events, axioms of probability, and counting rules (permutations and combinations). **Conditional Probability:** Independence, conditional probability formulas, law of total probability, and Bayes' Theorem with real-world classification applications.

Unit 3

Correlation: Covariance, Pearson product-moment correlation coefficient, Spearman rank correlation, and causation vs. correlation. **Simple Linear Regression (SLR):** The SLR model equation, Ordinary Least Squares (OLS) estimation method, and illustrative examples. **Multiple Linear Regression (MLR):** Matrix notation of MLR, interpretation of coefficients, and the coefficient of determination, usage of regression analysis in business intelligence.

4. Suggested Books and References

- [1] **OpenIntro Statistics** by David Diez, Mine Çetinkaya-Rundel, and Christopher Barr (OpenIntro).
- [2] **Introduction to Probability and Statistics for Engineers and Scientists** by Sheldon M. Ross (Academic Press).
- [3] **Applied Linear Regression Models** by Michael H. Kutner, Christopher J. Nachtsheim, and John Neter (McGraw-Hill).
- [4] **Doing Data Science: Straight Talk from the Frontline** by Cathy O'Neil and Rachel Schutt (O'Reilly Media).
- [5] **Practical Statistics for Data Scientists** by Peter Bruce, Andrew Bruce, and Peter Gedeck (O'Reilly Media).
- [6] **An Introduction to Statistical Learning** by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (Springer).

mcads-102 Database Technology	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Introduce foundational concepts of relational database management systems (RDBMS) and data modelling.

Teach advanced SQL for complex data retrieval, manipulation, and database administration.

Build skills in database design theory, normalization, and structural optimization.

Develop an understanding of transaction management, concurrency control, and crash recovery.

Provide exposure to modern non-relational (NoSQL) databases and distributed architectures.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Design accurate ER diagrams and map them to functional relational database schemas.

CO2: Formulate complex, optimized SQL queries to manipulate and retrieve data from a relational system.

CO3: Apply normalization techniques to eliminate data redundancy and anomalies.

CO4: Analyze transaction schedules to ensure concurrency control, data consistency, and system recovery.

CO5: Evaluate and select appropriate relational or NoSQL database solutions based on application requirements.

3. Course Syllabus

mcads-102 Database Technology	Duration: 3 hours Max Marks: 70
--	--

Unit 1

Introduction to DBMS: Database vs. file systems, three-schema architecture, and data independence. **Data Modelling:** Entity-Relationship (ER) model, extended ER features, and mapping ER diagrams to relational schemas. **Relational Algebra:** Fundamental operations (Select, Project, Join, Union, Set Difference). **Relational Database Design:** Functional dependencies, decomposition, and normalization forms (1NF, 2NF, 3NF, BCNF). **Data Integrity:** Domain constraints, referential integrity, assertions, and triggers.

Unit 2:

Structured Query Language (SQL): DDL, DML, DCL, complex joins, subqueries, and nested queries. **Advanced Programmability:** Stored procedures, user-defined functions, and cursors. **Storage and Indexing:** File organizations, primary/secondary indexing, B-trees, B+ trees, and hashing. **Query Processing & Optimization:** Query evaluation steps, cost estimation, and query execution plans.

Unit 3:

Transaction Management: ACID properties, states of execution, and schedules. **Concurrency Control:** Lock-based protocols (2PL), deadlocks, and timestamp-based protocols. **Recovery Systems:** Failure classification, log-based recovery, and checkpoints. **Database Security:** Authorization, access control (DAC, MAC, RBAC), and SQL injection prevention. **Beyond Relational:** Introduction to NoSQL architectures (Key-Value, Document, Columnar, Graph) and cloud databases.

4. Suggested Books and References

- [1] **Database System Concepts** by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan (McGraw-Hill).
- [2] **Fundamentals of Database Systems** by Ramez Elmasri and Shamkant B. Navathe (Pearson).
- [3] **Database Management Systems** by Raghu Ramakrishnan and Johannes Gehrke (McGraw-Hill).
- [4] **Designing Data-Intensive Applications** by Martin Kleppmann (O'Reilly Media)

mcads-103: Object Oriented Programming using Java	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the history, evolution, and foundational architecture of the Java programming language.

Teach core programming constructs including data types, variables, arrays, and operators.

Build skills in controlling program flow using conditional statements and looping structures.

Develop a deep understanding of the class mechanism as the absolute basis of Java's object-oriented paradigm.

Train students to design, implement, and instantiate basic user-defined classes with methods and constructors.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Explain the historical evolution of Java, the role of bytecode, and the architectural mechanics of the JVM.

CO2: Select and implement appropriate primitive data types, variables, and type casting rules in basic programs.

CO3: Construct robust mathematical and logical expressions using Java operators and structural precedence.

CO4: Design complex control flows using nested selection, loops, and jump statements to solve algorithmic problems.

CO5: Model real-world entities into custom Java classes by correctly implementing fields, parameterized constructors, and methods.

3. Course Syllabus

mcads-103: Object Oriented Programming using Java	Duration: 3 hours Max Marks: 70
--	--

Unit 1

The Genesis of Java: History, the lineage of Java, the creation of the internet, and Java's buzzwords, NetBeans IDE. **Architecture:** The JVM, bytecode, platform independence, servlets, and the evolution of Java SE. **Data Types & Variables:** Primitive types (Integers, Floating-point, Characters, Booleans), scope and lifetime of variables, dynamic initialization, and type conversion/casting. **Arrays:** One-dimensional arrays, multidimensional arrays, Encapsulation & Abstraction, Bundling data cleaning parameters or data-transformation steps within specific custom objects, Inheritance & Polymorphism

Unit 2

Operators: Arithmetic operators, bitwise operators, relational operators, boolean logical operators, assignment operators, and operator precedence. **Control Statements - Selection:** if-else conditionals, nested if statements, if-else-if ladders, and the switch statement (including nested switches). **Control Statements - Iteration:** while, do-while, and traditional for loops. **Jump Statements:** Using break (including labeled break), continue (including labeled continue), and return.

Unit 3

Class Fundamentals: The general form of a class, declaring objects, and assigning object reference variables. Introducing methods, adding methods to classes, returning values from methods, and working with parameters. **Constructors:** Parameterized constructors, the role of constructors in object initialization, and the implicit default constructor. **This Keyword:** Understanding instance variable hiding, garbage collection mechanics, and the finalization process (finalize()).

4. Suggested Books and References

- [1] **Java: The Complete Reference** by Herbert Schildt (McGraw-Hill)
- [2] **Java: A Beginner's Guide** by Herbert Schildt (McGraw-Hill).
- [3] **Core Java Volume I – Fundamentals** by Cay S. Horstmann (Pearson).

mcads-104: Data Visualization using Tableau	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the principles of visual design, human perception, and effective data storytelling.

Build foundational to advanced proficiency in using Tableau Desktop for data analysis.

Teach data connection, blending, shaping, and preprocessing techniques within Tableau.

Enable students to create diverse charts, maps, calculations, and interactive dashboards.

Train students to design professional, executive-ready data stories that drive business decisions.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Apply human perception and visual design principles to create accurate, ethical charts.

CO2: Connect, blend, and reshape disparate data sources within Tableau for downstream analysis.

CO3: Construct complex, analytical charts and geographic maps tailored to the structure of the underlying data.

CO4: Write calculated fields, table calculations, and LOD expressions to isolate deep business insights.

CO5: Author interactive, executive-ready dashboards and storytelling narratives that solve real-world industry problems.

3. Course Syllabus

mcads-104: Data Visualization using Tableau	Duration: 3 hours Max Marks: 70
--	--

Unit 1

Introduction to Visualization: Information visualization theory, visual encodings (color, size, shape), Gestalt principles, and avoiding misleading charts. **Tableau Ecosystem:** Overview of Tableau Desktop, Prep, Server, Cloud, and Public. Interface navigation and workspace terminology. **Connecting to Data:** Connecting to files (Excel, CSV, JSON) and relational databases. Live connections vs. Extracts.

Unit 2

Core Chart Types: Bar charts, line charts, scatter plots, pie charts, text tables, heat maps, and dual-axis charts. **Geographic Visualizations:** Geographic roles, plotting coordinates, filled maps, symbol maps, and custom geocoding. **Organizing Data:** Creating and managing Sets, Groups, and Hierarchies. Sorting and advanced filtering techniques.

Unit 3

Dashboard Design: Tiled vs. floating layouts, sizing, padding, and UI design best practices. **Dashboard Interactivity:** Configuring Actions (Filter, Highlight, URL, Go to Sheet, Parameter actions) and Tooltip customization. **Data Storytelling:** Creating linear narratives using Tableau Stories to convey specific insights and recommendations. **Publishing and Sharing:** Exporting static formats (PDF, Image), packaging workbooks (.twbx), publishing to Tableau Public, and introduction to data governance.

4. Suggested Books and References

- [1] Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic (Wiley).
- [2] Communicating Data with Tableau by Ben Jones (O'Reilly Media).
- [3] Learning Tableau (Latest Edition) by Joshua N. Milligan (Packt Publishing).
- [4] Practical Tableau by Ryan Sleeper (O'Reilly Media).
- [5] Tableau Training Videos (Official Tableau Free Training Suite).

mcads-201 Data-Driven Artificial Intelligence	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Connect data analytics, statistics, and data mining principles to modern Artificial Intelligence (AI) paradigms.

Teach classical search algorithms, logical reasoning, and probabilistic AI models.

Build advanced proficiency in implementing, tuning, and evaluating predictive Machine Learning models.

Introduce Neural Networks and Deep Learning concepts for pattern recognition in unorganized data.

Train students to deploy production-ready AI models and address algorithmic bias and ethics.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Design intelligent agents and solve complex state-space optimization problems using heuristic search methods.

CO2: Build probabilistic reasoning structures and Bayesian networks to handle environmental uncertainty.

CO3: Construct and fine-tune high-performance ensemble machine learning models to solve predictive classification tasks.

CO4: Implement basic Artificial Neural Networks and backpropagation routines to parse multi-dimensional feature sets.

CO5: Evaluate AI systems for algorithmic fairness, explainability, and alignment with modern ethical frameworks.

3. Course Syllabus

mcads-201	Duration: 3 hours
Data-Driven Artificial Intelligence	Max Marks: 70

Unit 1:

Introduction to AI: Definition, history, Turing test, rational agents, and the relationship between Data Science, Machine Learning, and AI. **Problem Solving & Search:** State-space representation, uninformed search (BFS, DFS), informed heuristic search, and adversarial search (Minimax, Alpha-Beta Pruning). **Knowledge Representation & Logic:** Propositional logic, first-order logic, inference, and knowledge-based systems. Introduction to NetworkX Python library for studying graphs and networks, visualizing state-space representations and executing custom search paths visually using NetworkX. **Reasoning under Uncertainty:** Probabilistic reasoning, Bayesian Networks, Hidden Markov Models (HMMs), and temporal data modelling. Introduction to pgmpy (Probabilistic Graphical Models)

Unit 2:

The ML Pipeline: Feature engineering, data leakage prevention, cross-validation (K-Fold), and hyperparameters vs. model parameters. **Advanced Supervised Learning:** Ensemble methods (Bagging, Boosting, Random Forests, AdaBoost, Gradient Boosting, XGBoost), and Support Vector Machines (SVM) with kernel tricks. Introduction to Scikit-Learn Python library for classical machine learning **Dimensionality Reduction & Advanced Clustering:** Manifold learning (t-SNE), Gaussian Mixture Models (GMMs), and Expectation-Maximization (EM) algorithm. **Model Evaluation Metrics:** Precision-Recall curves, F-Beta score, Log Loss, and computing bias-variance trade-offs. XGBoost / LightGBM gradient boosting libraries in python.

Unit 3

Foundations of Deep Learning: Biological vs. artificial neurons, Perceptron architecture, multi-layer perceptrons (MLPs), backpropagation algorithms, and activation functions (ReLU, Sigmoid, Softmax). Introduction to PyTorch (or TensorFlow/Keras) open-source deep learning framework. **Specialized Architectures (Introductory):** Brief overview of Convolutional Neural Networks (CNNs) for images and Recurrent Neural Networks (RNNs/LSTMs) for text **Modern Generative AI Foundations:** Introduction to attention mechanisms, transformers, Large Language Models (LLMs), and prompt engineering fundamentals. Introduction to Hugging Face (Transformers Ecosystem)

4. Suggested Books and References

- [1] Artificial Intelligence: A Modern Approach (Latest Edition) by Stuart Russell and Peter Norvig (Pearson).
- [2] Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron (O'Reilly Media).
- [3] Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (MIT Press).
- [4] Pattern Recognition and Machine Learning by Christopher M. Bishop (Springer).

mcads-202	Duration: 3 hours
Python for Data Science	Max Marks: 70

1. Course Objectives

The primary objectives of this course are to:

Establish professional-grade proficiency in Python core syntax and data structures.

Teach high-performance vector manipulation and numerical operations using NumPy.

Build advanced data engineering, wrangling, and cleansing skills using Pandas.

Develop mastery over statistical data visualization using Matplotlib and Seaborn.

Train students to execute end-to-end Exploratory Data Analysis (EDA) on messy, real-world datasets.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Configure computational Python environments and utilize Jupyter Notebooks to construct reproducible data pipelines.

CO2: Apply NumPy arrays and vectorized math operations to execute high-performance numerical algorithms without explicit loops.

CO3: Transform, merge, and clean structural datasets of varying formats using advanced Pandas transformations.

CO4: Design multi-dimensional, Publication-quality statistical charts using Matplotlib and Seaborn to communicate hidden patterns.

CO5: Execute a comprehensive Exploratory Data Analysis workflow on unorganized industry data to prepare it for machine learning algorithms.

3. Course Syllabus

mcads-202 Python for Data Science	Duration: 3 hours Max Marks: 70
--------------------------------------	------------------------------------

Unit 1:

Environment & Setup: Introduction to Anaconda ecosystem, Jupyter Notebooks, VS Code, and managing virtual environments using pip and conda. **Core Language Review:** Dynamic typing, built-in collections (Lists, Tuples, Sets, Dictionaries), list comprehensions, lambda functions, and exception handling. **Array Indexing and Operations:** Slicing, multi-dimensional indexing, boolean masking, universal functions (ufuncs), broadcasting rules.

Unit 2

Pandas Core Structures: Series and DataFrames, axis concepts, indexing, selecting, and filtering data. **Data Ingestion & Export:** Reading and writing files (CSV, Excel, JSON, Parquet) and fetching data via SQL databases using SQLAlchemy. **Data Cleaning:** Handling missing values (dropna, fillna), stripping white spaces, mapping values, removing duplicates, and data type casting.

Unit 3

Statistical Visualization with Matplotlib: Anatomy of a figure, axes, subplots, line plots, scatter plots, bar charts, histograms, and customizing chart aesthetics. **Statistical Graphics with Seaborn:** Distribution plots (kdeplot, displot), categorical plots (boxplot, violinplot), and matrix plots (heatmap) for correlation matrices. **Exploratory Data Analysis (EDA) Workflow:** Formulating hypotheses, univariate and bivariate analysis, detecting outliers using Interquartile Range (IQR), and scaling/normalizing features.

4. Suggested Books and References

- [1] Python Data Science Handbook (Latest Edition) by Jake VanderPlas (O'Reilly Media).
- [2] Python for Data Analysis by Wes McKinney (O'Reilly Media).
- [3] Effective Pandas: Patterns for Data Science by Matt Harrison (Treading on Python).
- [4] Fluent Python by Luciano Ramalho (O'Reilly Media).

mcads-203 Data Warehousing and Data Mining	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the architectural principles of Data Warehousing (DWH) and Online Analytical Processing (OLAP).

Teach structural design techniques for data repositories using multi-dimensional modeling.

Build foundational knowledge of the ETL (Extract, Transform, Load) pipeline architecture.

Develop mastery over data mining workflows and algorithmic pattern recognition.

Train students to implement classification, clustering, and association rule algorithms on enterprise datasets.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Differentiate between transactional database architectures (OLTP) and analytical data warehouses (OLAP).

CO2: Design multi-dimensional schemas (Star, Snowflake) to structurally represent complex enterprise business metrics.

CO3: Construct robust ETL frameworks to clean, transform, and aggregate raw operational records into optimized data warehouses.

CO4: Apply association rule algorithms and mining frameworks to discover hidden patterns in transactional data.

CO5: Evaluate and execute predictive classification and clustering models to group data and predict behavioral trends.

3. Course Syllabus

mcads-203	Duration: 3 hours
Data Warehousing and Data Mining	Max Marks: 70

Unit 1

Introduction to Data Warehousing: Definition, characteristics (subject-oriented, integrated, time-variant, non-volatile), and operational databases (OLTP) vs. Data Warehouses (OLAP). **Warehouse Architecture:** Operational data stores, data marts, metadata repositories, three-tier data warehouse architecture, and enterprise data models. **Multi-Dimensional Modeling:** Data cubes, dimensions, measures, Star Schema, Snowflake Schema, and Fact Constellation (Galaxy) Schema. **OLAP Operations:** Drill-down, roll-up, slice, dice, and pivot operations.

Unit 2

The ETL Process: Data Extraction techniques, Transformation mapping (cleansing, sorting, deriving metrics), Loading strategies (initial load vs. incremental refresh), and Data Quality Management. **Introduction to Data Mining:** Definition, knowledge discovery process (KDD lifecycle), data mining, and major challenges in data mining. Using Weka for Data Mining. **Data Preprocessing for Mining:** Data cleaning, data integration, data reduction (sampling, dimensionality reduction), and data transformation (discretization and normalization).

Unit 3

Association Rule Mining: Market Basket Analysis, frequent itemsets, support, confidence, and lift metrics. Usage of Apriori and FP-Growth algorithm variants in Weka for business insights. **Advanced Classification:** Overview of decision tree induction (ID3/C4.5), Naïve Bayes classification, and performance metrics (Confusion Matrix, ROC curves). Using Arff files and Csv files in Weka for classification in business datasets. **Advanced Clustering:** Partitioning methods (**K-Means**, K-Medoids), hierarchical methods (AGNES, DIANA), and density-based methods (DBSCAN). Brief introduction to other clustering options in Weka. **Outlier Analysis:** Types of outliers, statistical outlier detection methods, and proximity-based outlier mining.

4. Suggested Books and References

- [1] **Data Mining: Concepts and Techniques** by Jiawei Han, Micheline Kamber, and Jian Pei (Morgan Kaufmann).
- [2] **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling** by Ralph Kimball and Margy Ross (Wiley).
- [3] **Introduction to Data Mining** by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar (Pearson).
- [4] **Building a Data Warehouse: With Examples in SQL Server** by Vincent Rainardi (Apress).

mcads-204a Data Science for Web Intelligence	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the graph structure, topologies, and unique characteristics of the World Wide Web.

Teach automated data harvesting techniques via web scraping, crawling, and API integrations.

Build mathematical and algorithmic skills for web link analysis and search engine ranking.

Develop proficiency in structuring and analyzing unstructured textual web data using NLP.

Train students to extract business insights from clickstream logs, user behaviors, and social networks.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Build robust data engineering pipelines to harvest and parse complex data from static and dynamic web interfaces.

CO2: Compute PageRank and HITS scores to evaluate node importance within high-volume web graphs.

CO3: Implement vector space models and text mining workflows to classify and query unstructured text documents.

CO4: Extract and analyze web clickstream logs to trace user behavior patterns and compute key business performance metrics.

CO5: Architect collaborative and content-based recommendation systems tailored for personalized web deployment.

3. Course Syllabus

mcads-204a Data Science for Web Intelligence	Duration: 3 hours Max Marks: 70
---	--

Unit 1

Web Data Architecture: Characteristics of web data, HTTP request/response protocols, DOM trees, JSON/XML parsing, and RESTful web APIs. **Web Crawling Mechanics:** Crawling architectures, robot politeness (robots.txt), URL handling, deduplication. **Clickstream Analytics:** Parsing server log files (Apache/Nginx logs), user session identification, cookie tracking, and data processing pathways. **Web Intelligence Metrics:** Conversions, bounce rates, exit rates, conversion funnels, and page views. A/B testing methodologies on the web. **Real-world web analytics dashboards:** Using Matomo (open-source) for learning visual context for conversion funnels, page views, and traffic attribution

Unit 2

Information Retrieval (IR): Web search engine architecture, inverted indices, tokenization, stemming, lemmatization, and stop-word removal. **Opinion Mining & Sentiment Analysis:** Lexicon-based vs. machine-learning-based sentiment scoring on web text, reviews, and social media feeds. **Web Classification & Clustering:** Categorizing web pages automatically into topics, and clustering search engine query sessions. **NLTK (Natural Language Toolkit) :** Python libraries for natural language processing for the text preprocessing pipelines: tokenization, stemming, lemmatization, and stop-word removal.

Unit 3

Recommender Systems: Collaborative Filtering (User-based and Item-based), Content-based filtering, and handling cold-start problems in e-commerce. **Social Network Analysis:** Social graphs, community detection algorithms (Louvain, Girvan-Newman), influence propagation, and viral marketing metrics. **Surprise (Simple Python Recommendation System Engine):** Introduction to Surprise Python scikit for building and analyzing recommender systems. Collaborative Filtering (User-based and Item-based) and Content-based algorithms. Using built-in datasets (like MovieLens) to quickly evaluate prediction errors.

4. Suggested Books and References

- [1] Web Data Mining: Exploring Hyperlinks, Contents, and Usage by Bing Liu (Springer).
- [2] Introduction to Information Retrieval by Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze (Cambridge University Press).
- [3] Mining the Social Web by Matthew A. Russell and Mikhail Klassen (O'Reilly Media).
- [4] Networks, Crowds, and Markets: Reasoning About a Highly Connected World by David Easley and Jon Kleinberg (Cambridge University Press).

mcads-204b Data Analytics and Actionable Insights	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Formalize the standardized lifecycle models of data analytics used to solve structured business problems.

Teach the classification and application of different analytical types (Descriptive, Diagnostic, Predictive, Prescriptive).

Build foundational skills in designing measurable business KPIs, metrics, and data tracking frameworks.

Develop professional competency in data auditing, anomaly discovery, and structural validation workflows.

Train students to craft compelling, evidence-based business recommendations and data-driven narratives.

2. Course Outcomes (COs)

CO1: Map a complex, ambiguous corporate problem to a structured, repeatable analytical framework like CRISP-DM.

CO2: Formulate and construct measurable business KPIs and cohort segments.

CO3: Perform formal data profiling and quality audits to evaluate structural data pipelines.

CO4: Apply root-cause analysis methodologies to isolate and diagnose the drivers behind sudden shifts in business metrics.

CO5: Author professional data-driven narratives and executive slide-decks that clearly articulate technical findings to non-technical business leaders.

3. Course Syllabus

mcads-204b Data Analytics and Actionable Insights	Duration: 3 hours Max Marks: 70
--	--

Unit 1

Foundations of Data Analytics: Distinguishing Data Analytics from Data Science and Data Engineering. The value-complexity ladder of analytics. Introduction to Jira / Trello Agile Project Management and Kanban boards. **Standard Frameworks:** The CRISP-DM (Cross-Industry Standard Process for Data Mining) lifecycle and OSEMN (Obtain, Scrub, Explore, Model, iNterpret) frameworks and case studies. **Problem Scoping & Translation:** Mapping vague business problems to precise analytical queries. Defining target variables, objectives, and project constraints.

Unit 2

Descriptive & Diagnostic Deep-Dives: Conducting operational drill-downs, establishing cohort analysis frameworks, and using segmentation to isolate driver variables. Using Tableau for Descriptive & Diagnostic Deep-Dives, build interactive dashboards with drill-down capabilities. **Anomalies and Outlier Auditing:** Distinguishing true data anomalies (errors) from extreme business events (e.g., holiday sales spikes). Impact of outliers on baseline business metrics. **Experimental Foundations:** Basics of structured corporate evaluation. Introduction to conversion funnels, baseline measurements, and the operational logic of A/B testing.

Unit 3

Root-Cause Analysis (RCA): Applying structural frameworks (The 5 Whys, Fishbone Diagrams) to explain metric shifts observed in data reporting. **Translating Insights to Action:** Moving from "What happened" to "What should we do." Constructing executive summaries, action-oriented bullet points, and business case justifications. Create action-oriented executive summaries and business infographics using Canva / Infogram **Visual Presentation Frameworks:** Storyboarding analytical findings, structuring reports for non-technical stakeholders, and matching report detail to executive tiers. Preparing corporate presentations

4. Suggested Books and References

- [1] Data Analytics Basics: A Beginner's Guide to Data Analysis by N. P. Padhy (CRC Press).
- [2] Key Performance Indicators (KPI): The 75+ Measures Every Manager Needs to Know by Bernard Marr (Financial Times Prentice Hall)
- [3] Lean Analytics: Use Data to Build a Better Startup Faster by Alistair Croll and Benjamin Yoskovitz (O'Reilly Media).
- [4] Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things by Bernard Marr (Kogan Page).
- [5] Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic (Wiley).

mcads-301 Deep Learning	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Adapt deep learning topologies explicitly to solve high-dimensional structured and unstructured data science problems.

Teach the mathematical and structural application of neural networks to tabular data profiling and asset forecasting.

Build advanced skills in creating, managing, and indexing multi-dimensional vector embeddings for feature engineering.

Develop mastery over advanced deep time-series forecasting models using recurrence and attention mechanisms.

Train students to use deep learning for automated data cleaning, anomaly detection, and synthetic data generation.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Construct and configure specialized deep learning architectures (like TabNet) optimized to process unstructured and structured tabular matrices.

CO2: Transform sparse, complex high-dimensional categorical features into dense, continuous neural network entity embeddings.

CO3: Implement advanced temporal transformers and recurrent architectures to execute multi-step forecasting on multivariate data streams.

CO4: Understand variational autoencoders to isolate, measure, and flag complex, non-linear multi-dimensional data anomalies.

3. Course Syllabus

mcads-301 Deep Learning	Duration: 3 hours Max Marks: 70
--	--

Unit 1

Deep Tabular Models: Moving beyond Tree-Based models. Multi-Layer Perceptrons (MLPs) for tabular processing, CatBoost/XGBoost vs. Deep Learning, and specialized architectures like **TabNet** (attentive interpretable tabular learning). **Entity Embeddings:** Categorical data transformation through neural network embedding layers. Mapping sparse high-dimensional features into continuous, low-dimensional vector spaces.

Unit 2

Deep Time-Series Foundations: Recurrent Neural Networks (RNNs), LSTMs, and GRUs customized for multivariate financial and operational time-series forecasting. Handling structural data irregularities and long-range dependencies. **Attention and Transformers for Data:** Adapting Transformer architectures to numeric and temporal data streams. Implementing Temporal Fusion Transformers (TFT) and Informer models for enterprise-scale predictive modeling. **Deep Survival Analysis:** Combining neural networks with traditional statistical survival models (Cox Proportional Hazards) to predict customer churn risk windows and asset lifecycle depletion.

Unit 3

Deep Anomaly Detection: Implementing Variational Autoencoders (VAEs) and Reconstruction-based networks to parse structural logs, flag multi-dimensional outliers, and identify corporate fraud pipelines. **Synthetic Data Generation:** Utilizing deep generative models (Tabular GANs/CTGANs) to securely synthesize high-fidelity, privacy-compliant tabular datasets for downstream data science experiments. **Self-Supervised Pre-training:** Leveraging massive unlabeled corporate repositories for contrastive pre-training to boost model accuracy when labeled historical target are sparse.

4. Suggested Books and References

- [1] Deep Learning for Data Analytics: Foundations, Techniques, and Applications by Himansu Das, Radha Krishna, and B. K. Tripathy (Academic Press).
- [2] Deep Learning for Time Series Forecasting by Jason Brownlee (Machine Learning Mastery).
- [3] Deep Learning for Tabular Data: Mitigating the Deep Learning Tabular Dilemma by Boris Shishkov (Springer).
- [4] Research Paper Compendium: *TabNet: Attentive Interpretable Tabular Learning* (Arik & Pfister) and *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting* (Lim et al.).

mcads-302 Data Science for Financial Technology (FinTech)	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

Introduce the core components, evolution, and disruptive business models of FinTech.

Teach financial time-series modeling techniques to process asset prices and market volatility.

Build advanced data-driven skills for credit scoring, underwriting, and automated financial risk management.

Enable students to design and evaluate algorithmic trading strategies and quantitative portfolios.

Provide foundational knowledge of blockchain architectures, smart contracts, and decentralized data analytics.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Process and clean structured and unstructured alternative financial datasets for automated analytical pipelines.

CO2: Forecast financial market volatility and future asset pricing metrics using linear and non-linear time-series algorithms.

CO3: Construct robust credit risk scoring and anomaly detection architectures capable of mitigating institutional lending risk and transaction fraud.

CO4: Design, build, and statistically backtest automated algorithmic trading frameworks under simulated real-world market constraints.

CO5: Analyze blockchain network transactional records and audit automated financial models for regulatory safety compliance.

3. Course Syllabus

mcads-302	Duration: 3 hours
Data Science for Financial Technology (FinTech)	Max Marks: 70

Unit 1

The FinTech Landscape: Evolution from traditional finance to FinTech, digital banking, peer-to-peer lending, neo-brokers, and regulatory technology (RegTech). **Financial Ingestion & Market Data:** API integrations for real-time market data (e.g., Yahoo Finance, Alpha Vantage), cleaning financial data, and managing irregular sampling frequencies. **Time-Series Analysis Foundations:** Stationarity, autocorrelation, partial autocorrelation functions (ACF/PACF), differencing, and white noise tests.

Unit 2

Credit Risk Modeling: Applying logistic regression, decision trees, and ensemble models to calculate Probability of Default (PD) and Loss Given Default (LGD). **Imbalanced Data Challenges:** Handling highly skewed distributions in financial datasets using SMOTE (Synthetic Minority Over-sampling Technique), cost-sensitive learning, and anomaly detection algorithms (Isolation Forests, One-Class SVM).

Unit 3

Portfolio Optimization: Modern Portfolio Theory (MPT), efficient frontiers, Sharpe ratio maximization, and data-driven asset allocation strategies. **Blockchain Data Analytics:** Public ledger architectures (Bitcoin, Ethereum), parsing blockchain transactions, smart contract execution traces, and decentralized finance (DeFi) telemetry tracking.

4. Suggested Books and References

[1] **Advances in Financial Machine Learning** by Marcos López de Prado (Wiley) — *The industry-standard guide for quantitative data modeling.*

[2] **Introduction to Financial Technology** by Roy S. Freedman (Academic Press).

[3] **Data Science for Finance** by Yves Hilpisch (O'Reilly Media) — *Excellent practical reference using Python ecosystems.*

mcads-303a Enterprise Cloud Architecture and Virtualization	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Evaluate cloud computing implementations from both corporate business value and technical IT scalability perspectives.

Analyze the operational boundaries, security trade-offs, and service models of public, private, and hybrid cloud structures.

Teach the design and deployment of enterprise-grade cloud systems using Service-Oriented Architecture (SOA) principles.

Build foundational skills in governing automated cloud provisioning, billing structures, and high-availability disaster recovery.

Deeply analyze virtualization layers, hardware partitioning, and decoupled network/storage architectures within modern virtualized data centers.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Critique cloud migration plans by balancing IT operational requirements against business agility and adoption costs.

CO2: Architect secure, multi-tenant public and private cloud models featuring automated infrastructure self-service provisioning.

CO3: Implement Service-Oriented Architecture (SOA) components to map changing consumer service demands to elastic cloud ecosystem solutions.

CO4: Design automated governance pipelines to track cloud assets, report operational usage metrics, and enforce metered billing routines.

3. Course Syllabus

mcads-303a	Duration: 3 hours
Enterprise Cloud Architecture and Virtualization	Max Marks: 70

Unit 1

Cloud Dynamics: Cloud computing introduction from business and IT perspectives. The alignment of cloud architectures with dynamic infrastructures, core computing characteristics, and industry adoption frameworks. **Structural Cloud Models:** Analysis of cloud characteristics including Measured Service mechanics, public versus private cloud setups, cloud infrastructure self-service portals, and explicit threat modeling for security in a public cloud, cloud solutions, principal technologies.

Unit 2

Enterprise Cloud Solutions: Cloud ecosystem, implementing cloud business process management (BPM), cloud service management operations, cloud stack frameworks, provisioning computing on demand, and optimizing cloud sourcing strategies. **Specialized Cloud Offerings:** Architecting cloud analytics environments, orchestrating automated software testing under cloud setups, managing information security parameters, provisioning virtual desktop infrastructure (VDI), and implementing scalable storage clouds. **Operational Cloud Management:** Ensuring systemic resiliency, automated provisioning, cloud asset management, and unified cloud governance. Engineering high availability and disaster recovery pathways, defining cloud charging models, and configuring usage reporting, billing, and metering structures.

Unit 3

Core Server Virtualization: Virtualization defined and its business/technical benefits. Implementing server virtualization, managing hypervisor abstraction layers through hypervisor management software, and configuring logical partitioning (LPAR). **Virtual I/O and Infrastructure Controls:** Implementing Virtual I/O (VIO) servers, mapping foundational virtual infrastructure requirements, and managing shared hardware access loops. **Storage and Network Virtualization:** Designing storage virtualization matrices, building high-speed Storage Area Networks (SAN), and configuring Network Attached Storage (NAS) fabrics. Integrating cloud server virtualization to operate a modern, software-defined virtualized data center.

4. Suggested Books and References

- [1] **Cloud Computing: Concepts, Technology & Architecture** by Thomas Erl, Ricardo Puttini, and Zaigham Mahmood (Prentice Hall)
- [2] **Virtualization Essentials** by Matthew Portnoy (Sybex)
- [3] **Enterprise Cloud Computing: Technology, Architecture, Applications** by Gautam Shroff (Cambridge University Press).
- [4] **IBM PowerVM Virtualization Introduction and Configuration** by Scott Vetter et al. (IBM Redbooks)

mcads-303b Data Science for Retail Intelligence	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the fundamental data structures, KPIs, and operational systems driving modern retail analytics.

Teach advanced customer segmenting, behavioral tracking, and lifetime value estimation methodologies.

Build predictive optimization workflows for store location selection, product assortment, and dynamic pricing.

Develop mastery over personalized multi-layered recommendation systems and affinity analysis.

Train students to design inventory forecasting networks to optimize supply chains and minimize stockouts.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Clean and structure heterogeneous POS and clickstream records into unified customer-centric metrics.

CO2: Evaluate and implement predictive BG/NBD models to accurately forecast long-term Customer Lifetime Value (CLV).

CO3: Apply association rule mining to transaction matrices to design data-driven promotional Bundling strategies.

CO4: Construct dynamic pricing models by mathematically deriving price elasticity values from historical sales history.

CO5: Build hierarchical SKU-level demand forecasting models to optimize safety stock parameters and distribution logistics routes.

3. Course Syllabus

mcads-303b Data Science for Retail Intelligence	Duration: 3 hours Max Marks: 70
--	--

Unit 1

The Retail Data Ecosystem: Architecture of Point-of-Sale (POS) logs, e-commerce clickstreams, loyalty program databases, and inventory logs. Core metrics: Average Order Value (AOV), Gross Margin Return on Investment (GMROI), and conversion rates. **Advanced Customer Segmentation:** Beyond basic demographics. Implementing Recency, Frequency, Monetary (RFM) modeling, and behavioral clustering using unsupervised learning algorithms. **Association Rule Mining Deep-Dive:** Market Basket Analysis mechanics. Advanced extraction of frequent itemsets using the Apriori and FP-Growth algorithms; interpreting support, confidence, lift, and conviction to guide storefront layout and cross-selling.

Unit 2

Retail Demand Forecasting: Building hierarchical time-series models (Prophet, LightGBM, DeepAR) to forecast SKU-level item demand under heavy seasonal fluctuations, promotions, and markdown events. **Product Assortment and Space Planning:** Algorithmic product mix selection using constrained optimization techniques. Solving multi-item shelf space allocation problems under store physical footprint constraints. **Price Elasticity & Dynamic Pricing:** Estimating price elasticity of demand using regression modeling. Formulating price optimization frameworks that dynamically adjust prices based on real-time competitor tracking, inventory constraints, and decay velocity.

Unit 3

Personalized Recommendation Frameworks: Implementing collaborative filtering (Matrix Factorization, Alternating Least Squares) and content-based recommendation pipelines for digital retail storefronts. **Clickstream Analytics & Attribution Modeling:** Mapping the complete consumer multi-touch attribution journey (First-touch, Last-touch, Linear, Time-decay) across web, mobile, and physical store interactions. **Inventory Optimization & Supply Chain Intelligence:** Calculating optimal safety stock levels, reorder points, and EOQ. Spatial data analytics for brick-and-mortar site selection and distribution route optimization.

4. Suggested Books and References

- [1] Retail Analytics: The Secret Weapon by Emmett Cox (Wiley).
- [2] Data Science for Retail: A Practical Guide to Enhancing Customer Experience and Operational Efficiency by Dmitry Efimov (Apress).
- [3] Predictive Analytics for Marketers: Using Data Mining for Business Advantage by Barry Leventhal (Kogan Page).
- [4] Open Dataset Frameworks: *The Kaggle Instacart Market Basket Analysis dataset or the UCI Machine Learning Repository Online Retail Dataset*

mcads-304a Data Science for Smart Cities	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the pillars, architectures, and spatial data infrastructures (SDI) that define modern smart cities.

Teach advanced geospatial data engineering and vector/raster manipulation using programmatic GIS frameworks.

Build predictive modeling skills for urban mobility, public transit networks, and intelligent traffic management.

Develop analytical workflows for environmental tracking, localized air quality forecasting, and utility distribution optimization.

Train students to implement data-driven public safety response systems while managing spatial data privacy constraints.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Process, re-project, and manipulate multi-layered vector and raster geographic datasets using programmatic Python GIS libraries.

CO2: Formulate city transportation street networks as topological graphs to evaluate urban accessibility and shortest path logistics.

CO3: Build and fine-tune spatial-temporal predictive models to forecast localized vehicular traffic flow and public transit arrival delays.

CO4: Apply spatial interpolation algorithms to heterogeneous sensor networks to construct real-time environmental monitoring maps.

3. Course Syllabus

mcads-304a Data Science for Smart Cities	Duration: 3 hours Max Marks: 70
---	--

Unit 1

The Smart City Data Architecture: Components of cyber-physical urban systems, sensor deployments (IoT grids), spatial databases, and municipal open data portals. **Geospatial Data Wrangling:** Processing vector data types (Points, Linestrings, Polygons) and raster grids. Interfacing with programmatic GIS ecosystems using Python (GeoPandas, Shapely, Fiona, and PyProj). **Urban Telemetry Ingestion:** Querying and streaming open-source urban data infrastructures (e.g., OpenStreetMap via OSMnx, sensor readings via MQTT, and historical transit feeds).

Unit 2

Network Analysis Foundations: Representation of city streets as topological graphs. Calculating shortest paths, network centrality, and accessibility scores. **Traffic Flow Predictive Modeling:** Utilizing high-frequency loop detector data, GPS trajectories, and floating car data to forecast localized traffic congestion via time-series networks (LSTM, Spatio-Temporal Graph Neural Networks). **Ride-Hailing & Demand Allocation:** Spatial-temporal clustering (DBSCAN, HDBSCAN) to identify urban hotspot distributions, predict taxi/ride-sharing demand, and build dynamic fleet relocation algorithms.

Unit 3

Environmental Sensor Analytics: Mapping and forecasting air quality indexes (AQI), urban heat island (UHI) intensities, and noise pollution using spatial interpolation methods (Kriging, Inverse Distance Weighting). **Smart Grid & Utility Optimization:** Building regression and machine learning pipelines to forecast real-time municipal water consumption and electrical grid demands. Predictive maintenance modeling for urban pipe systems and drainage infrastructure. **Data-Driven Public Safety:** Geospatial modeling of emergency resource allocation, optimal fire station placement algorithms using Location-Allocation optimization models, and flood vulnerability zoning.

4. Suggested Books and References

- [1] Data Science for Smart Cities by Kevin Wang and Mark J. Thompson (CRC Press).
- [2] Urban Informatics by Wenzhong Shi, Michael F. Goodchild, Michael Batty, and Kwan-Mei P. (Springer)
- [3] Geocomputation with Python by Robin Lovelace, Jakub Nowosad, and Jannes Muenchow (CRC Press).
- [4] Open Dataset Frameworks: *The New York City TLC (Taxi and Limousine Commission) Trip Record Data or OpenStreetMap*

mcads-304b Data Science for Social media	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the underlying data structures, API communication architectures, and unique structural properties of social media ecosystems.

Teach advanced Natural Language Processing (NLP) techniques for mining unstructured, informal social text.

Build foundational practical skills in modeling social media graphs and detecting online communities.

Develop predictive workflows for viral propagation, misinformation tracking, and user behavior analytics.

Train students to handle large-scale social listening datasets ethically while navigating data privacy parameters.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Build robust API ingestion scripts to securely harvest, clean, and structure high-velocity social media data streams.

CO2: Implement specialized NLP pipelines to parse emojis, slang, and contextual sentiment from informal user-generated text.

CO3: Construct topological network graphs to isolate high-influence users and map hidden online communities using clustering algorithms.

CO4: Apply mathematical diffusion models to forecast the viral propagation velocity of information cascades across social channels.

3. Course Syllabus

mcads-304b Data Science for Social media	Duration: 3 hours Max Marks: 70
---	--

Unit 1

Social Data Ecosystems: Structure of user profiles, follower networks, timelines, and streaming engagement metadata. Navigating social platform APIs (REST vs. Streaming APIs) and web crawling platforms. **Social Text Preprocessing:** Handling informal language, short texts, emojis, hashtags, user mentions (@), custom abbreviations, typos, and multilingual code-switching. **Opinion Mining & Sentiment Analysis:** Lexicon-based approaches (VADER, SentiWordNet) optimized for social expressions. Implementing machine-learning and deep learning text classification models for aspect-based sentiment scoring. **Topic Modeling & Trend Analysis:** Extracting latent discussion themes from high-velocity text streams using Latent Dirichlet Allocation (LDA) and modern embedding-based methods like BERTopic.

Unit 2

Graph Modeling of Social Platforms: Representing social interactions as mathematical graphs (directed/undirected, weighted/unweighted, bipartite structures like user-hashtag graphs). **Community Detection Algorithms:** Identifying online subcultures and echo chambers using modularity maximization, the Louvain algorithm, Girvan-Newman edge-betweenness division, and Infomap.

Unit 3

Information Diffusion & Virality: Modeling how memes, hashtags, and marketing campaigns spread across social graphs using epidemic frameworks (SI, SIR, SIS models) and Independent Cascade Models. **User Behavioral Profiling & Analytics:** Customer segmentation using engagement telemetry. Designing churn prediction models for online communities and optimizing content recommendation feeds using collaborative filtering. **Misinformation & Trust Infrastructure:** Algorithmic detection of coordinated inauthentic behavior (CIB), automated bots, deepfake text generation, and fake news propagation paths.

4. Suggested Books and References

- [1] **Social Media Mining: An Introduction** by Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu (Cambridge University Press)
- [2] **Analyzing Social Networks** by Stephen P. Borgatti, Martin G. Everett, and Jeffrey C. Johnson (SAGE Publications).
- [3] **Mining the Social Web** by Matthew A. Russell and Mikhail Klassen (O'Reilly Media)
- [4] **Open Dataset Frameworks: The Center for Spatial Data Science (CSDS) data repository or municipal open data platforms (such as NYC Open Data or Chicago Data Portal)**

mcads-401 Data Science with R	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Establish professional-grade proficiency in R syntax, functional programming, and data structures.

Teach high-performance data engineering, wrangling, and cleaning using the Tidyverse ecosystem.

Build advanced skills in exploratory data analysis and publication-quality visualization using ggplot2.

Develop mastery over statistical modeling, hypothesis testing, and machine learning pipelines using Tidymodels.

Train students to design and deploy interactive, production-ready web applications using Shiny.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Formulate efficient R programs using functional computing principles and vectorized data processing.

CO2: Transform and clean complex, unorganized transactional datasets using advanced Tidyverse verbs and pipelines.

CO3: Construct publication-quality, multi-layered statistical charts using ggplot2 to communicate complex data patterns.

CO4: Implement and evaluate robust machine learning and forecasting workflows using the unified tidymodels architecture.

CO5: Architect and deploy interactive, low-latency web applications using Shiny to expose model insights to corporate stakeholders.

3. Course Syllabus

mcads-401 Data Science with R	Duration: 3 hours Max Marks: 70
--	--

Unit 1

The R Environment: Setting up RStudio, managing projects, installing CRAN packages, understanding vectorization, data types (vectors, factors, matrices, data frames, tibbles), and functional programming with purrr. **Ingestion & Text Handling:** Reading flat files, Excel sheets, and SQL databases using readr, readxl, and DBI. Manipulating strings and categorical variables using stringr and forcats. **Visual Analytics with ggplot2:** The Grammar of Graphics framework. Mapping aesthetics, configuring geometric layers (bars, lines, scatter plots, box plots), faceting panels, and customizing publication-quality themes.

Unit 2

Parametric & Non-Parametric Statistics: Formulating and executing statistical tests natively in R: ANOVA, t-tests, Chi-Square independence checks, and correlation matrices. **Regression Diagnostics:** Building Ordinary Least Squares (OLS) and Generalized Linear Models (GLM - Logistic Regression) using lm() and glm(). Evaluating residual distributions, Cook's distance, and multicollinearity (VIF). **The tidymodels Framework:** Modern machine learning workflows in R. Data splitting, recipe-based feature engineering, model specification, hyperparameter tuning, and cross-validation pipelines (Random Forests, XGBoost). **Time-Series Analytics:** Processing temporal streams using tsibble and forecasting macro trends using the fable ecosystem (ARIMA, Exponential Smoothing).

Unit 3

Reproducible Research with Quarto: Authoring dynamic, data-driven documents, executive PDF summaries, and automated HTML briefings combining live code, LaTeX mathematics, and markdown text. **Interactive Dashboard Architecture (shiny):** Designing web applications natively in R. Structuring the User Interface (UI) layout and backend Server logic. **R in Production:** Interfacing R pipelines with Python via reticulate. Optimizing execution speeds, profiling code efficiency, and deploying containerized R scripts or Shiny applications to cloud servers (Shinyapps.io or Posit Connect).

4. Suggested Books and References

Standard Textbooks

- [1] R for Data Science (2nd Edition) by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund (O'Reilly Media).
- [2] Tidy Modeling with R by Max Kuhn and Julia Silge (O'Reilly Media).
- [3] Mastering Shiny by Hadley Wickham (O'Reilly Media).

mcads-403a Data Science for Healthcare	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the unique structures, formats, and regulatory governance models of healthcare data.

Teach advanced processing techniques for Electronic Health Records (EHR) and clinical text.

Build foundational skills in biological sequence analysis and genomic data science.

Develop machine learning pipelines for medical image classification and diagnostics.

Train students to handle predictive modeling for patient risk, survival analysis, and public health epidemiology.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Query, parse, and clean clinical records matching standard interoperability frameworks like FHIR and ICD-10.

CO2: Apply specialized clinical NLP toolkits to extract features from unstructured medical reports.

CO3: Implement sequence alignment tools and survival analysis models to quantify health risks based on genomic and lifestyle markers.

CO4: Construct deep learning classification and segmentation pipelines over multi-dimensional DICOM medical image sets.

CO5: Evaluate predictive public health models while strictly complying with global health data privacy and algorithmic fairness standards.

3. Course Syllabus

mcads-403a Data Science for Healthcare	Duration: 3 hours Max Marks: 70
---	--

Unit 1

The Healthcare Data Landscape: Architecture of Electronic Health Records (EHR), claims data, wearable sensor streams, and public health registries. **Privacy, Ethics, and Governance:** Legal frameworks (HIPAA, GDPR), data de-identification algorithms (k-anonymity, l-diversity), and managing algorithmic bias in clinical triage. **Clinical Natural Language Processing (NLP):** Parsing unstructured physician notes, discharge summaries, and clinical trial documents using specialized NLP frameworks (e.g., MedSpacy, John Snow Labs Spark NLP).

Unit 2

Foundations of Bioinformatics: Introduction to genomic structures, DNA/RNA sequencing technologies (Next-Generation Sequencing), and biological sequence formats (FASTA, FASTQ, SAM/BAM). **Sequence Alignment & Variant Analysis:** Pairwise and multiple sequence alignment algorithms (BLAST, Smith-Waterman). Identifying Single Nucleotide Polymorphisms (SNPs) and mapping mutation profiles to diseases.

Unit 3

Medical Imaging Data Engineering: Handling DICOM and NIfTI file formats. Preprocessing medical images (intensity normalization, spatial registration, and data augmentation). **Deep Learning for Clinical Vision:** Implementing Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for multi-class classification and semantic segmentation (e.g., U-Net for tumor boundary isolation) in X-rays, MRIs, and CT scans.

4. Suggested Books and References

- [1] Healthcare Data Analytics by Chandan K. Reddy and Charu C. Aggarwal (CRC Press)
- [2] Introduction to Genomic Data Science by Julia Stewart Lowndes and Stephanie Hampton (Princeton University Press).
- [3] Deep Learning for Genomics by Ramsundar, Bharath et al. (O'Reilly Media).
- [4] Bioinformatics and Functional Genomics by Jonathan Pevsner (Wiley-Blackwell).

mcads-403b Data Science for Industry 4.0	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the pillars, architectures, and data-driven paradigms of Industry 4.0 and the Industrial Internet of Things (IIoT).

Teach edge data ingestion, cleaning, and windowing techniques for high-frequency sensor telemetry.

Build advanced diagnostic and predictive skills for industrial asset health and Predictive Maintenance (PdM).

Develop computer vision workflows for automated assembly line anomaly tracking and quality inspection.

Train students to build scalable industrial optimization algorithms, digital twins, and supply chain logistics models.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Formulate data pipelines that ingest and clean telemetry from heterogeneous industrial communication nodes like OPC-UA.

CO2: Transform raw vibration and thermal signals into time-frequency domain features optimized for downstream predictive analytics.

CO3: Construct robust machine learning and survival models to estimate the Remaining Useful Life (RUL) of physical assets.

CO4: Implement deep-learning vision pipelines to automate non-destructive defect identification during real-time manufacturing.

CO5: Build linear programming solvers to dynamically optimize production schedules, logistics routes, and utility consumption within a smart factory.

3. Course Syllabus

mcads-403b Data Science for Industry 4.0	Duration: 3 hours Max Marks: 70
---	--

Unit 1

The Industry 4.0 Ecosystem: Cyber-Physical Systems (CPS), smart factories, automation hierarchies (Automation Pyramid / ISA-95 standard), and cloud-to-edge topologies. **Industrial Protocols & Ingestion:** Interfacing with industrial data layers: MQTT, OPC-UA (Open Platform Communications Unified Architecture), Modbus, and streaming architectures via Kafka. **Time-Series Sensor Wrangling:** Handling high-frequency telemetry data streams, downsampling, alignment of asynchronous sensors, interpolation of missing readings, and handling data drift. **Edge Computing Frameworks:** Data preprocessing at the edge vs. cloud data lakes; stream windowing mechanics (tumbling, sliding, session) for continuous anomaly calculations.

Unit 2

Reliability Engineering Foundations: Failure modes, Mean Time Between Failures (MTBF), and degrading asset indicators. **Feature Engineering for Telemetry:** Extracting time-domain features (RMS, skewness, kurtosis) and frequency-domain features (Fast Fourier Transform - FFT, Power Spectral Density - PSD) from vibration, acoustic, and thermal arrays. **Predictive Maintenance Modeling:** Estimating Remaining Useful Life (RUL) using survival analysis, regression networks, and degrading health indices.

Unit 3

Automated Quality Inspection: Deploying Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) on assembly-line imagery for surface scratch, crack, and dimensional defect tracking. **Digital Twin Analytics:** Conceptualizing and building digital twin models; real-time synchronization between physical telemetry streams and physics-based analytical simulations. **Supply Chain and Process Optimization:** Prescriptive scheduling models using Linear Programming (LP) and Mixed-Integer Linear Programming (MILP) to optimize manufacturing queues, energy expenditure, and inventory supply routes.

4. Suggested Books and References

- [1] Data Science for Industry 4.0: Methods and Applications by Christoph Gröger and Stefan DeBloch (Springer).
- [2] Predictive Maintenance in the Industry 4.0 Era by Alireza Ghasemi and Mohammadreza Alizadeh (CRC Press).
- [3] The Internet of Things: Key Applications and Protocols by Olivier Hersent, David Boswarthick, and Omar Elloumi (Wiley).
- [4] Industrial Data Analytics and Machine Learning by Johannes Michaelis (Academic Press).

mcads-404a Data Science for Urban Analytics	Duration: 3 hours Max Marks: 70
--	--

1. Course Objectives

The primary objectives of this course are to:

Master the programmatic tools required to clean, fuse, and model multi-dimensional urban administrative and demographic datasets.

Teach the application of spatial econometrics to analyze housing markets, real estate dynamics, and neighborhood changes.

Build foundational practical skills in network analysis to model infrastructure accessibility, job access, and pedestrian flow.

Develop predictive pipelines for resource allocation, neighborhood change, and public budget distribution.

Train students to use data ethically within the public sphere, balancing transparency with spatial privacy.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Clean, cross-reference, and integrate heterogeneous municipal administrative records with national census demographic registries.

CO2: Apply spatial econometric models (SAR, SEM, GWR) to accurately measure localized socio-economic and real estate dynamics.

CO3: Construct topological network graphs to quantify structural accessibility gaps to essential public resources across varying urban neighborhoods.

CO4: Build machine learning classification frameworks to forecast indicators of structural neighborhood change, gentrification, and economic displacement.

CO5: Evaluate urban analytical predictive systems for compliance with civic transparency mandates, algorithmic equity, and spatial privacy frameworks.

3. Course Syllabus

mcads-404a Data Science for Urban Analytics	Duration: 3 hours Max Marks: 70
--	--

Unit 1

The Civic Data Landscape: Navigating municipal microdata, national census repositories (e.g., American Community Survey, Census Bureau datasets), open data portals, and crowd-sourced geographic information. **Administrative Data Wrangling:** Methods for stitching together disparate, non-standardized municipal data (e.g., merging business registries, tax assessments, and 311 service logs). **Spatial Autocorrelation:** The First Law of Geography. Quantifying spatial patterns using Global and Local Moran's I, Geary's C, and Getis-Ord Gi* statistics to detect geographic clusters.

Unit 2

Topological Urban Networks: Modeling city street networks, public transit grids, and utility paths as mathematical graphs using Python (NetworkX, OSMnx). **Accessibility and Centrality Modeling:** Computing closeness, betweenness, and straightness centralities on city street networks. Applying cumulative opportunity and gravity-based models to evaluate spatial access to healthcare, jobs, and healthy food options. **Active Transportation & Pedestrian Analytics:** Modeling pedestrian and micro-mobility flow dynamics, bike-share network optimization, and predicting sidewalk level-of-service metrics using topographic and behavioral constraints.

Unit 3

Predictive Risk Modeling: Building machine learning pipelines for predictive building inspections, identifying high-risk structural zones, and optimizing municipal service delivery (e.g., waste collection routes). **Modeling Neighborhood Change:** Applying classification and survival analysis models to historical census and real estate transactional trends to forecast patterns of gentrification, displacement, and commercial vacancy. **Urban Simulation & Land-Use Change:** Introduction to cellular automata and agent-based modeling (ABM) to simulate city growth patterns, zoning regulation impacts, and population migration.

4. Suggested Books and References

- [1] Urban Analytics by Alex D. Singleton, Seth Spielman, and David C. Folch (SAGE Publications)
- [2] Spatial Econometrics: Methods and Models by Luc Anselin (Springer).
- [3] Geocomputation with Python by Robin Lovelace, Jakub Nowosad, and Jannes Muenchow (CRC Press)

mcads-404b Revenue Management through Data Science	Duration: 3 hours Max Marks: 70
---	--

1. Course Objectives

The primary objectives of this course are to:

Introduce the core paradigms of Revenue Management and Pricing (RMP) along with the structural industry metrics that govern them.

Teach the mathematical derivation and algorithmic estimation of price elasticity and consumer demand curves.

Build advanced skills in optimization techniques for dynamic pricing, markdown strategies, and bundle generation.

Develop practical mastery over capacity allocation algorithms, booking limits, and overbooking risk models.

Train students to handle high-frequency competitor tracking, market telemetry, and algorithmic fairness constraints.

2. Course Outcomes (COs)

Upon successful completion of this course, students will be able to:

CO1: Map raw e-commerce and reservation transaction records into mathematically calibrated Multinomial Logit choice models.

CO2: Compute optimal inventory protection levels and booking limits using Littlewood's Rule and EMSR-b algorithms.

CO3: Formulate marginal risk optimization models to establish safety overbooking thresholds that balance spoil and spill costs.

CO4: Construct high-frequency dynamic pricing and markdown optimization engines utilizing time-variant elasticity metrics.

CO5: Implement Multi-Armed Bandit algorithms to perform low-risk price exploration in highly volatile or unknown market segments.

3. Course Syllabus

mcads-404b Revenue Management through Data Science	Duration: 3 hours Max Marks: 70
---	--

Unit 1

Foundations of Revenue Management: Historical evolution (deregulation of airlines), core pillars (perishable inventory, segmentable markets, advance bookings), and baseline metrics (RevPAR, Yield, Load Factor). **Price Elasticity & Demand Curves:** Continuous vs. discrete demand functions. Formulating and programmatically fitting linear, log-linear, and constant-elasticity demand models using regression toolkits. **Willingness-to-Pay (WTP) Estimation:** Leveraging historical transaction data, bidding metadata, and web clickstream interactions to estimate customer-specific reservation prices.

Unit 2

Quantity-Based Revenue Management: Managing fixed, perishable capacity. Single-resource allocation mechanics, booking limits, protection levels, and the mathematical implementation of **Littlewood's Rule**. **Network Revenue Management:** Scaling capacity control across multi-resource legs (e.g., connecting flights, multi-night hotel stays). Formulating and solving the Expected Marginal Seat Revenue (EMSR-b) algorithm and Linear Programming (LP) network models. **Overbooking Optimization Models:** Balancing the cost of empty inventory (spill) against the financial and reputational penalties of denied service (spoil). Determining optimal overbooking limits using marginal risk analysis.

Unit 3

Dynamic Pricing Algorithms: Formulating multi-period pricing paths for finite-horizon perishable inventories. Implementing time-dependent pricing, markdown optimization for retail inventory clearance, and surge pricing engines for matching two-sided marketplaces (e.g., ride-hailing). **Reinforcement Learning for Pricing:** Utilizing Multi-Armed Bandits (MAB), Thompson Sampling, and Upper Confidence Bound (UCB) algorithms for real-time price exploration vs. exploitation in uncertain demand environments. **Competitor Stream Ingestion & Telemetry:** Building real-time data ingestion pipelines to scrape and track competitor pricing variations, adjust localized strategies, and monitor price dispersion indexes.

4. Suggested Books and References

- [1] The Theory and Practice of Revenue Management by Kalyan T. Talluri and Garrett J. van Ryzin (Springer).
- [2] Pricing and Revenue Optimization by Robert L. Phillips (Stanford Business Books).
- [3] Dynamic Pricing and Automated Market Making by Junyi Chai (Academic Press).
- [4] Introduction to Scientific Computing with Python by Joakim Sundnes (Springer).
- [5] Open Dataset Frameworks: *The NASA Prognostics repositories, or airline/hospitality transaction logs available on Kaggle (e.g., Expedia hotel pricing datasets).*