

Rayat Shikshan Sanstha's
Karmaveer Bhaurao Patil College Vashi, Navi Mumbai
Autonomous College
[University of Mumbai]

Syllabus

Sr. No.	Heading	Particulars
1	Title of Course	M.Sc. II Data Science
2	Eligibility for Admission	M.Sc. I
3	Passing Marks	40%
4	Ordinances/Regulations (if any)	
5	No. of Years/Semesters	One year/Two semester
6	Level	P.G.
7	Pattern	Semester
8	Status	Revised
9	To be implemented from Academic year	2022-23

Scheme of examination for Each Semester:

Continuous Internal Evaluation: 40 Marks (Common Test-20 Marks & 20 Marks for- Mini Projects, Presentation, Online Course, Case Study, Assignment, Analysis In Statistics, Report Writing, Interviews, Paper Review, Surprise Test, Research Paper, Data Analysis).

Semester End Examination: 60 Marks will be as follows-

I.	Theory: The Semester End Examination for theory course work will be conducted as per the following scheme.	
	Each theory paper shall be of two and half hour duration.	
	All questions are compulsory and will have internal options.	
	Q – I	From Unit – I (having internal options.) 12 M
	Q – II	From Unit – II (having internal options.)12 M
	Q – III	From Unit – III (having internal options.)12 M
	Q – IV	From Unit – IV (having internal options.)12 M
	Q – V	Questions from all the FOUR Units with equal weightage of marks allotted to each Unit. 12 M
II.	Practical	The Semester End Examination for practical course will be conducted as per the following scheme.
Sr. No.	Particulars of Semester End Practical Examination	Marks
1	Laboratory Work	40
2	Journal	05
3	Viva	05
	TOTAL	50

AC - 25/10/2021
Item No - 7.7



**Rayat Shikshan Sanstha's
KARMAVEER BHURAO PATIL COLLEGE, VASHI.
NAVI MUMBAI
(AUTONOMOUS COLLEGE)
Sector-15- A, Vashi, Navi Mumbai - 400 703**

Syllabus for M.Sc. in Data Science

Program: M.Sc. Data Science

Course: M.Sc. Data Science

**(Choice Based Credit, Grading and Semester System
with effect from the academic year 2021-22)**

Rayat Shikshan Sanstha's
KARMAVEER BHAURAO PATIL COLLEGE, VASHI.
NAVI MUMBAI (Autonomous)
Department of Computer Science
M. Sc. Data Science

Program Outcomes (POs)

Learners are able to–

PO-1	Disciplinary Knowledge	Acquire the comprehensive and in-depth knowledge of various subjects in sciences such as Physics, Chemistry, Mathematics, Microbiology, Bio-analytical Science, Computer Science, Data Science, Information Technology and disciplinary skills and ability to apply these skills in the field of science, technology and its allied branches
PO-2	Communication and Presentation Skills	Develop various communication skills including presentation to express ideas evidently to achieve common goals of the organization.
PO-3	Creativity and Critical Judgment	Facilitate solutions to current issues based on investigations, evaluation and justification using evidence based approach.
PO-4	Analytical Reasoning and Problem Solving	Build critical and analytical attitude in handling the problems and situations.
PO-5	Sense of Inquiry	Curiously raise relevant questions based on highly developed ideas, scientific theories and its applications including research.
PO-6	Use of Digital Technologies	Use various digital technologies to explore information/data for business, scientific research and related purposes.
PO-7	Research Skills	Construct, collect, investigate, evaluate and interpret information/data relevant to science and technology to adapt, evolve and shape the future.
PO-8	Application of Knowledge	Develop a scientific outlook to create consciousness against the social myths and blind faith.
PO-9	Moral and Ethical Reasoning	Imbibe ethical, moral and social values to develop virtues such as justice, generosity and charity as beneficial to individuals and society at large.
PO-10	Leadership and Teamwork	Work cooperatively and lead proactively to achieve the goals of the organization by implementing the plans and projects in various field-based situations related to science, technology and society at large.
PO-11	Environment and Sustainability	Create social awareness about the environment and develop sustainability for betterment of the future.
PO-12	Lifelong Learning	Realize that pursuit of knowledge is a lifelong activity and in combination with determined efforts, positive attitude and other qualities to lead a successful life.

Department of Data Science
Program Specific Outcomes(PSO)

PSO-1	Understand the various processes involved in application development in the context of Data science.
PSO-2	Develop realistic solutions to meet the requirements of the society and the industry using the acquired data analytics skills

Preamble

M.Sc Data Science course syllabus is designed in a manner that covers all the aspects of Data Science.

The syllabus proposes to have three core Compulsory Courses, one Skill Enhancement Course and one Discipline Specific Elective course in Semester III. and Semester IV. The course gives insights into the practical and theoretical aspects of data science, Machine Learning, Business Analytics, Real-Time Processing, Neural Networks, Artificial Intelligence. The primary focus of the course is to equip the candidates of the course with principal concepts of data science and application of the same in real-time processing and applications.

Data science combines the knowledge of mathematics, computer science and statistics to solve exciting data-intensive problems in industry and in many fields of science. As data is collected and analyzed in all areas of society, demand for professional data scientists is high and will grow higher.

The syllabus also offers an Internship and Project implementation in the semester – IV, each of which has weights equivalent to a full course. This will definitely equip the student with industry readiness as an internship in an IT or IT-related organization gives a practical exposure to what is learned and what is practiced. The strong foundation given in the core courses in different semesters will give enough confidence to the learner to face and adapt to the changing trends and requirements of industry and academia.

As one can easily notice, the syllabus offers lots of emphasis on student driven learning and learning through experience. Research is embedded in the course structure.

We thank all the industry experts, senior faculties and our colleagues of different colleges as well as BOS members who have given their valuable comments and suggestions, which we tried to incorporate.

M.Sc. Data Science Syllabus
Choice Based Credit, Grading and Semester System
Academic year 2022-2023

SEMESTER - III

CODE	COURSE TYPE	SUBJECT	SCHEME OF INSTRUCTION		SCHEME OF EXAMINATION			NO. OF CREDITS
			(PERIOD PER WEEK)		(MAX MARKS)			
			TH	LAB	C I A	SE E	TOTA L	
PGDS301	CORE	MACHINE LEARNING	4	-	40	60	100	4
PGDSP301	Core Subject Practical	PRACTICAL ON MACHINE LEARNING	-	4	50			2
PGDS302	CORE	ARTIFICIAL INTELLIGENCE	4	-	100			4
PGDSP302		PRACTICAL ON ARTIFICIAL INTELLIGENCE	-	4	50			2
PGDS303	CORE	NATURAL LANGUAGE PROCESSING	4	-	100			4
PGDSP303		PRACTICAL ON NATURAL LANGUAGE PROCESSING	-	4	50			2
PGDS304	Skill Enhancement Course	DATA HANDLING AND VISUALIZATION	2	-		50	50	2
PGDSP304	Skill Enhancement Subject Practical	PRACTICAL ON DATA HANDLING AND VISUALIZATION	-	4	50			2
PGDS305A	Discipline Specific Elective - I OR	CASE STUDIES IN DATA SCIENCE	4	-	100			4

PGDSP305 A	Discipline Specific Elective - I Subject Practical OR	PRACTICAL ON CASE STUDIES IN DATA SCIENCE						4		50	2
PGDS305B	Discipline Specific Elective - II	DATA WRANGLING	4	-	4 0	60	100				4
PGDSP305 B	Discipline Specific Elective - I Subject Practical	PRACTICAL ON DATA WRANGLING	-	4						50	2
TOTAL										750	28

SEMESTER - IV

CODE	COURSE TYPE	SUBJECT	SCHEME OF INSTRUCTION		SCHEME OF EXAMINATION			NO. OF CREDITS
			(PERIOD PER WEEK)		(MAX MARKS)			
			TH	LAB	CIA	SEE	TOTAL	
PGDS401	CORE	Deep Learning	4	-	40	60	100	4
PGDSP401	Core Subject Practical	PGDS401	-	4			50	2
PGDS402	CORE	Internship	4	-			100	4
PGDSP402		Internship	-	4			50	2
PGDS403	CORE	Project	4	-			100	4
PGDSP403		Project	-	4			50	2
PGDS404	Skill Enhancement Course	Data Visualization using Tableau	2	-		50	50	2
PGDSP404	Skill Enhancement Subject Practical	PGDS404	-	4			50	2

PGDS405A	Discipline Specific Elective - I OR	MOOCs OR	4	-	100			4
PGDSP405 A	Discipline Specific Elective - I Subject Practical OR	PGDS405A		4	50			2
PGDS405B	Discipline Specific Elective - II	Cloud Computing	4	-	40	60	100	4
PGDSP405 B	Discipline Specific Elective - I Subject Practical	PGDS405B	-	4	50			2
TOTAL							750	28

Note: TH-Theory, CIA- Continuous Internal Assessment, SEE-Semester End Examination.

Detailed syllabus of Semester – III

Class: M.Sc Part II	Branch: Data Science	Semester: III	
Subject : Machine Learning			
Period per Week(Each 48 min)	Lecture	04	
	Practical	04	
Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40
	Semester End Practical Examination	2	50
	Total	—	150

PGDS301 Machine Learning

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Explain real-world applications that needs machine learning based solutions

CO-2: Implement and apply machine learning algorithms

CO-3: Recognize the characteristics of machine learning techniques that are useful to solve real-world problems

CO-PO Mapping Matrix

CO\PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS301	Machine Learning (Credits : 4 Lectures/Week: 4)
---------------------------	--

Unit I:	Introduction to Data and Machine Learning Essentials of Data and its analysis, Framework of Data Analysis, Machine Learning Basics: History of Machine Learning, Machine Learning Vs Statistical Learning, Types of Machine Learning Algorithms, Supervised Learning, Unsupervised Learning, Reinforcement Learning	12 L
Unit II:	Understanding Regression Analysis Linear Regression, Multiple Regression, Logistic Regression, Classification Techniques, Decision Tree, SVM, Naïve Bayes, KNN	12 L
Unit III:	Clustering K means clustering, Hierarchical Clustering, Practical Issues in Clustering. Probabilistic Models Uncertainty, Normal distribution and its geometric interpretations, Baye's theorem, Naïve Bayes Classifier, Bayesian network, Discriminative learning with maximum likelihood, Probabilistic models with hidden, variables, Hidden Markov model, Expectation Maximization methods, Gaussian Mixtures and compression based models	12L
Unit IV	Model Evaluation Introduction, Performance Measures, Confusion Matrix, Ensemble Methods Introduction, Bagging, Cross Validation	12L

Text book:

- Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, 3rd Edition
- Margaret H. Dunham, S. Sridhar, Data Mining - Introductory and Advanced Topics, Pearson Education
- Tom Mitchell, Machine Learning, McGraw-Hill, 1997
- R.O. Duda, P.E. Hart, D.G. Stork., Pattern Classification, Second edition. John Wiley and Sons, 2000.

References:

- Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer 2006
- Ian H. Witten, Data Mining: Practical Machine Learning Tools and Techniques, Eibe Frank Elsevier / (Morgan Kauffman)
- Bing Liu: Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, Springer (2006).
- Soumen Chakrabarti: Mining the Web: Discovering knowledge from hypertext data, Elsevier (2003).
- Christopher D Manning, Prabhakar Raghavan and Hinrich Schütze: An Introduction to Information Retrieval, Cambridge University Press (2009)

Links:

- <https://www.ibm.com/downloads/cas/GB8ZMQZ3>
- <https://www.edx.org/course/machine-learning-fundamentals-2>

[PGDSP301 Practical on Machine Learning](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Implement and apply machine learning algorithms

CO-2: Recognize the characteristics of machine learning techniques that are useful to solve real-world problems

CO-PO Mapping Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	1
CO2	2	2	-	2	-	-	-	-	-	-	-	1

Sr. No.	List of Practical Experiments on PGDSP301
1	Data Preprocessing – I
2	Data Preprocessing – II
3	Perform Feature Selection
4	Regression Analysis- Linear regression
5	Regression Analysis- Multiple regression
6	Regression Analysis- Logistic Regression
7	Classification Techniques- Decision tree
8	Classification Techniques- SVM
9	Classification Techniques- Naïve Bayes
10	Classification Techniques- KNN
11	Clustering- K- Means clustering
12	Clustering- Hierarchical Clustering
13	Implement Naive baye’s classifier
14	Implement Gaussian Mixtures based models
15	Implement Hidden Markov model
16	Market Basket Analysis
Note: Machine Learning Practicals Using R / Python	

Class: M.Sc Part II	Branch: Data Science	Semester: III
Subject : Artificial Intelligence		
Period per Week(Each 48 min)	Lecture	04
	Practical	04

Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40
	Semester End Practical Examination	2	50
	Total	—	150

[PGDS302 Artificial Intelligence](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Describe different types of problem-solving agents and its applications

CO-2: Solve problems using informed and uninformed search strategies

CO-3: Compare various Knowledge Representation Logic using scripts and frames

CO-4: Use expert system tools to realize the concepts and components of expert system.

CO-PO Mapping Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS302	Artificial Intelligence (Credits : 4 Lectures/Week: 4)	
	Expected Course Outcomes : After successful completion of this course, students will be able to: 1. Describe different types of problem-solving agents and its applications 2. Solve problems using informed and uninformed search strategies. 3. Compare various Knowledge Representation Logic using scripts and frames. 4. Use expert system tools to realize the concepts and components of expert system.	
Unit I:	Fundamentals of Artificial Intelligence Introduction, A.I. Representation, Non-AI & AI Techniques, Representation of Knowledge, KnowledgeBase Systems, State Space Search, Production Systems, Problem Characteristics, types of production systems, Intelligent Agents and Environments, concept of rationality, the nature of environments, structure of agents, problem solving agents, problem formulation	15 L
Unit II:	Uninformed Search Strategies Formulation of real world problems, Breadth First Search, Depth First Search,	15 L

	Depth Limited Search, Iterative Deepening Depth First Search, Bidirectional Search, Comparison of Uninformed search Strategies, Searching with partial information, Sensor-less problems, Contingency problems. Informed Search Strategies Generate& test, Hill Climbing, Best First Search, A* and AO* Algorithm, Constraint satisfaction, Game playing: Minimax Search, Alpha-Beta Cutoffs, Waiting for Quiescence	
Unit III:	Knowledge Representation Knowledge based agents, Wumpus world. Propositional Logic: Representation, Inference, Reasoning Patterns, Resolution, Forward and Backward Chaining. First order Logic: Representation, Inference, Reasoning Patterns, Resolution, Forward and Backward Chaining. Basics of PROLOG: Representation, Structure, Backtracking. Expert System: Case study of Expert System in PROLOG	15 L
Unit IV:	Introduction to Planning and ANN Blocks world, STRIPS, Implementation using goal stack, Introduction to Neural networks:- basic, comparison of human brain and machine, biological neuron, general neuron model, activation functions, Perceptron learning rule, applications and advantages of neural networks. Brief introduction to single layer and multiplayer networks. Uncertainty Non Monotonic Reasoning, Logics for Non Monotonic Reasoning, Justification based Truth Maintenance Systems, Semantic Nets, Statistical Reasoning, Fuzzy logic: fuzzy set definition and types, membership function, designing a fuzzy set for a given application. Probability and Bayes' theorem, Bayesian Networks.	15 L
Text book:		
<ul style="list-style-type: none"> ● Elaine Rich and Kevin Knight: "Artificial Intelligence." Tata McGraw Hill ● Stuart Russell & Peter Norvig : "Artificial Intelligence : A Modern Approach", Pearson Education, 2nd Edition. 		
References:		
<ul style="list-style-type: none"> ● Ivan Bratko : "Prolog Programming For Artificial Intelligence" , 2nd Edition Addison Wesley, 1440. ● Eugene, Charniak, Drew Mcdermott: "Introduction to Artificial Intelligence.", Addison Wesley ● Patterson: —Introduction to AI and Expert Systems, PHI ● Nilsson : —Principles of Artificial Intelligence, Morgan Kaufmann. ● Carl Townsend, —Introduction to turbo Prolog, Paperback, 14836. Jacek M. Zurada, Introduction to artificial neural systems, Jaico Publication 		
Links:		
<ul style="list-style-type: none"> ● http://www.eecs.qmul.ac.uk/~mmh/AINotes/AINotes4.pdf ● https://www.slideshare.net/JismyKJose/conceptual-dependency-70129647 		

PGDSP302 Practicals on Artificial Intelligence

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Solve problems using informed and uninformed search strategies

CO-2: Use expert system tools to realize the concepts and components of expert system.

CO-PO Mapping Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-

Sr. No.	List of Practical Experiments on PGDSP302
1	Implement Non-AI and AI Techniques
2	Implement Breadth First Search
3	Implement Depth First Search
4	Implement Depth Limited Search
5	Implement Iterative Deepening Depth Limited Search
6	Implement Best First Search A* algorithm
7	Implement AO* algorithm
8	Implement Hill Climbing
9	Implement Perceptron learning algorithm
10	Implement a real life application in Prolog.
11	Expert System in Prolog-new application
12	Implement any two Player game using min-max search algorithm.
13	Design a fuzzy set for shape matching of handwritten character
14	Implement a general neuron model.
15	Implementation using goal stack.

Class: M.Sc Part II	Branch: Data Science	Semester: III	
Subject : Natural Language Processing			
Period per Week(Each 48 min)	Lecture	04	
	Practical	04	
Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40

	Semester End Practical Examination	2	50
	Total	—	150

[PGDS303 Natural Language Processing](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Describe the basics of Natural language processing

CO-2: Analyze the text syntactically

CO-3: Analyze the text content Semantically

CO-4: Implement recurrent network for language models.

CO-PO Mapping Matrix

CO\PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS303	Natural Language Processing (Credits : 4 Lectures/Week: 4)	
Unit I:	INTRODUCTION Introduction to NLP, Regular Expressions, Words, Corpora, Text Normalization, Minimum Edit distance, N gram Language Models, Evaluating Language Models	15 L
Unit II:	SYNTACTIC ANALYSIS English Word Classes, The Penn Treebank Part-of-Speech Tagset, Part-of-Speech Tagging, HMM Part-of-Speech Tagging, Maximum Entropy Markov Models, Grammar Rules for English, Treebanks, Grammar Equivalence and Normal form, Lexicalized Grammar.	15 L
Unit III:	SEMANTIC ANALYSIS Representation of Sentence Meaning: Computational Desiderata for Representations, Model-Theoretic Semantics, First-Order Logic, Event and State Representations, Description Logics, Semantic roles, Semantic role labeling.	15 L
Unit IV:	SEQUENCE PARSING WITH RECURRENT NETWORKS Simple Recurrent Networks, Applications of RNNs, Deep Networks: Stacked and Bidirectional RNNs, Managing Context in RNNs: LSTMs and GRUs, Words, Characters and Byte-Pairs.	15 L

	Case Study Sentiment Classification, Dialog Systems and Chatbots
Text book:	
<ul style="list-style-type: none"> • Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft), 2019. 	
References:	
<ul style="list-style-type: none"> • Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, First Edition, O'reilly, 2009 • Yoav Goldberg, University of Toronto, Neural Network Methods for Natural language Processing, Morgan & Claypool, 2017 • Christopher D. Manning, and Hinrich Schütze. Foundations of statistical natural language processing. First Edition, MIT press, 1999 	
Links:	
<ul style="list-style-type: none"> • https://www.cs.vassar.edu/~cs366/docs/Manning_Schuetze_StatisticalNLP.pdf • https://www.nltk.org/book/ 	

<u>PGDSP303 Practicals on Natural Language Processing</u>												
Course Outcomes: After successful completion of this course, students will be able to:												
CO-1: Analyze the text syntactically and text content Semantically.												
CO-2: Implement recurrent network for language models.												
CO-PO Mapping Matrix												
CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	1
CO2	2	2	-	2	-	-	-	-	-	-	-	1

Sr. No.	List of Practical Experiments on PGDSP303
1	a. Convert the text into tokens b. Find the word frequency c. Demonstrate a bigram language model d. Demonstrate a trigram language model e. Generate regular expression for a given text f. Text Normalization
2	a. Perform Lemmatization b. Perform Stemming c. Identify parts-of Speech using Penn Treebank tag set. d. Implement HMM for POS tagging e. Build a Chunker f. Text summarization
3	a. Find the synonym of a word using WordNet b. Find the antonym of a word

CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS304	Data Handling and Visualization (Credits : 2 Lectures/Week: 2)	
Unit I:	INTRODUCTION TO VISUALIZATION Visualizing Data-Mapping Data onto Aesthetics, Aesthetics and Types of Data, Scales Map Data Values onto Aesthetics, Coordinate Systems and Axes-Cartesian Coordinates, Nonlinear Axes, Coordinate Systems with Curved Axes, Color Scales-Color as a Tool to Distinguish, Color to Represent Data Values ,Color as a Tool to Highlight, Directory of Visualizations- Amounts, Distributions, Proportions, x-y relationships, Geospatial Data	15 L
Unit II:	VISUALIZING DISTRIBUTIONS Visualizing Amounts-Bar Plots, Grouped and Stacked Bars, Dot Plots and Heatmaps, Visualizing Distributions: Histograms and Density Plots- Visualizing a Single Distribution, Visualizing Multiple Distributions at the Same Time, Visualizing Distributions: Empirical Cumulative Distribution Functions and Q-Q Plots-Empirical Cumulative Distribution Functions, Highly Skewed Distributions, Quantile-Quantile Plots, Visualizing Many Distributions at Once-Visualizing Distributions Along the Vertical Axis, Visualizing Distributions Along the Horizontal Axis	15 L
Unit III:	VISUALIZING ASSOCIATIONS Visualizing Proportions-A Case for Pie Charts, A Case for Side-by-Side Bars, A Case for Stacked Bars and Stacked Densities, Visualizing Proportions Separately as Parts of the Total ,Visualizing Nested Proportions- Nested Proportions Gone Wrong, Mosaic Plots and Treemaps, Nested Pies ,Parallel Sets. Visualizing Associations Among Two or More Quantitative Variables-Scatterplots, Correlograms, Dimension Reduction, Paired Data.	15 L
Text book:		
<ul style="list-style-type: none"> ● Claus Wilke, “Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures”, 1st edition, O’Reilly Media Inc, 2019. 		
References:		
<ul style="list-style-type: none"> ● Tony Fischetti, Brett Lantz, R: Data Analysis and Visualization,O’Reilly ,2016 ● Ossama Embarak, Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems,Apress, 2018 		
Links:		
<ul style="list-style-type: none"> ● https://www.netquest.com/hubfs/docs/ebook-data-visualization-EN.pdf ● https://www.coursera.org/learn/data-visualization ● https://clauswilke.com/dataviz/time-series.html 		

PGDSP304 Practicals on Data Handling and Visualization

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Implement visualization of distributions

CO-2: Write programs on visualization of time series, proportions & associations

CO-3: Apply visualization on Trends and uncertainty.

CO-PO Mapping Matrix

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	3	1	-	-	-	-	-
CO2	2	2	-	2	-	3	1	-	-	-	-	-
CO3	2	1	2	1	-	3	1	-	-	-	-	-

Sr. No.	List of Practical Experiments on PGDSP304
1	Consist of at least 10 practical experiments covering the topics of the syllabus.

Class: M.Sc Part II		Branch: Data Science		Semester: III	
Subject : Case Studies in Data Science					
Period per Week(Each 48 min)	Lecture		04		
	Practical		04		
Evaluation System			Hours	Marks	
	Semester End Examination		2 hrs 30 min	60	
	Continuous Internal Assessment		—	40	
	Semester End Practical Examination		2	50	
	Total		—	150	

PGDS305A Case Studies in Data Science

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Define steps for solving case studies of different domains.

CO-2: Illustrate analytical tool's features for problem solving, filtering data and visualization.

CO-3: Apply the data manipulation and transformation techniques to prepare data ready for processing.

CO-4: Analyze data with the help of visualization and predict results.

CO-PO Mapping Matrix

CO/P O	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS305 A	Case Studies in Data Science (Credits : 4 Lectures/Week: 4)	
	<p>Expected Course Outcomes : After successful completion of this course, students will be able to:</p> <ol style="list-style-type: none"> 1. Define steps for solving case studies of different domains. 2. Illustrate analytical tool's features for problem solving, filtering data and visualization. 3. Apply the data manipulation and transformation techniques to prepare data ready for processing. 4. Analyze data with the help of visualization and predict results. 	
Unit I:	<p>Brief Introduction to Data Science What is Data Science, Why Now? - The importance of Data Science in today's business environment., Difference between Data Science, Business Intelligence and Data Analysis, Real World Applications of Data Science, Popular Tools and Languages to Apply Data Science, A Typical Data Science Team, Team Structures</p>	15 L
Unit II:	<p>Brushing Up on your MS Excel Skills Basic Functions, Advanced Functions, Sorting, Filtering and Pivot tables Charts and other visualization tools ,Handling Date formats for time series analysis, Short Cut Keys, Making simple Dashboards</p>	15 L
Unit III:	<p>Making Data Work for You! What can Data do?, Types of Data, Data Exploration ,Data Sources and Risks, Data Cleaning, Manipulating Time Series Data, Data Storage and Retrieval Analysis, Prediction and Visualization Generating Charts and plots to better understand the output, Interpretation, Prediction, Conclusion Communicate Results Creating Dashboards (MS Excel), Generate High End Presentations in MS PowerPoint</p>	15 L
Unit IV:	<p>The Data Science Workflow Understand the steps in the life cycle of a Data Science project, Theoretical Vs Applied Data Science Structure of Case Study Problem Statement – Domain and scope of the study, Data</p>	15 L

	Source, Data Preparation, Choice of Model, Model Building, Type of tools to use, Analysis, Presentation	
Text book:		
<ul style="list-style-type: none"> Microsoft Excel Data Analysis and Business Modeling, By Wayne L. Winston, Published by Microsoft Press 2019 		
References:		
<ul style="list-style-type: none"> Ryan Sleeper, Practical Tableau, O'Reilly publication, 2019. 		
Links:		
<ul style="list-style-type: none"> Python Data Science Hand Book https://tanthiamhuat.files.wordpress.com/2018/04/pythondatasciencehandbook.pdf 		

PGDSP305A Practicals on Case Studies in Data Science

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Illustrate analytical tool's features for problem solving, filtering data and visualization.

CO-2: Apply the data manipulation and transformation techniques to prepare data ready for processing.

CO-3: Analyze data with the help of visualization and predict results.

CO-PO Mapping Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-

Practical Experiments on PGDSP305A

Any five Case Study on Data Science

Class: M.Sc Part II	Branch: Data Science	Semester: III	
Subject : Data Wrangling			
Period per Week(Each 48 min)	Lecture	04	
	Practical	04	
Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40

	Semester End Practical Examination	2	50
	Total	—	150

[PGDS305B Data Wrangling](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Identify and execute the basic data format.

CO-2: Perform the computations with Excel and pdf file

CO-3: Describe the concepts of data cleanup.

CO-4: Analyze the Image and video data.

CO-PO Mapping Matrix

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS305 B	Data Wrangling (Credits : 4 Lectures/Week: 4)	
Unit I:	<p>INTRODUCTION TO DATA WRANGLING: What Is Data Wrangling?- Importance of Data Wrangling -How is Data Wrangling performed?- Tasks of Data Wrangling-Data Wrangling Tools-Introduction to Python-Python Basics-Data Meant to Be Read by Machines-CSV Data-JSON Data-XML Data.</p> <p>WORKING WITH EXCEL FILES AND PDFS Installing Python Packages-Parsing Excel Files-Parsing Excel Files -Getting Started with Parsing-PDFs and Problem Solving in Python-Programmatic Approaches to PDF Parsing-Converting PDF to Text-Parsing PDFs Using pdf miner-Acquiring and Storing Data-Databases: A Brief Introduction-Relational Databases: MySQL and PostgreSQL-Non-Relational Databases: NoSQL-When to Use a Simple File-Alternative Data Storage.</p>	15 L
Unit II:	<p>DATA CLEANUP Why Clean Data?- Data Cleanup Basics-Identifying Values for Data Cleanup-Formatting Data-Finding Outliers and Bad Data-Finding Duplicates-Fuzzy Matching-RegEx Matching-Normalizing and Standardizing the Data-Saving the Data-Determining suitable Data Cleanup-Scripting the Cleanup-Testing with New Data</p>	15 L

Unit III:	<p>DATA EXPLORATION AND ANALYSIS</p> <p>Exploring Data-Importing Data-Exploring Table Functions-Joining Numerous Datasets-Identifying Correlations-Identifying Outliers-Creating Groupings-Analyzing Data-Separating and Focusing the Data-Presenting Data-Visualizing the Data-Charts-Time-Related Data-Maps-Interactives-Words-Images, Video, and Illustrations-Presentation Tools-Publishing the Data-Open Source Platforms.</p>	15 L
Unit IV:	<p>WEB SCRAPING</p> <p>What to Scrape and How-Analyzing a Web Page-Network/Timeline-Interacting with JavaScript-In-Depth Analysis of a Page-Getting Pages-Reading a Web Page-Reading a Web Page with LXML-XPath-Advanced Web Scraping-Browser-Based Parsing-Screen Reading with Selenium-Screen Reading with Ghost.Py-Spidering the Web-Building a Spider with Scrapy-Crawling Whole Websites with Scrapy.</p>	
<p>Text book:</p> <ul style="list-style-type: none"> • Jacqueline Kazil & Katharine Jarmul,” Data Wrangling with Python”, O’Reilly Media, Inc,2016 		
<p>References:</p> <ul style="list-style-type: none"> • Dr. Tirthajyoti Sarkar, Shubhadeep,” Data Wrangling with Python: Creating actionable data from raw sources”, Packt Publishing Ltd,2019. • Stefanie Molin,” Hands-On Data Analysis with Pandas”, Packt Publishing Ltd,2019 • Allan Visochek,” Practical Data Wrangling”, Packt Publishing Ltd,2017 • Tye Rattenbury, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, Connor Carreras,” Principles of Data Wrangling: Practical Techniques for Data Preparation”, O’Reilly Media, Inc,2017 <p>Links:</p> <ul style="list-style-type: none"> • http://www.gbv.de/dms/ilmeneau/toc/827365454.PDF • https://www.udemy.com/course/data-wrangling-with-python/ 		

PGDSP305B Practicals on Data Wrangling

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Identify and execute the basic data format.

CO-2: Perform the computations with Excel and pdf file

CO-3: Analyze the Image and video data.

CO-PO Mapping Matrix

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	2	-	-	-	-	-	1
CO2	2	2	-	2	-	2	-	-	-	-	-	1
CO3	2	1	2	1	-	2	-	-	-	-	-	1

Sr. No.	List of Practical Experiments on PGDSP305B
1	Write the python script to Read the XML file
2	Develop the python script to parse the pdf files using pdfminer.
3	Extract the Table from the child labour and child marriage data.xlsx using pdfables library
4	Write a Python data wrangling scripts to insert the data into SQLite database
5	Develop the Python Shell Script to do the basic data cleanup on child labour and child marriage data.xlsx a. Check duplicates and missing data b. Eliminate Mismatches c. Cleans line breaks, spaces, and special characters
6	Import the data into `agate` then explores the table using agate methods and perform statistical correlations
7	Draw the chart between perceived corruption scores compared to the child labour percentages using matplotlib.
8	Write the python script to Map the Child Labour Worldwide using pygal.
9	Write a Python program to download and display the content of robot.txt for en.wikipedia.org

Detailed syllabus of Semester – IV

Class: M.Sc Part II	Branch: Data Science	Semester: IV	
Subject : Deep Learning			
Period per Week(Each 48 min)	Lecture	04	
	Practical	04	
Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40
	Semester End Practical Examination	2	50
	Total	—	150

[PGDS401 Deep Learning](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Design simple Neural Networks using Linear Perceptron.

CO-2: Implement Convolutional Neural Networks using TensorFlow.

CO-3: Develop an application based on Recurrent Neural Network.

CO-PO Mapping Matrix												
CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-

Course: PGDS401	Deep Learning (Credits : 4 Lectures/Week: 4)											
Unit I:	NEURAL NETWORK Mechanics of Machine Learning-Neuron-Linear Perceptron-Feed-Forward Neural Networks-Sigmoid, Tanh, and ReLU Neurons- Training Feed-Forward Neural Networks-Fast-Food Problem-Gradient Descent-Delta Rule and Learning Rates.											15 L
Unit II:	CONVOLUTIONAL NEURAL NETWORKS TensorFlow: Creating and Manipulating TensorFlow Variables-TensorFlow Operations-Neurons in Human Vision-Convolutional Layer-Building a Convolutional Network-Visualizing Learning in Convolutional Networks-Learning Lower Dimensional Representations- Principal Component Analysis- Autoencoder Architecture- Implementing an Autoencoder in TensorFlow.											15 L
Unit III:	RECURRENT NEURAL NETWORKS Recurrent Neural Networks- Challenges with Vanishing Gradients- Long Short-Term Memory (LSTM) Units- TensorFlow Primitives for RNN Models- Implementing a Sentiment Analysis Model- Solving seq2seq Tasks with Recurrent Neural Networks- Memory Augmented Neural Networks:Neural Turing Machines, Attention-Based Memory Access, Differentiable neural Computers (DNC) -Memory Reuse - Temporal Linking - DNCController Network – Visualizing – Implementing the DNC in TensorFlow.											15 L
Unit IV:	DEEP REINFORCEMENT LEARNING Deep Reinforcement Learning - Masters Atari Games-Markov Decision Processes- Policy Versus Value Learning, Pole-Cart with Policy Gradients-Q-Learning and Deep Recurrent-v-Q-Networks. APPLICATIONS Applications in Object Recognition and Computer Vision- Unsupervised or generative feature learning- Supervised feature learning and classification- Applications in Multimodal and Multi-task Learning- Multi- modalities: Text and image-Speech and image- Multi-task learning within the speech, NLP or image domain											15 L
Text book:												
<ul style="list-style-type: none"> Nikhil Buduma, Nicholas Locascio, “Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms”, O'Reilly Media, 2017. 												

- Li Deng and Dong Yu “Deep Learning Methods and Applications”, Foundations and Trends in Signal Processing, 2013.

References:

- Ian Goodfellow, YoshuaBengio, Aaron Courville, ”Deep Learning (Adaptive Computation and Machine Learning series”, MIT Press, 2017.
- SandroSkansi“Introduction to Deep Learning From Logical Calculus to Artificial Intelligence”Springer, 2018.
- Michael Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.

Links:

- <https://www.deeplearningbook.org/>
- <https://pythonmachinelearning.pro/free-ebook-deep-learning-with-python/>
- <https://www.getfreebooks.com/deep-learning/>

PGDSP401 Practicals on Deep Learning

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Design simple Neural Networks using Linear Perceptron.

CO-2: Implement Convolutional Neural Networks using TensorFlow.

CO-3: Develop an application based on Recurrent Neural Network.

CO-PO Mapping Matrix

CO\PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-

Sr. No.	List of Practical Experiments on PGDSP401
1	a. Write a program in Python to Calculate the output of a simple neuron b. Construct a Perceptron for the classification of data in Python c. Develop the Python code to Classify the 4-class problem with Multi-layer Perceptron d. Implement the back propagation algorithm for neural networks and apply it to the task of hand-written digit recognition
2	a. Build a simple deep neural network with many layers in Python using TensorFlow b. Implement binary classification for medical diagnosis for a single medical condition like say disease vs. no disease based on a battery of tests. c. Explore multi-class with Rock Paper Scissors dataset d. Implement an Autoencoder in TensorFlow.
3	a. Implementing a Sentiment Analysis Model in TensorFlow b. Solve seq2seq Tasks with Recurrent Neural Networks using TensorFlow c. Implementing the DNC in TensorFlow

4	<ul style="list-style-type: none"> a. Implement a policy-gradient agent to solve pole-cart-reinforcement learning problem. b. Implementing Experience Replay in Q-Network using TensorFlow
5	<ul style="list-style-type: none"> a. Build a model to classify movie reviews as positive or negative using TensorFlow b. Develop the CNN Model for Image Classification

GUIDELINES FOR INTERNSHIP IN SEMESTER – IV (PGDS402)

- Internship should be of 2 to 3 months with 8 to 12 weeks duration.
- A student is expected to find an internship by himself or herself. However, the institution should assist their students in getting internships in good organizations.
- The home institution cannot be taken as the place of internship.
- A student is expected to devote at least 300 hours physically at the organization.
- Internship can be on any topic covered in the syllabus mentioned in the syllabus, not restricted to the specialization.
- Internship can be done, in one of the following, but not restricted to, types of organizations:
 - Software development firms
 - Hardware/ manufacturing firms
 - Any small scale industries, service providers like banks
 - Clinics/ NGOs/professional institutions like that of CA, Advocate etc
 - Civic Depts like Ward office/post office/police station/ punchayat.
 - Research Centres/ University Depts/ College as research Assistant for research projects or similar capacities.

GUIDELINES FOR MAKING INTERNSHIP REPORT IN SEMESTER –IV (PGDSP402)

A student is expected to make a report based on the internship he or she has done in an organization. It should contain the following:

- **Certificate:** A certificate in the prescribed Performa (given in appendix 1) from the organization where the internship was done.
- **Evaluation form:** The form filled by the supervisor or to whom the intern was reporting, in the prescribed Performa (given in appendix 2).
- **Title:** A suitable title giving the idea about what work the student has performed during the internship.
- **Description of the organization:** A small description of 1 to 2 pages on the organization where the student has interned
- **Description about the activities done by the section where the intern has worked:** A description of 2 to 4 pages about the section or cell of the organization where the intern actually worked. This should give an idea about the type of activity a new employee is expected to do in that section of the organization.
- **Description of work allotted and actually done by the intern:** A detailed description of the work allotted and actual work performed by the intern during the internship period. Intern may give a weekly report of the work by him or her if needed. It shall be of around 7 to 10 pages.

- **Self assessment:** A self assessment by the intern on what he or she has learnt during the internship period. It shall contain both technical as well as inter personal skills learned in the process. It shall be of around 2 to 3 pages.

The internship report may be around 15 pages and this needs to be submitted to the external examiner at the time of University examination.

GUIDELINES FOR RESEARCH IMPLEMENTATION IN SEMESTER – IV (PGDS403)

- A student is expected to devote at least 3 to 4 months of efforts for the implementation.
- Student should submit a detailed project implementation report at the time of viva.

GUIDELINES FOR DOCUMENTATION OF PROJECT PROPOSAL IN SEMESTER –IV (PGDSP403)

A Student should submit project implementation report with following details:

- **Title:** Title of the project
- **Implementation details:** A description of how the project has been implemented. It shall be of 2 to 4 pages.
- **Experimental set up and results:** A detailed explanation on how experiments were conducted, what software used and the results obtained. Details like screenshots, tables and graphs can come here. It shall be of 6 to 10 pages.
- **Analysis of the results:** A description on what the results means and how they have been arrived at. Different performing measures or statistical tools used etc may be part of this. It shall be of 4 to 6 pages.
- **Conclusion:** A conclusion of the project performed in terms of its outcome (May be half a page).
- **Future enhancement:** A small description on what enhancement can be done when more time and resources are available (May be half a page).
- **Program code:** The program code may be given as an appendix. The report may be of around 20 pages (excluding program code), which needs to be signed by the teacher in charge and head of the Department. Students should submit the signed project implementation report along with an evaluated copy of the project proposal documentation at the time of Project evaluation and viva.

Appendix 1

(Proforma for the certificate for internship in official letter head)

This is to certify that Mr/Ms _____ of
_____ College/Institution worked as an intern as part of her MSc
course in Computer Science of University of Mumbai. The particulars of internship are
given below:

Internship starting date: _____

Internship ending date: _____

Actual number of days worked: _____

Tentative number of hours worked: _____ Hours

Broad area of work: _____

A small description of work done by the intern during the period:

Signature:

Name:

Designation:

Contact number:

Email:

(seal of the organization)

Appendix 2

(Proforma for the Evaluation of the intern by the supervisor/to whom the intern was reporting in the organization)

Professional Evaluation of intern

Name of intern: _____

College/institution: _____

[Note: Give a score in the 1-5 scale by putting √ in the respective cells]

Sr. No.	Particular	Excellent	Very Good	Good	Moderate	Satisfactory
1	Attendance					
2	Punctuality					
3	Adaptability					
4	Ability to shoulder responsibility					
5	Ability to work in a team					
6	Written and oral communication skills					
7	Problem solving skills					
8	Ability to grasp new concepts					
9	Ability to complete task					
10	Quality of work done					

Comments:

Signature:

Name:

Designation:

Contact number:

Email:

(seal of the organization)

Class: M.Sc Part II	Branch: Data Science	Semester: IV	
Subject : Data Visualization in Tableau			
Period per Week(Each 48 min)	Lecture	04	
	Practical	04	
Evaluation System		Hours	Marks
	Semester End Examination	2 hrs 30 min	60
	Continuous Internal Assessment	—	40
	Semester End Practical Examination	2	50
	Total	—	150

[PGDS404 Data Visualization in Tableau](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Understand types of data and data visualization methods

CO-2: Understand the need of data visualization.

CO-3: Create and design visualizations and dashboards

CO-4: Evaluate the performance of visualization technique

CO-5: Apply data visualization using open source tool Tableau

CO-PO Mapping Matrix

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4	1	1	1		-	1	-	-	-	-	-	-
CO5	3	2	1	1		1		1				1

Course: PGDS404	Data Visualization in Tableau (Credits : 2 Lectures/Week: 2)
---------------------------	---

Unit I:	<p>Introduction to Data Visualization: Need for data visualization. Visualization an aspect of business analytics, importance of data visualization. Types of Data, Stages of Data visualization, Fitts Law, Human visual perception and cognition Comparison between tableau and Power BI</p> <p>Installation of Tableau: Using the Workspace Control Effectively, Tableau Desktop Workspace, Navigation, Data Terminology, Data Types & Default Properties, different aggregation types, File Types</p> <p>Data Connection: Data Connection with Text File, Connection with Microsoft Excel, Extracting data, data joining, data blending, sorting and replacing data source.</p>	15 L
Unit II:	<p>Tableau Calculation: Operations, functions, string, Numeric, date, table calculation, LOD Expressions</p> <p>Filter data: Basic filter, filter operation, Extract filter, Quick Filters, Context filter, conditional filters, data source filters, Sort data, Build Groups, Hierarchy and sets</p> <p>Tableau Charts and Graphs: Bar chart, Line Graphs with Date & Without Date, Pie Chart, Tree maps, Scatter Plots, Individual Axes, Blended Axes, Dual Axes & Combination chart, Edit axis, Bins/Histograms, Parts of Views, Sorting, Trend lines, Reference Lines, Forecasting, View data & Actions (across sheets), latitude and longitude, Default location/Edit locations, Symbol Map & Filled Map</p>	15 L
Unit III:	<p>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</p>	15 L
<p>Text Books:</p> <ul style="list-style-type: none"> ● Learning Tableau 10 – Second Edition, by Joshua Milligan ● Practical Tableau by Ryan Sleeper ● Communicating Data with Tableau by Ben Jones ● Mastering Tableau by David Baldwin 		
<p>References:</p> <ul style="list-style-type: none"> ● Big data black book, Dream tech publication ● Handbook for visualizing : a handbook for data driven design by Andy krik 		
<p>Links:</p> <p>https://www.tutorialspoint.com/tableau/index.htm</p> <p>https://medium.com/analytics-vidhya/quick-notes-on-tableau-8596f1b6009</p>		

[PGDSP404 Practicals on Data Visualization in Tableau](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Create and design visualizations and dashboards

CO-2: Evaluate the performance of visualization technique

CO-3: Apply data visualization using open source tool Tableau

CO-PO Mapping Matrix

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
--------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------	-------------	-------------

CO1	2	1	2	1	-	-	-	-	-	-	-	-
CO2	1	1	1		-	1	-	-	-	-	-	-
CO3	3	2	1	1		1		1				1

Sr. No.	List of Practical Experiments on PGDSP404
1	Installation of Tableau and working with Tableau Desktop Workspace
2	Data Connection with various File
3	Import and manage data (join, relationship, replace)
4	Perform various Tableau Calculation
5	Adding filters and quick filters to dashboards
5	Implement Bar charts, Pie chart, Line chart, Multiple chart and distribution
6	Implement Highlight tables, Scatterplot, Trendline
7	Implement Heatmap, Geographic mapping, Impressive bar chart, bullet graph
8	Implement Gantt chart, data calendar, circle view, general operation
9	Building Interactive Dashboards

Class: M.Sc Part II		Branch: Data Science		Semester: IV	
Subject : Cloud Computing					
Period per Week(Each 48 min)	Lecture		04		
	Practical		04		
Evaluation System			Hours	Marks	
	Semester End Examination		2 hrs 30 min	60	
	Continuous Internal Assessment		—	40	
	Semester End Practical Examination		2	50	
	Total		—	150	

[PGDS405B Cloud Computing](#)

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Describe the concepts of Cloud Computing and its Service Models & Deployment Models

CO-2: Classify the types of Virtualization

CO-3: Describe cloud security

CO-4:

CO-PO Mapping Matrix

CO\PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-
CO2	2	2	-	2	-	-	-	-	-	-	-	-
CO3	2	1	2	1	-	-	-	-	-	-	-	-
CO4					-	-	-	-	-	-	-	-

Course: PGDS405 B	Cloud Computing (Credits : 4 Lectures/Week: 4)	
	Expected Course Outcomes : After successful completion of this course, students will be able to: 1.	
Unit I:	Introduction to Cloud Computing Overview, Layers and Types of Cloud, Desired Features of a Cloud, Benefits and Disadvantages of Cloud Computing, Cloud Infrastructure Management, Infrastructure as a Service Providers, Platform as a Service Providers, Multitenant Technology. Cloud-Enabling Technology: Broadband Networks and Internet Architecture, Data Center Technology, Virtualization Technology. Infrastructure as a Service, Platform as a Service, Software as a Service, Cloud Deployment Models.	15 L
Unit II:	Abstraction and Virtualization Introduction to Virtualization Technologies, Load Balancing and Virtualization, Understanding Hyper visors, Virtual Machines Provisioning and Manageability Virtual Machine Migration Services, Provisioning in the Cloud Context Virtualization of CPU, Memory , I/O Devices, Virtual Clusters and Resource management	15 L
Unit III:	Programming, Environments and Applications Features of Cloud and Grid Platforms, Programming Support of Google App Engine, Programming on Amazon AWS and Microsoft Azure, Emerging Cloud Software Environments, Applications: Moving application to cloud, Microsoft Cloud Services, Google Cloud Applications, Amazon Cloud Services, Cloud Applications.	15 L
Unit IV:	Security In The Cloud	15 L

	Security Overview – Cloud Security Challenges and Risks – Software-as-a-Service Security – Security Governance – Risk Management – Security Monitoring – Security Architecture Design – Data Security – Application Security – Virtual Machine Security - Identity Management and Access Control, Disaster Recovery in Clouds.	
--	--	--

Text Books:

- Cloud Computing: Technologies and Strategies of the Ubiquitous Data Center, Brian J.S. Chee and Curtis Franklin, CRC Press, ISBN :9781439806128
- Rajkumar Buyya, Christian Vecchiola, S. ThamaraiSelvi, Mastering Cloud Computing: Foundations and Applications Programming, McGraw Hill, ISBN: 978 1259029950, 1259029956

References:

- Kai Hwang, Geoffrey C Fox, Jack G Dongarra, Distributed and Cloud Computing, From Parallel Processing to the Internet of Things, Morgan Kaufmann Publishers, 2012.

Links:

https://mrcet.com/pdf/Lab%20Manuals/IT/R15A0529_CloudComputing_Notes-converted.pdf
https://www.iare.ac.in/sites/default/files/lecture_notes/CC%20LECTURE%20NOTES.pdf

PGDS405B Cloud Computing

Course Outcomes: After successful completion of this course, students will be able to:

CO-1: Implement Cloud Service Models & Deployment Models

CO-PO Mapping Matrix

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	2	1	1	-	-	-	-	-	-	-

Sr. No	Practical on PGDS405B
1.	Working and Implementation of Infrastructure as a service.
2.	Working and Implementation of Software as a service.
3.	Working and Implementation of Platform as a services.
4.	Practical Implementation of Storage as a Service.
5.	Working of Google drive to make spreadsheet and notes.
6.	Working and Implementation of identity management.
7.	Write a program for web feed.
8.	Execute the step to Demonstrate and implementation of cloud on single sign on.

9.	Practical Implementation of cloud security.
10.	Installing and Developing Application Using Google App Engine.
11.	Implement VMWareESXi Server
12.	Using OpenNebula to manage heterogeneous distributed data center Infrastructure.
13.	Implementation of Cloud Failure Cluster.
14.	Managing and working of cloud xen server.
15.	Working with Aneka and demonstrate how to Managing cloud computing Resources .
16.	Installation and configuration of cloud Hadoop and demonstrate simple query.
17.	Create a sample mobile application using Amazon Web Service (AWS) account as a cloud service. Also provide database connectivity with implemented mobile application.