Bachelor of Science
(Honours)

Data Science & Analytics

Followed by

National Education Policy- 2020

AY: 2021- 22

Program and Course Structure

Department of Mathematics

School of Basic Sciences & Research

B.Sc. (H)
Data Science & Analytics

SBR0308

Batch 2021-24

1.1 Vision, Mission and Core Values of the University

Vision of the University

To serve the society by being a global University of higher learning in pursuit of academic excellence, innovation and nurturing entrepreneurship.

Mission of the University

1. Transformative educational experience
2. Enrichment by educational initiatives that encourage global outlook
3. Develop research, support disruptive innovations and accelerate entrepreneurship
4. Seeking beyond boundaries

Core Values

1. Integrity
2. Leadership
3. Diversity
4. Community

1.2 Vision and Mission of the School

Vision of the School

Achieving excellence in the realm of science to address the challenges of evolving society.

Mission of the School

1. Equip the students with knowledge and skills
2. Capacity building by providing academic flexibility to student and faculty members
3. To establish Centre of excellence for innovative research
4. Address the deficiencies of the society pertaining to environment
5. To strengthen academic- industry collaboration for better employability
6. Developing a culture for continued betterment in all facets of life

Core Values

1. Integrity
2. Leadership
3 Diversity
3. Community

1.3 Vision and Mission Department of Mathematics

Vision of the Department

To become a globally recognized destination for education in applied mathematics and research.

Mission of the Department

1. To develop mathematical skills in students and make them employable across a wide range of professions and promote interest research.
2. To develop entrepreneurial skills in students to serve the society at large.
3. To develop skills for the applications of mathematics in the various fields.

Core Values

1. Integrity
2. Leadership
3. Diversity
4. Community

1.4 Programme Educational Objectives (PEO's)

PEO1: Prepare professionals conversant with current and advanced technological tools to carry out Investigation, analysis and synthesis by identifying various compute oriented solutions.
PEO2: To develop positive attitude and skills which enable them to become a multi facet personality.
PEO3: To prepare students in such a way so that they perform excellently in national label entrance examinations conducted by various well known institution like IIT's/ central Universities/other academic institutes etc. to pursue their PG/MS/Dual PG and Ph. D. programs.
PEO4: To make them aware of effective machine learning and Artificial Intelligence based data analytics and inference required for Industrial Application.
PEO5: To inculcate passion for lifelong learning by introducing principles of group dynamics, public policies, environmental and societal context.

1.4.1Program Outcomes (PO's)

PO1: Data Science knowledge: Application of Data Science knowledge in various fields of science, engineering and management etc.
PO2: Nature of Data Science: Understand the concise, precise and rigorous nature of Data Science.
PO3: `Critical thinking: Develop the skill to think critically on abstract concepts of Data Science.
PO4: Problem analysis: Develop the ability to analyze a problem logically and dissect into micro-parts and thus resolving the problem to accessible components.
PO5: Presentation skill: Develop the skill to pleasant exposition for successful presentation for any career interview with confidence.
PO6: Data Science logic: Formulates and develops data analysis arguments in logical manner.
PO7: Team Work: Work as a team player and strive for self-excellence.
PO8: **Ethics:** Realize and understand professional, ethical and cultural responsibilities.
PO9: Communication: Communicate effectively with an elite audience.
PO10: Life-long learning: Engage in life-long learning towards enduring professional development.

Program Structure
Department of Mathematics
School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024
TERM: I

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular 5. Training/Survey /Project
	THEORY COURSES								
			L	**T**	**P**	**TOTAL (hrs.)**	**TOTAL (CREDITS)**		
1.	BDA 101	Statistics I	4	0	0	4	4	Co Requisite	Major 1
2.	MSM 101	Foundation course in Mathematics	4	0	0	4	4	Co Requisite	Major 2
3.	BDA 103	Fundamentals of Computers & Problem solving using C	4	0	0	4	4	Pre-Requisite	Major 3
4.	COC 101	Food, nutrition and hygiene	2	0	0	2	2	Co Requisite	Compulsory Co-curricular 1
	PRACTICAL COURSES								
5.	BDA 151	Statistics Lab - 1	0	0	4	4	2	Co Requisite	Major Lab 1
6.	MSM 151	Mathematics Lab - 1	0	0	4	4	2	Co Requisite	Major Lab 2
7.	BDA 152	C Programming Lab	0	0	4	4	2	Co Requisite	Major Lab 3
8.	BHM 152	Basic Statistical Analysis	0	0	5	5	3	Pre - Requisite	Vocational course 1
TOTAL							**23**		

Program Structure
Department of Mathematics

School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024
TERM: II

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular, 5. Training/Survey /Project
	THEORY COURSES								
			L	**T**	**P**	**TOTAL (hrs.)**	**TOTAL (CREDITS)**		
1.	BDA 105	Statistics II	4	0	0	4	4	Co Requisite	Major 4
2.	BHM 103	Linear Algebra & Discrete Mathematics	6	0	0	6	6	Co Requisite	Major 5
3.	BDA 104	Programming R	4	0	0	4	4	Pre-Requisite	Major 6
4.	OPE	Elective (Other dept./ other School)	4	0	0	4	4	Pre-Requisite	Minor 1
5.	COC 201	First Aid and Health	2	0	0	2	2	Pre-Requisite	Compulsory Co-curricular 2
	PRACTICAL COURSES								
6.	BDA 153	Statistics Lab - 2	0	0	4	2	2	Co Requisite	Major Lab 4
7.	BDA154	R Programming Lab	0	0	4	2	2	Pre- Requisite	Major Lab 5
8.	BHM 153	Statistical thinking using Analysis Software	0	0	5	3	3	Pre- Requisite	Vocational course 2
TOTAL							**27**		

Program Structure
Department of Mathematics
School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024
TERM: III

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular 5. Training/ Survey /Project
	THEORY COURSES								
			L	**T**	**P**	**TOTAL Hrs.**	**TOTAL (CREDITS)**		
1.	BDA 212	Statistics III	4	0	0	4	4	Co-requisite	Major 7
2.	BHM 308	Numerical Analysis & Operations Research	4	0	0	4	4	Co-requisite	Major 8
3.	BDA 213	Data Structure & Theory of Algorithms	4	0	0	4	4	Co-requisite	Major 9
4.	COC 301	Human Values and Environment studies	2	0	0	2	2	Co-requisite	Compulsory Co-curricular 3
	PRACTICAL COURSES								
5.	BDA 251	Statistics Lab - 3	0	0	4	4	2	Co Requisite	Major Lab 6
6.	BHM 354	Mathematics Lab - 6	0	0	4	4	2	Pre- Requisite	Major Lab 7
7.	BDA 252	Data Structure Lab	0	0	4	4	2	Co Requisite	Vocational course 3
8.	BHM 252	Prediction and forecasting management	0	0	5	5	3	Co Requisite	Major Lab 6
TOTAL							**23**		

Program Structure
Department of Mathematics
School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024
TERM: IV

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular 5. Training/ Survey /Project
	THEORY COURSES								
			L	**T**	**P**	**TOTAL (hrs.)**	**TOTAL (CREDITS)**		
1.	BDA 201	Data preparation and Data Cleaning	4	0	0	4	4	Co Requisite	Major 10
2.	BDA 202	Database Management Systems	4	0	0	4	4	Co Requisite	Major 11
3.	BDA 204	Operating System	4	0	0	4	4	Pre-Requisite	Major 12
4.	OPE	Elective (Other dept./ other School)	4	0	0	4	4	Pre-Requisite	Minor 2
5.	COC 401	Physical Education and Yoga	2	0	0	2	2	Co Requisite	Compulsory Co-curricular 4
	PRACTICAL COURSES								
	BDA 253	Data Cleaning Lab	0	0	4	4	2	Co Requisite	Major Lab 8
	BDA 254	DBMS Lab	0	0	4	4	2	Pre-Requisite	Major Lab 9
	BDA 255	Operating System Lab	0	0	4	4	2	Co Requisite	Vocational course 4
	BHM 254	Advance Statistical Analysis	0	0	5	5	3	Co Requisite	Major Lab 8
TOTAL							**27**		

Program Structure
Department of Mathematics
School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024

TERM: V

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular 5. Training/ Survey /Project
	THEORY COURSE								
			L	**T**	**P**	**TOTAL (hrs.)**	**TOTAL (CREDITS)**		
1.	BDA 312	Data Ware housing and Data mining	5	0	0	5	5	Co Requisite	Major 13
2.	BDA 313	Regression, time series, forecasting and Index numbers	5	0	0	5	5	Co Requisite	Major 14
3.	BDA 314	Machine learning Deep learning	5	0	0	5	5	Co Requisite	Major 15
4.	BDA 315	Multivariate Analysis	5	0	0	5	5	Co Requisite	Major 16
5.	COC 501	Analytic Ability and Digital Awareness	2	0	0	2	2	Co Requisite	Compulsory Co-curricular 5
	PRACTICAL COURSES								
6.	BHM 352	**Part- I:** Community connect (2) + Summer internship/Industrial Training/ Survey/Project (1) (Will be completed after 4th Semester)	0	0	6	6	3	Co Requisite	Summer internship/ Industrial Training/ Survey/Project
TOTAL							**25**		

Program Structure
Department of Mathematics
School of Basic Sciences & Research
B. Sc. (H) Data Science & Analytics
Batch: 2021-2024

TERM: VI

S. No.	Course Code	Course Name	Teaching Load				CREDITS	PRE-REQUISITE/ CO-REQUISITE	Type of Course: 1. Major,2. Minor/ Elective, 3.Vocational 4. Compulsory Co-curricular 5. Training/ Survey /Project
	THEORY COURSES								
			L	**T**	**P**	**TOTAL (hrs.)**	**TOTAL (CREDITS)**		
1.	BDA 316	Statistical Analysis & simulation	4	0	0	4	4	Co Requisite	Major 17
2.	BDA 302	Data Scientist Toolbox	4	0	0	4	4	Co Requisite	Major 18
3.	BDA 317	Statistical Inference	4	0	0	4	4	Co Requisite	Major 19
4.	BDA 318	Data Visualization	4	0	0	4	4	Co Requisite	Major 20
5.	COC 601	Communication Skills and Personality Development	2	0	0	2	2	Pre- Requisite	Compulsory Co-curricular 6
	PRACTICAL COURSES								
6.	BDA 351	Statistical Simulation Lab	0	0	4	4	2	Co Requisite	Major Lab 10
7.	BDA 352	Data Visualization Lab	0	0	4	4	2	Co Requisite	Major Lab 11
8.	BHM 355	**Part-II:** Community connect (2) + Summer internship/Industrial Training/ Survey/Project (1) (Will be completed after 4th Semester)	0	0	6	6	3	Co Requisite	Summer internship/ Industrial Training/ Survey/Project
TOTAL							**25**		

Year- wise Program structure
B.Sc. (H) Data Science & Analytics
Department of Mathematics
School of Basic Sciences & Research, Sharda University

		Subject 1	Subject 2	Subject 3	Subject 4	Vocational	Compulsory Co-curricular	Training/ Survey/ Project	Credits	(Total Credits) After completion {Minimum Credits} [Max Duration in years]
		Major	Major	Major	Minor/ Elective	Minor	Minor	Major		
		Credits 4/5/6	Credits 4/5/6	Credits 4/5/6	Credits 4	3 Credits	2 Credits	Credits 3/6/8		
Year	Sem.	Own Faculty	Own Faculty	Any Faculty	Othe Departm ent/ Faculty	Vocation al Faculty	Co Curricular Course	Related to main Subject		
1	I	Statistics I Credit: 6 (1 Th-4 +1 P- 2)	Foundation course in Mathematics Credit: 6 (1 Th-4 +1 P- 2)	Fundamentals of Computers & Problem solving using C Credit: 6 (1 Th-4 +1 P- 2)		Basic Statistical Analysis	Health, nutrition and hygiene		23 (18+3+2)	(50-52) {46} [4] Certificate in Faculty
	II	Statistics II Credit: 6 (1 Th-4 +1 P- 2)	Linear Algebra & Discrete Mathematics Credit: 6 (1 Th-6)	Programming R Credit: 6 (1 Th-4 +1 P- 2)	Elective	Statistical thinking using Analysis Softwares	First Aid and Health		27 (18+3+2) +4	

2	III	Statistics III Credit: 6 (1 Th-4 +1 P- 2)	Numerical Analysis & Operationas Research Credit: 6 (1 Th-4 +1 P- 2)	Data Structure & Theory of Algorithms / Others Credit: 6 (1 Th-4 +1 P- 2)		Prediction and forecasting management	Human Values and Environment studies		23 (18+3+2)	(100-104) {92} [7] Diploma in Faculty
	IV	Data preparation and Data Cleaning Credit: 6 (1 Th-4 +1 P- 2)	Database Management Systems Credit: 6 (1 Th-4 +1 P- 2)	Operating Systems	Elective	Advance Statistical Analysis	Physical Education and Yoga		27 (18+3+2) +4	
3	V	Data Ware housing and Data mining + Regression, time series, forecasting and Index numbers OR any course from elective courses Credit: 10 (2Th- 5)	(Machine learning & Deep learning)+ Multivariate Analysis OR any course from elective courses Credit: 10 (2Th- 5)				Analytic Ability and Digital Awareness	**PART 1** Practical hands on Industrial Training/ Survey/Project. 1 (3)	25 (20+3+2)	(150-154) {138} [10] Bachelor in Faculty
	VI	(Statistical Analysis (Count Data and survival Analysis) & statistical simulation) + Data Scientist Toolbox OR any course from elective courses Credit: 10 (2Th- 4+ 1P- 2)	Statistical Inference (non- parametric) + Data Visualization OR any course from elective courses Credit: 10 (2Th- 4+ 1P- 2)				Communicati on Skills and Personality Development	**PART 2** Practical hands on Industrial Training/ Survey/Project. 1 (3)	25 (20+3+2)	
Total Credit: 150 **National Education Policy- 2020** **(w. e. f. Academic Session 2021- 22)**										

Note: Students can opt of the following courses as elective in V & VI Semester:

1. Big Data Analytics 2. Digital Marketing, 3. Business Intelligence, 4. IOT 5. Econometrics 6. Artificial Intelligence, 7. Epidemiological methods 8. Statistics in Forensic Analytics 9. Statistics in Agriculture 10. Health Statistics

Statistics I (BDA 101)

School: SBSR		**Batch: 2021- 2024**
Program: B. Sc. (H)		**Current Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: I**
1	Course Code.	BDA101
2	Course Title	STATISTICS I
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course status	Compulsory
5	Course Objectives	1. To introduce basic statistical concepts, logics and analytical tools, analyze and communicate quantitative data verbally, graphically, symbolically and numerically. 2. To make students familiar with the concept of Probability and Statistics and display data by means of various tables, charts, and graphs.
6	Course Outcomes	CO1: Describe the process and particular steps in designing studies, collecting and analyzing data, interpreting and presenting results; and develop skills in presenting quantitative data using appropriate diagrams, tabulations and summaries. (K2, K5) CO2: Describe the properties of discrete and continuous distribution functions. (K2) CO3: Calculate the measures of central tendency and dispersion of a data and describe the method used for analysis, including a discussion of advantages, disadvantages, and necessary assumptions. (K2, K3) CO4: Calculate and interpret the correlation between two variables and Calculate the simple linear regression equation for a set of data and know the basic assumptions behind regression analysis. (K2, K3) CO5: Understand the line of best fit as a tool for summarizing a linear relationship and predicting future observed values, develop the ability to use formal mathematical argument in the context of probability. (K2, K5) CO6: Develop the skills to interpret the results of statistical analysis. (K2, K5).
7	Course Description	This is an introductory course in statistics. Students are introduced to the fundamental concepts involved in using sample data to make inferences about populations. Included are the study of measures of central tendency and dispersion, finite probability, statistical inferences from large and small samples, linear regression, and correlation.

8	**Outline syllabus:**
UNIT 1	**Presentation of data**
A	Classification, tabulation, diagrammatic & graphical representation of grouped data.
B	Frequency distributions, cumulative frequency distributions
C	Histogram, Ogives, frequency polygon, Tree and leaf diagram.
UNIT 2	**Descriptive statistics**
A	Measures of central tendency – arithmetic mean, median, quartiles, mode, harmonic mean, geometric mean.
B	Their properties, merits and demerits
C	Measures of dispersion – range, quartile deviation, mean deviation, standard deviation and coefficient of variation.
UNIT 3	**Moments**
A	Moments, Skewness, Measures of skewness: Karl Pearson's coefficient of skewness.
B	Quartile coefficient of skewness, Measure of skewness based on moments.
C	Kurtosis, measure of Kurtosis.
UNIT 4	**Bi-variate data analysis**
A	Bivariate data, principles of least squares, fitting of polynomial curves and fitting of curves reducible to polynomial form.
B	Correlation: Spearman's rank correlation, Partial and Multiple Correlation (only two independent variables case).
C	Regression lines.
UNIT 5	**Probability**
A	Probability: Introduction, random experiment, outcomes, sample space, events, various definitions of probability, laws of total and compound probability. Boole's inequality. Conditional probability, independence of events. Bayes theorem and its applications.
B	Random variables: discrete and continuous random variables, probability mass function (p.m.f), probability density function (p.d.f) and cumulative distribution function (c.d.f), illustrations and properties of random variables, univariate transformations with illustrations.
C	Mathematical Expectation: Expectation of single and bivariate random variables, properties of expectation, conditional expectation and its properties. Moments and cumulants. Moment generating function, probability generating function.
Mode of Examination	Theory Max. Marks: 25+ 25+ 50

Weightage distribution	CA	MTE	MTE
	25 Marks	25 Marks	50 Marks

Text books	1. 1. Gupta,S.C and Kapoor,V.K, "Fundamental of Mathematical Statistics".
Other references	1. Daniel,WayneW.,"Biostatistics": Basic concept and Methodology for Health Science. 2. Grewal,B.S, "Higher Engineering Mathematics".

Foundation Course in Mathematics (MSM 101)

School: SBSR	**Batch : 2021- 2024**
Program: B.Sc. (H)	**Academic Year: 2021-22**
Branch: B. Sc. Data Science & Analytics	**Semester: I**
Course Code	MSM 101
Course Title	Foundation Course in Mathematics
Credits	4
Contact Hours (L-T-P)	4-0-0
Course Status	
Course Objective	1. To familiarise the students with basic concepts of matrices, determinants and solving the system of linear equations. 2. To understand the basic concept of sets theory, co-ordinate geometry, complex number and vector algebra.
Course Outcomes	CO1: Explain the concept of matrices and solve systems of linear equations and determinants. (K2,K3, K4) CO2: Explain the concept of complex numbers and calculate the nth roots of complex numbers and illustrate the solutions of simple Polynomial equations. (K2, K3, K4) CO3: Memorize the basic of Cartesian coordinate system and use algebraic techniques to explain intercepts and explore equations of lines on the number plane. (K1, K3, K4) CO4: Describe and differentiate the symmetries from graphs of conic sections. (K1, K2) CO5: Describe and use the concepts of set theory, relation and functions. (K1,K2,K3) CO6: Explain the basic concepts of vector algebra and use to find area of parallelogram and quadrilateral, Vector triple product.(K2,K3,K4)
Course Description	This course is an introduction to the fundamental of Mathematics. The primary objective of the course is to develop the basic understanding of linear algebra, complex number, co-ordinate geometry, sets theory and vector algebra.
Outline syllabus	
Unit 1	Evaluation of determinants, Properties of determinants,
A	Matrices: types of matrices, addition, subtraction and multiplication of matrices, symmetric and skew symmetric matrix. Inverse of matrix.
B	Rank of a matrix, Consistency of system of equations, Characteristic equation, Cayley -Hamilton theorem.
C	**Complex Numbers**
Unit 2	Representation of complex number in Argand plane, Modulus and argument of complex number
A	Algebraic operations, De- Moivre's theorem
B	Nth root of complex number, Euler's formula
C	**Co-ordinate geometry**
Unit 3	Cartesian coordinate system, Distance between two points Equations of line in various forms
A	Equation of circle in various forms, Equation of tangent and normal to the circle.
B	Equation of ellipse, parabola and hyperbola
C	**Sets Theory**
Unit 4	Definition of set, types of sets, Union and intersection of sets, Venn diagram, De-Morgan's law.
A	Relation and functions.
B	Composite function and inverse function.

C	**Vector Algebra**		
Unit 5	Addition and subtraction of vectors and their geometric application.		
A	Scalar and vector product, their physical application, Projection of vector on another vector, area of triangle.		
B	Area of parallelogram and quadrilateral, Vector triple product.		
C	Evaluation of determinants, Properties of determinants,		
Mode of examination	Theory		
Weightage Distribution	CA	MTE	ETE
	25 marks	25	75 marks
Text book/s*	1. Kreyszig, E., "Advanced Engineering Mathematics", John Wiley & Sons Inc. Jain, M.K., and Iyengar, S.R.K., "Advanced Engineering Mathematics", Narosa Publications		
Other References	1. Thomas, B.G., and Finny R.L., "Calculus and Analytical geometry", Pearson Education Asia, AdisonWisley. Simmons, G.F., "Differential Equations with applications with applications", Tata McGraw-Hill.		

Fundamentals of Computers & Problem-Solving using C (BDA 103)

School: SBSR	**Batch : 2021- 2024**
Program: B.Sc. (H)	**Academic Year: 2021-22**
Branch: B. Sc. Data Science & Analytics	**Semester: I**
Course Code	BDA 103
Course Title	Fundamentals of Computers & Problem-Solving using C
Credits	4
Contact Hours (L-T-P)	4-0-0
Course Status	
Course Objective	To understand and demonstrate how to solve logical and scientific problems using programming.
Course Outcomes	CO1: Explain the concept of key components of a computer system (K2,K3, K4). CO2: Apply and practice logical ability to solve the problems. (K2, K3, K4). CO3: Describe how to generate efficient and schematic solution to the problems. (K1, K2). CO4: **Demonstrate** the algorithm, Pseudo-code and flow chart for the given problem (K2, K3,K4). CO5: **Create** and implement logic using Operators and control statements. CO6: **Develop** better understanding of basic concepts of C programming and Computer Organization.
Course Description	To understand and demonstrate how to solve logical and scientific problems using programming.
Outline syllabus	
Unit 1	Computer Fundamentals
A	Introduction to Computers: Characteristics of Computers, Uses of computers, Types and generations of Computers. Overview of basic digital building blocks
B	Number representation, Characters-ASCII coding, other coding schemes, IEEE754, Assembly language programming for some processor
C	Basic building blocks for the ALU, Registers, CPU buses; Control path microprogramming (only the idea), hardwired, logic.
Unit 2	Basic Computer Organization
A	External interface, Memory Subblock, Memory organization; Technology-ROM, RAM, EPROM, Flash etc. Cache
B	Cache coherence protocol for uniprocessor (simple), I/O Subblock, I/O techniques -interrupts, polling, DMA; Synchronous vs. Asynchronous I/O, Controllers
C	Peripherals, Disk drives; Printers- impact, dot matrix, ink jet, laser; Plotters; Keyboards; Monitors.
Unit 3	Techniques of Problem Solving
A	Structured programming concepts, Programming methodologies viz. top-down and bottom-up programming.
B	Introduction to Programming, Flowchart, Pseudo code with examples, From algorithms to programs
C	Structure of a C program, Datatypes, Variables, Constants, Identifiers and Keywords, Different Operators, operator precedence and associativity,
Unit 4	C Language-I
A	Storage classes, Logical Errors in compilation, object and executable code
B	Types of Statements: Assignment, Control, jumping,
C	Control statements: Decisions, Loops, break, continue, Nested Loop
Unit 5	C Language-II
A	Functions, passing values between functions, scope rules of functions.
B	Structures and Unions
C	Arrays: 1D Array, 2D array
Mode of	Theory

examination			
Weightage Distribution	CA	MTE	ETE
	25 marks	25 Marks	50 Marks
Text book/s*	Yashavant Kanetkar, “Let Us C”, BPB.		
Other References	1. Byron Gottfried, “Programming with C”,TMH. 2.R. G. Dromey, “How to Solve It by Computer”,Pearson.		

SEMESTER - II

Statistics II (BDA 105)

School: SBSR		**Batch : 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2021-22**
Branch: **Data Science & Analytics**		**Semester: II**
1	Course Code	BDA 105
2	**Course Title**	**Statistics II**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	**Compulsory**
5	Course Objective	To make students familiar with the concept of sample and population, complete enumeration versus sampling. The concept of Systematic Sampling, estimates of population mean and total, variances of these estimates along with the brief of present official statistical system in India, methods of collection of official statistics, their reliability and limitations has been introduced.
6	Course Outcomes	CO1: Explain and illustrate the concepts of sample and population. (K2, K3, K4) CO2: Describe the properties of complete enumeration versus sampling; explain random sampling with and without replacement. (K1, K2, K3) CO3: Describe estimates of population mean, explain its application and estimates of theses variances and sample size determination. (K2, K3, K4) CO4: Describe stratified random sampling, estimates of population mean and total and explain its application; and illustrate systematic sampling. (K2, K3, K4) CO5: Describe the ratio and regression methods of estimation and evaluate variances in terms of correlation coefficient between X and Y for regression method and their comparison with SRS. (K2, K3, K6) CO6: Describe and analyze the basic concepts present official statistical system in India, methods of collection of official statistics. (K1,K2, K4)
7	Course Description	This course is an initiate the advance concept of sample and population, complete enumeration versus sampling. The concept of Systematic Sampling, estimates of population mean and total, variances of these estimates along with the brief of present official statistical system in India, methods of collection of official statistics, their reliability and limitations has been introduced. .
8	Outline syllabus : **Statistics -II**	
	Unit 1	
	A	Concept of sample and population, complete enumeration versus sampling

	B	Sampling and non-sampling errors, requirements of a good sample,		
	C	Simple random sampling with and without replacement.		
	Unit 2			
	A	Estimates of population mean, total and proportion,		
	B	Variances of these estimates		
	C	Estimates of theses variances and sample size determination.		
	Unit 3			
	A	Stratified random sampling, estimates of population mean and total variances of these estimates.		
	B	Proportional and optimum allocations and their comparison with SRS.		
	C	Systematic Sampling, estimates of population mean and total, variances of these estimates.		
	Unit 4			
	A	Ratio and regression methods of estimation, estimates of population mean and total (for SRS of large size),		
	B	Variances of these estimates and estimates of theses variances,		
	C	Variances in terms of correlation coefficient between X and Y for regression method and their comparison with SRS.		
	Unit 5			
	A	Present official statistical system in India, Methods of collection of official statistics, their reliability and limitations.		
	B	Principal publications containing data on the topics such as population, industry and finance.		
	C	Various official agencies responsible for data collection and their main functions.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25 marks	25 marks	50 marks
	Text book/s*	1. Goon A.M., Gupta M.K. and Dasgupta B (2001): Fundamentals of Statistics (Vol.2), Word Press. 2. Murthy M.N. (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta 3. Des Raj and Chandhok P.(1998): Sample Survey Theory, Narosa Publishing House. Cochran W.G (1984):Sampling Techniques (3rd Ed.), Wiley Eastern.		

	Other References	1. Mukhopadhyay P.(1998): Theory and Methods of Survey Sampling, Prenctice Hall 2. Sampat S.(2001): Sampling Theory and Methods, Narosa Publishing House 3. Guide to current Indian Official Statistics, Central Statistical Organization , GOI,New Delhi. Saluja,M.P. (1972): Indian official statistical systems, Statistical Pub. Society, Calcutta.

LINEAR ALGEBRA & DISCRETE MATHEMATICS (BHM 103)

School: SBSR		**Batch: 2021- 2024**
Program: B. Sc.(H)		**Academic Year: 2021-22**
Branch: Data Science &Analytics		**Semester: II**
1	Course Code.	BHM 103
2	**Course Title**	LINEAR ALGEBRA & DISCRETE MATHEMATICS
3	Credits	6
4	Contact Hours (L-T-P)	6-0-0
	Course status	**Compulsory**
5	Course Objectives	To familiarise students with basics algebra of matrices, and its applications, vector space, Linear transformation and its properties, matrix representation of a linear transformation.
	Course Outcomes	CO1: Describe the concept of algebra of matrices and elementary row operations and calculate the rank of matrix and analyse consistency of a linear system. (K1, K2, K3, K4) CO2: Calculate the eigenvalues, eigenvectors, diagonalization of a matrix. (K2, K3) CO3: Explain and illustrate Cayley - Hamilton theorem and its applications. (K2,K3, K4). CO4: Discuss vector space and subspace, explain linear dependence and independence of vectors and calculate linear span, basis and dimension, sums and direct sums(K2, K3, K4) CO5: Discuss the concept of sets, un-countably infinite sets, principle of inclusion and exclusion, multisets, propositions, conditional propositions and evaluate normal forms, Mathematical induction.(K2,K3, K4,K5) CO6: Describe the concept functions, composition of function, invertible functions, discrete properties of binary relations and check the closure of relations. (K3, K6)
7	Course Description	This course is an introduce basics algebra of matrices, and its applications, vector space, Linear transformation and its properties, matrix representation of a linear transformation.
8	Outline syllabus	Linear Algebra
	Unit 1	**Algebra of matrices-1**
	A	Algebra of matrices, elementary row operations
	B	Row reduced Echelon form, rank of a matrix
	C	Consistency of a linear system, inverse of a matrix (using elementary row operations.
	Unit 2	**Algebra of matrices-2**
	A	Eigenvalues and eigenvectors
	B	Diagonalization of a matrix
	C	Cayley - Hamilton theorem (without proof) and its applications

	UNIT 3	**Vector Spaces**		
	A	Vector space and subspace of vector space.		
	B	Linear dependence and independence of vectors, linear span.		
	C	Basis and dimension, sums and direct sums.		
	Discrete Mathematics			
	Unit 4	**Sets and Propositions -**		
	A	Sets, Un-countably infinite sets, Principle of inclusion and exclusion, multisets, propositions,		
	B	Conditional propositions. Logical connectivity, Propositional, calculus,		
	C	Universal and existential quantifiers, Normal forms, methods of proofs, Mathematical induction.		
	Unit 5	**Relations and Functions -**		
	A	Functions , Composition of function , invertible functions, Discrete properties of binary relations, closure of relations		
	B	Warshall's algorithm, Equivalence relations and partitions, Ordered Sets and Lattices: Introduction, Ordered set,		
	C	Hasse diagram of partially ordered set, Consistent enumeration, Isomorphic ordered set, Well ordered set, Lattices, Properties of lattices, Bounded lattices, Distributive lattices, and Complemented lattices. Chains, and Anti-chains.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25 marks	25 marks	50 marks
	Text book/s*	1. Hoffman, K &Kunze, R. , Linear Algebra, 2nd edition, Prentice Hall of India, 1975. 2.Lipshutz, S., Lipsom, M., Linear algebra, 3rd edition, Schaum series, 2001. 3. Liu C.L. and Mohapatra, D.P., " Elements of Discrete Mathematics" , SiE edition, TMH, 2008		
	Other References	1. Strang, G., Linear Algebra and its applications, 3rd edition, Thomson,1998. 2. Kreyszig , E., Advanced Engineering Mathematics, John Wiley & Sons. 3. V. Krishnamurthy, V.P. Mainra and J.L. Arora: An Introduction to Linear Algebra. 3. Kenneth H.R.,' Discrete Mathematics and its Applications", Mc-graw hill. 5. Biggs N., "Discrete Mathematics", 3rd edition, Oxford University		

Programming R (BDA 104)

School: SBSR		**Batch: 2021-2024**
Program: B. Sc.(H)		**Academic Year: 2021-22**
Branch: Data Science &Analytics		**Semester: II**
1	Course Code.	BDA 104
2	**Course Title**	**Programming R**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course status	Compulsory
5	Course Objectives	To familiarise students with basics programming in R, and its applications in data analysis.
	Course Outcomes	CO1: Explain the R Windows Environment and describe various data types. (K1, K2, K3, K4). CO2: Explain and describe Outliers, Combining Datasets. (K2, K3) CO3: Explain and illustrate R Functions and loops, Summary Statistics –Summarizing data with R. (K2,K3, K4). CO4: Discuss how to load data, plot a graph and illustrate different types of graphs with graphical summaries of data. (K2, K3, K4) CO5: Discuss how to generate automated reports giving detailed basic statistics using R and evaluate measures of central tendency and dispersion. Covariance, correlation and lines of regression in R.(K2, K3, K4) CO6: Explain fitting of polynomials and exponential curves and illustrate Normal probability plot. (K 4, K6)
7	Course Description	This course is an introduce basics programming in R, and its applications in data analysis.
8	**Outline syllabus: Programming R**	
	Unit 1	
	A	Introduction to R, R-Studio (GUI): R Windows Environment, introduction to various data types,
	B	Numeric, Character, date, data frame, array, matrix etc.,
	C	Reading Datasets, Working with different file types .txt, .csv etc.
	Unit 2	
	A	Outliers, Combining Datasets,
	B	R Functions and loops,
	C	Summary Statistics –Summarizing data with R.
	UNIT 3	
	A	Vector space and subspace of vector space.
	B	Linear dependence and independence of vectors, linear span.
	C	Basis and dimension, sums and direct sums.
	Unit 4	
	A	Learn how to load data, plot a graph viz.
	B	histograms (equal class intervals and unequal class intervals), box plot, stem-leaf, frequency polygon, pie chart, ogives with graphical summaries of data,
	C	customization of plot setting, adding text, saving to a file, adding a legend.
	Unit 5	
	A	Random number generation and sampling procedures.
	B	Fitting of polynomials and exponential curves.

	C	Application Problems based on fitting of suitable distribution, Normal probability plot.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Gardener, M (2012): Beginning R: The Statistical Programming Language, Wiley Publications. 2. Braun W J, Murdoch D J (2007): A First Course in Statistical Programming with R. Cambridge University Press. New York		
	Other References	1. Crawley, M.J. (2015): Statistics: An Introduction Using R, 2nd Edition. Wiley. 2. Crawley, M.J. (2012): The R Book, 2nd Edition. Wiley.		

SECOND YEAR THEORY COURSES

Statistics III (BDA 212)

School: SBSR		**Batch: 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2022-23**
Branch: Data Science& Analytics		**Semester: III**
1	Course Code.	BDA 212
2	Course Title	STATISTICS III
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course status	**Compulsory**
5	Course Objectives	To introduce concepts of statistical analysis of descriptive statistics, logics and analytical tools, analyze and communicate quantitative data verbally, graphically, symbolically and numerically. To make students familiar with the concept of Probability and Statistics and hypothesis.
6	Course Outcomes	CO1: Describe the process Statistical analysis of descriptive statistics, principle of least square, lines of regression, simple linear regression and evaluate multiple linear regression, coefficient of multiple determination. (K2, K5) CO2: Describe the process of fitting of polynomials and exponential curves. (K2) CO3: Explain the criteria for obtaining a good estimator . (K2, K3) CO4: Calculate and interpret the point estimation, confidence interval, construction of confidence intervals using pivotal, shortest expected length confidence interval. (K2, K3) CO5: Understand the null hypothesis, alternative hypothesis, type I error, type II error, level of significance, p-value and power of test, develop the ability to use one sample t-test, two-sample t-test, paired-sample t-test. Tests for variance based on normal distribution – one sample and two-sample problem. (K2, K5) CO6: Develop the skills to interpret the results of statistical analysis by using Z-test, F-test, Chi-square test for goodness of fit. One-way and Two-way analysis of variance (ANOVA) techniques. (K2, K5)
7	Course Description	This is an advances course in statistics. Students are introduced to the f concepts involved in using sample data to make inferences about populations. Included are the study of measures of central tendency and dispersion, finite probability, statistical inferences from large and small samples, linear regression, and correlation and hypothesis.
8	Outline syllabus:	

UNIT 1				
A	Statistical analysis of descriptive statistics, principle of least square, lines of regression, simple linear regression			
B	coefficient of determination. Multiple linear regression, coefficient of multiple determination.			
C	Fitting of polynomials and exponential curves.			
UNIT 2				
A	Criteria for obtaining a good estimator: unbiasedness, consistency, efficiency, sufficiency.			
B	Minimal sufficient statistic.			
C	Uniformly minimum variance unbiased estimator, complete statistic.			
UNIT 3				
A	Method of point estimation: Method of moments, maximum likelihood estimator and its properties, mean square error (MSE).			
B	Interval estimation: Confidence interval, construction of confidence intervals using pivotal			
C	Shortest expected length confidence interval.			
UNIT 4				
A	Null hypothesis, alternative hypothesis, type I error, type II error, level of significance, p-value and power of test.			
B	Tests for mean based on normal distribution – one sample t-test, two-sample t-test, paired-sample t-test.			
C	Tests for variance based on normal distribution – one sample and two-sample problem			
UNIT 5				
A	The large sample size test: Z-test, F-test,			
B	Chi-square test for goodness of fit.			
C	One-way and Two-way analysis of variance (ANOVA) techniques.			
	Mode of Examination	Theory		
	Weightage distribution	CA	MTE	ETE
		25	25	50
	Text books	1. Gupta,S.C and Kapoor,V.K, "Fundamental of Mathematical Statistics".		
	Other references	2.Daniel,WayneW.,"Biostatistics": Basic concept and Methodology for Health Science. 3. Grewal,B.S, "Higher Engineering Mathematics".		

Numerical Analysis & Operations Research (BHM 308)

School: SBSR		**Batch : 2021- 2024**
Program: B.Sc.(H)		**Academic Year: 2022-23**
Branch: Mathematics		**Semester: III**
1	Course Code	BHM 308
2	Course Title	Numerical Analysis & Operations Research
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	1. To provide the student with numerical methods of solving the non-linear equations, interpolation, differentiation, and integration. 2.To improve the student's skills in numerical methods by using the MATLAB.3. To provide the students are able to formulate a real-world problem as a mathematical programming model, understand the theoretical workings of the simplex method for linear programming and perform iterations of it by hand, relationship between a linear program and its dual, including strong duality and complementary slackness and solve specialized linear programming problems like the transportation and assignment problems.
6	Course Outcomes	**CO1:** The aim of this course is to teach the student the application of various numerical technique for variety of problems occurring in daily life. At the end of the course the student will be able to understand the basic concept of Numerical Analysis and to solve algebraic and differential equation. (K1,K2,K3.) **CO2:** The main outcome will be that students will be able to handle problems and finding approximated solution. Later he can opt for advance course in Numerical Analysis in higher Mathematics. (K2,K3). **CO3:** The student will be able to solve various problems based on convex sets and linear programming. After successful completion of this paper will enable the students to apply the basic concepts of transportation

		problems and its related problems to apply in further concepts and application of operations research. (K2,K3,K4) **CO4:** Identify and develop operational research models from the verbal description of the real system. (K3,K4) **CO5:** Discuss the concept of duality and formulate and solve Dual of LPP. (K3,K4,K5) **CO6:** Describe the numerical differentiation and evaluate the differentiation. (K4,K5,K6)		
7	Course Description	This course is an introduction to the numerical analysis. The primary objective of the course is to develop the basic understanding of numerical algorithms and skills to implement algorithms to solve mathematical problems in MATLAB. Operations research (OR) have many applications in science, engineering, economics, and industry and thus the ability to solve OR problems are crucial for both researchers and practitioners. Being able to solve the real life problems and obtaining the right solution requires understanding and modelling the problem correctly and applying appropriate optimization tools and skills to solve the mathematical model. The goal of this course is to teach students to formulate, analyze, and solve mathematical models that represent real-world problems.. In particular, we will cover linear programming.		
8	Outline syllabus			
	Unit 1			
	A	Solution of equations: bisection, Secant, Regular Falsi, Newton Raphson's method, Newton's method for multiple roots.		
	B	Interpolation, Lagrange and Hermite interpolation.		
	C	Difference schemes, Divided differences, Interpolation formula using differences.		
	Unit 2			
	A	Numerical differentiation, Numerical Quadrature: Newton Cotes Formulas. Gaussian Quadrature Formulas.		
	B	System of Linear equations: Direct method for solving systems of linear equations (Gauss elimination, LU Decomposition, Cholesky Decomposition).		
	C	Iterative methods (Jacobi, Gauss Seidel, Relaxation methods). The Algebraic Eigen value problem: Jacobi's method, Givens method, Power method.		
	Unit 3			
	A	Numerical solution of Ordinary differential equations: Euler method, single step methods.		
	B	Runge-Kutta method, Multi-step methods: Milne-Simpson method, Types of approximation: Last Square polynomial approximation, Uniform approximation, Chebyshev polynomial approximation.		
	C	Difference Equations and their solutions, Shooting method and Difference equation method for solving Linear second order differential equation with boundary conditions of first, second and third type.		
	Unit 4			
	A	Introduction, Linear programming problems, statement and formation of general linear programming problems, graphical method, slack and surplus variables, standard and matrix forms of linear programming problem, basic feasible solution.		
	B	Convex sets, fundamental theorem of linear programming, basic solution, Simplex method.		
	C	Introduction to artificial variables, two phase method Big-M method and their comparison.		
	Unit 5			
	A	Resolution of degeneracy, duality in linear programming problems.		
	B	Primal dual relationships, revised simplex method, sensitivity analysis.		
	C	Transportation problems, assignment problems.		
	Mode of examination	Theory/Jury/Practical/Viva		
	Weightage Distribution	CA	MTE	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	**Suggested Readings(Part-A Numerical Analysis):**		

		1. Numerical Methods for Engineering and scientific computation by M. K. Jain, S.R.K. Iyengar & R.K. Jain. 2. An Introduction to Numerical Analysis by EndreSuli, David F. Mayers, Cambridge University Press, 2003. 3. Applied Numerical Analysis by C. F. Gerald, Pearson Education, 2009. 4. Elements of Numerical Analysis by R. S. Gupta, Macmillan India Ltd, 2009. **Suggested Readings(Part-B Operation Research):** **1.**Taha, Hamdy H, "Opearations Research- An Introduction ", Pearson Education. **2.**Kanti Swarup , P. K. Gupta , Man Mohan Operations research, Sultan Chand & Sons **3.**Hillier Frederick S and Lieberman Gerald J., "Operations Research", McGraw Hill Publication.
	Other References	1. Introductory methods of Numerical Analysis by S. S. Sastry 2. Suggested digital plateform:NPTEL/SWAYAM/MOOCs 3.Winston Wayne L., "Operations Research: Applications and Algorithms", Cengage Learning, 4 th Edition. **4.**Hira D.S. and Gupta Prem Kumar, "Problems in Operations Research: Principles and Solutions", S Chand & Co Ltd. **5.** Kalavathy S., "Operations Research", S Chand.

Data Structure & Theory of Algorithms (BDA 213)

School: SBSR		**Batch : 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2022-23**
Branch: **Data Science & Analytics**		**Semester: III**
1	Course Code	BDA 213
2	Course Title	Data Structure & Theory of Algorithms
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	To make students familiar with the data structure & algorithms. The concept of data organizations, data structure operations; analysis of an algorithm; Stacks and Queues; Linked Lists; Sorting and Hashing; Graph.
6	Course Outcomes	CO1: Explain and illustrate the concepts basic terminologies: elementary data organizations, data structure operations: insertion, deletion, traversal etc. (K2, K3, K4) CO2: Describe the analysis of an algorithm, asymptotic; notations, time-space trade off. (K1, K2, K3) CO3: Describe Linear Search and Binary Search Techniques and explain their complexity analysis. (K2, K3, K4) CO4: Describe ADT Stack and its operations: Algorithms and their complexity analysis, Applications of Stacks; Types of Queue; Algorithms and their analysis. (K2, K3, K4) CO5: Describe the Singly linked lists; trees; algorithms and analysis. (K2, K3, K6) CO6: Describe and analyze the basic concepts of Sorting and Hashing; Graphs. (K1,K2, K4)
7	Course Description	This course an introduce data structure & algorithms. The concept of data organizations, data structure operations; analysis of an algorithm; Stacks and Queues; Linked Lists; Sorting and Hashing; Graph.
8	Outline syllabus :	

	Unit 1			
	A	Basic Terminologies: Elementary Data Organizations,		
	B	Data Structure Operations: insertion		
	C	deletion, traversal etc.		
	Unit 2			
	A	Analysis of an Algorithm, Asymptotic;		
	B	Notations, Time-Space trade off. **Searching:** Linear Search		
	C	Binary Search Techniques and their complexity analysis.		
	Unit 3			
	A	**Stacks and Queues**: ADT Stack and its operations: Algorithms and their complexity analysis,		
	B	Applications of Stacks: Expression Conversion and evaluation – corresponding algorithms and complexity analysis.		
	C	ADT queue, Types of Queue: Simple Queue, Circular Queue, Priority Queue; Operations on each types of Queues: Algorithms and their analysis.		
	Unit 4			
	A	**Linked Lists:** Singly linked lists: Representation in memory, Algorithms of several operations: Traversing, Searching, Insertion into, Deletion from linked list;		
	B	Linked representation of Stack and Queue, Header nodes, Doubly linked list: operations on it and algorithmic analysis; Circular Linked Lists: all operations their algorithms and the complexity analysis.		
	C	**Trees:** Basic Tree Terminologies, Different types of Trees: Binary Tree, Threaded Binary Tree, Binary Search Tree, AVL Tree; Tree operations on each of the trees and their algorithms with complexity analysis. Applications of Binary Trees. B Tree, B+ Tree: definitions, algorithms and analysis.		
	Unit 5			
	A	**Sorting and Hashing:** Objective and properties of different sorting algorithms: Selection Sort, Bubble Sort, Insertion Sort, Quick Sort, Merge Sort, Heap Sort;		
	B	Performance and Comparison among all the methods, Hashing.		
	C	**Graph:** Basic Terminologies and Representations, Graph search and traversal algorithms and complexity analysis.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50

	Text book/s*	1.Fundamentals of Data Structures", Illustrated Edition by Ellis Horowitz, Sartaj Sahni, Computer Science Press.
	Other References	1.Algorithms, Data Structures, and Problem Solving with C++", Illustrated Edition by Mark Allen Weiss, Addison-Wesley Publishing Company. 2. How to Solve it by Computer", 2nd Impression by R. G. Dromey, Pearson Education.

DATA PREPARATION AND DATA CLEANING (BDA 201)

School: SBSR		**Batch : 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2022-23**
Branch: **Data Science & Analytics**		**Semester: IV**
1	Course Code	BDA 201
2	Course Title	Data preparation and Data Cleaning
3	Credits	6
4	Contact Hours (L-T-P)	6-0-0
	Course Status	Compulsory
5	Course Objective	To make students familiar with the concepts of preparing your data; Working with dates and times, Data Cleaning, Data Structure, Cleaning Text Data.
6	Course Outcomes	CO1: Describe preparing data: Rearranging and removing variables, Renaming variables, Variable classes, Calculating new numeric variables and explain how to Dividing a continuous variable into categories, Working with factor variables. (K1, K3) CO2: Discuss how to working with dates and times, adding and removing observations and explain about removing duplicate observations, selecting a subset of the data, selecting a random sample from a dataset, sorting a dataset. (K2, K3, K4) CO3: Explain the data cleaning and technical representation of data. (K2,K3, K4) CO4: Discuss about the data structure. (K2, K6) CO5: Describe Character Normalization, Encoding Conversion and Unicode Normalization, Character Conversion and Transliteration. (K1, K2) CO6: Discuss and evaluate Generating Regular Expressions in R, Common String Processing Tasks in R, Approximate Text Matching, String Metrics, String Metrics and Approximate Text Matching in R.
7	Course Description	This course is an introduces preparing your data; Working with dates and times, Data Cleaning, Data Structure, Cleaning Text Data..
8	Outline syllabus	
	Unit 1	
	A	Preparing your data: Rearranging and removing variables, Renaming variables, Variable classes, Calculating new numeric variables,
	B	Dividing a continuous variable into categories, Working with factor variables,
	C	**Manipulating character variables:** Concatenating character strings, Extracting a substring, Searching a character variable.
	Unit 2	
	A	Working with dates and times**,** Adding and removing observations,
	B	Removing duplicate observations, Selecting a subset of the data,
	C	Selecting a random sample from a dataset, Sorting a dataset.
	Unit 3	

	A	**Data Cleaning:** The Statistical Value Chain, Raw Data, Input Data, Valid Data, Statistics, Output.		
	B	Technical Representation of Data: Numeric Data, Integers, Integers in R, Real Numbers, Double Precision Numbers, The Concept of Machine Precision, Consequences of Working with Floating Point Numbers, Dealing with the Consequences,		
	C	Numeric Data in R, Text Data, Terminology and Encodings, Unicode, Textual Data in R: Objects of Class Character, Encoding in R, Reading and Writing of Data with Non-Local Encoding, Detecting Encoding, Collation and Sorting, Times and Dates. Time and Date Notation, Time and Date Storage in R, Time and Date Conversion in R, Leap Days, Time Zones, and Daylight Saving Times.		
	Unit 4			
	A	**Data Structure:** Introduction, Tabular Data, data.frame, Databases, dplyr, Matrix Data, Time Series,		
	B	Graph Data, Web Data, Web Scraping, Web API, Other Data, Tidying Tabular Data,		
	C	Variable Per Column, Single Observation Stored in Multiple Tables.		
	Unit 5			
	A	Cleaning Text Data: Character Normalization, Encoding Conversion and Unicode Normalization, Character Conversion and Transliteration,		
	B	Pattern Matching with Regular, Expressions, Basic Regular Expressions, Practical Regular Expressions, Generating Regular Expressions in R,		
	C	Common String Processing Tasks in R, Approximate Text Matching, String Metrics, String Metrics and Approximate Text Matching in R.		
	Mode of examination	Theory		
	Weightage Distribution	CA 25	MTE 25	ETE 50
	Text book/s*	1. Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work by Q. Ethan McCallum Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data by Jason W Osborne		
	Other References	1) Data Wrangling with Python by Jacqueline Kazil Principles of Data Wrangling: Practical Techniques for Data Preparation by Tye Rattenbury		

DATABASE MANAGEMENT SYSTEMS (BDA 202)

School: SBSR		**Batch : 2021- 2024**
Program: B.Sc. (H)		**Academic Year: 2022-23**
Branch: Data Science		**Semester: IV**
1	Course Code	BDA 202
2	Course Title	DATABASE MANAGEMENT SYSTEMS
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory

5	Course Objective	To make students familiar with the basic concepts of Databases and Transactions and Data Models, Database Design ,ER-Diagram and Unified Modeling Language, Relational Algebra and Calculus, Constraints, Views and SQL, Transaction management and Concurrency control.
6	Course Outcomes	CO1: Discuss the basics of Databases and Transactions and Data Models. (K1, K2, K3) CO2: Discuss about Database Design ,ER-Diagram and Unified Modeling Language. (K1, K3) CO3: Explain relational algebra and calculus, describe Domain relational Calculus, calculus vs algebra, computational capabilities. (K3, K4) CO4: Explain and illustrate Constraints, Views and SQL. (K3,K6) CO5: Evaluate different types of transaction management. (K4,K5) CO6: Explain concurrency control, time stamping methods, optimistic methods, database recovery management. (K2, K4, K5)
7	Course Description	This course is an introduce the basic concepts of Databases and Transactions and Data Models, Database Design ,ER-Diagram and Unified Modeling Language, Relational Algebra and Calculus, Constraints, Views and SQL, Transaction management and Concurrency control..
8	Outline syllabus : **DATABASE MANAGEMENT SYSTEMS**	
	Unit 1	**Introduction to Databases and Transactions and Data Models**
	A	What is database system, purpose of database system, view of data, relational databases, database architecture,
	B	Transaction management, The importance of data models, Basic building blocks,
	C	Business rules, The evolution of data models, Degrees of data abstraction.
	Unit 2	**Database Design ,ER-Diagram and Unified Modeling Language**
	A	Database design and ER Model: overview, ER-Model, Constraints, ER-Diagrams, ERD Issues, weak entity sets, Codd's rules, Relational Schemas,
	B	Introduction to UML Relational database model: Logical view of data, keys, integrity rules.
	C	Relational Database design: features of good relational database design, atomic domain and Normalization (1NF, 2NF, 3NF, BCNF).
	Unit 3	**Relational Algebra and Calculus**
	A	Relational algebra: introduction, Selection and projection, set operations, renaming, Joins, Division, syntax, semantics.
	B	Operators, grouping and ungrouping, relational comparison.
	C	Calculus: Tuple relational calculus, Domain relational Calculus, calculus vs algebra, computational capabilities.
	Unit 4	**Unit-IV Constraints, Views and SQL**
	A	What is constraints, types of constrains, Integrity constraints.
	B	Views: Introduction to views, data independence, security, updates on views, comparison between tables.

	C	Views SQL: data definition, aggregate function, Null Values, nested sub queries, Joined relations. Triggers.		
	Unit 5	**Unit-V Transaction management and Concurrency control**		
	A	Transaction management: ACID properties, serializability and concurrency control,		
	B	Lock based concurrency control (2PL, Deadlocks), Time stamping methods.		
	C	Optimistic methods, database recovery management.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE 50
		25	25	
	Text book/s*	1."Database System Concepts", 6th Edition by Abraham Silberschatz, Henry F. Korth, S. Sudarshan, McGraw-Hill		
	Other References	1 "Principles of Database and Knowledge – Base Systems", Vol 1 by J. D. Ullman, Computer science Press. 2 "Fundamentals of Database Systems", 5th Edition by R. Elmasri and S. Navathe, Pearson Education 3 "Foundations of Databases", Reprint by Serge Abiteboul, Richard Hull, VictorVianu, Addison-Wesley		

Operating System (BDA 204)

School: SBSR		**Batch: 2021- 2024**
Program: B.Sc. (H)		**Academic Year: 2022-23**
Branch: Data Science & Analytics		**Semester: IV**
1	Course Code	BDA 204
2	Course Title	OPERATING SYSTEM
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Code	**Compulsory**
5	Course Objective	To familiarise students with basic concepts of Operating Systems, Process Management Processes, Interprocess Communication Race Conditions, Deadlocks, Memory Management, I/O Management Principles of I/O Hardware, File Management.
6	Course Outcomes	CO1: Describe the concept of operating systems and process management processes. (K2) CO2: Explain the concept of interprocess communication race conditions, deadlocks (K2, K4) CO3: Recognize and decide basic memory management and virtual memory. (K1, K6) CO4: Define and discriminate I/O Management Principles of I/O Hardware and I/O Software. (K1, K6). CO5: Discuss about file management and directory implementation efficiency & performance. (K1,K2,K5) CO6:Explain Unix/Linux operating system and development of Unix/Linux. (K2,K4, K6)
7	Course	This course will cover basic concepts of Operating Systems, Process Management Processes,

	Description	Interprocess Communication Race Conditions, Deadlocks, Memory Management, I/O Management Principles of I/O Hardware, File Management.		
8	Outline syllabus			
	Unit 1			
	A	Introduction: Basics of Operating Systems: Definition – Generations of Operating systems – Types of Operating Systems, OS Service, System Calls, OS structure: Layered, Monolithic, Microkernel Operating Systems – Concept of Virtual Machine.		
	B	Process Management Processes: Definition , Process Relationship , Process states , Process State transitions , Process Control Block ,Context switching – Threads – Concept of multithreads , Benefits of threads – Types of threads		
	C	Process Scheduling: Definition , Scheduling objectives ,Types of Schedulers ,Scheduling criteria : CPU utilization, Throughput, Turnaround Time, Waiting Time, Response Time (Definition only) , Scheduling algorithms : Pre emptive and Non , pre emptive , FCFS – SJF – RR , Multiprocessor scheduling : Types , Performance evaluation of the scheduling		
	Unit 2			
	A	Interprocess Communication Race Conditions, Critical Section, Mutual Exclusion, Hardware Solution, Strict Alternation , Peterson's Solution,		
	B	The Producer Consumer Problem, Semaphores, Event Counters, Monitors, Message Passing, Classical IPC Problems: Reader's & Writer Problem, Dinning Philosopher Problem etc., Scheduling , Scheduling Algorithms.		
	C	Deadlocks: Definition,Deadlock characteristics , Deadlock Prevention , Deadlock Avoidance :banker's algorithm, Deadlock detection and Recovery		
	Unit 3			
	A	Memory Management Basic Memory Management: Definition ,Logical and Physical address map , Memory allocation : Contiguous Memory allocation – Fixed and variable partition – Internal and External fragmentation and Compaction ,		
	B	Paging : Principle of operation – Page allocation – Hardware support for paging –,Protection and sharing – Disadvantages of paging.		
	C	Virtual Memory: Basics of Virtual Memory – Hardware and control structures – Locality of reference, Page fault , Working Set , Dirty page/Dirty bit – Demand paging (Concepts only) – Page Replacement policies : Optimal (OPT) , First in First Out (FIFO), Second Chance (SC), Not recently used (NRU) and Least Recently used (LRU)		
	Unit 4			
	A	I/O Management Principles of I/O Hardware: I/O devices, Device controllers		
	B	, Direct memory access Principles of I/O Software: Goals of Interrupt handlers , Device drivers , Device		
	C	independent I/O software , Secondary-Storage Structure: Disk structure ,Disk scheduling algorithm		
	Unit 5			
	A	File Management File concept, Aaccess methods, File types, File operation, Directory structure, File System structure, Allocation methods (contiguous,linked, indexed),		
	B	Free-space management (bit vector, linked list, grouping), directory implementation (linear list, hash table),efficiency & performance.		
	C	Unix/Linux Operating System Development Of Unix/Linux, Role & Function Of Kernel, System Calls, Elementary Linux command & Shell Programming, Directory Structure, System Administration Case study: Linux, Windows Operating System.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book	1."Operating System Concepts" by Avi Silberschatz and Peter Galvin		

		2."Operating Systems: Internals and Design Principles" by William Stallings
	Other References	1."Operating Systems: A Concept-Based Approach" by D M Dhamdhere 2."Operating System: A Design-oriented Approach" by Charles Crowley

DATA WARE HOUSING AND DATA MINING (BDA 312)

School: SBSR		**Batch: 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2023-24**
Branch: Data Science		**Semester: V**
1	Course Code	BDA 312
2	Course Title	DATA WARE HOUSING AND DATA MINING
3	Credits	5
4	Contact Hours (L-T-P)	5-0-0
	Course Status	Compulsory
5	Course Objective	Familiarise students with basic concepts of data warehousing, business analysis, data mining, association rule mining and classification, clustering and trends in data mining.
6	Course Outcomes	CO1: Discuss about the Data warehousing Components, Cleanup and transformation Tools - Metadata. (K3, K5) CO2: Explain methods of business analysis, reporting and query tools and applications. (K2, K3, K4) CO3: Describe the OLAP guideline multidimensional versus multirelational OLAP, categories of tools, OLAP tools and the internet. (K2, K4) CO4: Explain and illustrate data mining functionalities, interestingness of patterns, integration of a data mining system with a data warehouse issues, data preprocessing . (K2, K3) CO5: Explain the basic concepts of decision tree induction, bayesian classification, rule based classification, classification by back propagation and apply support vector machines, associative classification, lazy learners, other classification methods, prediction. (K2, K3, K4) CO6: Explain and evaluate clustering and trends in data mining. (K2, K4, K6)
7	Course Description	This course is an introduce the basic concepts of data warehousing, business analysis, data mining, association rule mining and classification, clustering and trends in data mining..
8	Outline syllabus	
	Unit 1	**DATA WAREHOUSING**
	A	Data warehousing Components –Building a Data warehouse.
	B	Mapping the Data Warehouse to a Multiprocessor Architecture – DBMS Schemas for Decision Support
	C	Data Extraction, Cleanup, and Transformation Tools - Metadata.
	Unit 2	**BUSINESS ANALYSIS**
	A	Reporting and Query tools and Applications, Cognos Impromptu, Online Analytical Processing (OLAP).
	B	Multidimensional Data Model, OLAP Guideline Multidimensional versus Multirelational OLAP,
	C	Categories of Tools, OLAP Tools and the Internet.
	Unit 3	**DATA MINING**
	A	Introduction, Data, Types of Data, Data Mining Functionalities,
	B	Interestingness of Patterns, Classification of Data Mining Systems, Data Mining Task Primitives,
	C	Integration of a Data Mining System with a Data Warehouse Issues, Data Pre processing
	Unit 4	**ASSOCIATION RULE MINING AND CLASSIFICATION**
	A	Mining Frequent Patterns, Associations and Correlations, Mining Methods, Mining various

		Kinds of Association Rules, Correlation Analysis,		
	B	Constraint Based Association Mining Classification and Prediction, Basic Concepts, Decision Tree Induction, Bayesian Classification, Rule Based Classification, Classification by Back propagation,		
	C	Support Vector Machines, Associative Classification, Lazy Learners, Other Classification Methods, Prediction.		
	Unit 5	**CLUSTERING AND TRENDS IN DATA MINING**		
	A	Cluster Analysis, Types of Data, Categorization of Major Clustering Methods, K-means, Partitioning Methods, Hierarchical Methods,		
	B	Density-Based Methods, Grid Based Methods, Model Based Clustering Methods, Clustering High Dimensional Data, Constraint, Based Cluster Analysis, Outlier Analysis.		
	C	Data Mining Applications. Apply data mining techniques and methods to large data sets, Use data mining tools, Compare and contrast the various classifiers.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Alex Berson and Stephen J.Smith, "Data Warehousing, Data Mining and OLAP", Tata McGraw – Hill Edition, Thirteenth Reprint 2008. 2. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.		
	Other References	1. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", Person Education, 2007. 2. K.P. Soman, Shyam Diwakar and V. Aja, "Insight into Data Mining Theory and Practice", Eastern Economy Edition, Prentice Hall of India, 2006. 3. G. K. Gupta, "Introduction to Data Mining with Case Studies", Eastern Economy Edition, Prentice Hall of India, 2006. 4. Daniel T.Larose, "Data Mining Methods and Models", Wiley-Interscience, 2006.		

REGRESSION, TIME SERIES, FORECASTING AND INDEX NUMBERS (BDA 313)

School: SBSR		**Batch : 2021- 2024**
Program: B.Sc. (H)		**Academic Year: 2023-24**
Branch: **Data Science & Analytics**		**Semester: V**
1	Course Code	BDA 313
2	Course Title	Regression, time series, forecasting and Index numbers
3	Credits	5
4	Contact Hours (L-T-P)	5-0-0
	Course Status	Compulsory
5	Course Objective	The objective of the course is to explain basic concepts of regression, time series, forecasting and index numbers.
6	Course Outcomes	CO1: Explain and illustrate the nature and uses of forecasts, some examples of time series, the forecasting process, resources for forecasting, statistics background for forecasting: graphical displays, numerical description of time series data (K2, K3) CO2: Describe how to evaluate least squares estimation in linear regression models, statistical inference in linear regression, prediction of new observations, model adequacy checking, model adequacy checking, generalized and weighted least squares, regression models for general time series data. (K6) CO3: Explain and illustrate first-order exponential smoothing, modeling time series data, second-order exponential smoothing, higher-order exponential smoothing. (K3, K6)

		CO4: Use forecasting: constant process, linear trend process and evaluate estimation of σ_e^2, adaptive updating of the discount factor, model assessment. (K3, K6) CO5: Describe autoregressive integrated moving average (arima) models. (K2) CO6: Explain and illustrate index numbers with application. (K6)
7	Course Description	This course will cover the fundamental concepts of Regression, time series, forecasting and Index numbers.
8	Outline syllabus	
	Unit 1	
	A	Introduction to Forecasting: The Nature and Uses of Forecasts, Some Examples of Time Series, The Forecasting Process, Resources for Forecasting,
	B	Statistics Background for Forecasting: Graphical Displays, Numerical Description of Time Series Data, Use of Data Transformations and Adjustments,
	C	General Approach to Time Series Modeling and Forecasting, Evaluating and Monitoring Forecasting Model Performance
	Unit 2	
	A	Regression Analysis and Forecasting: Least Squares Estimation in Linear Regression Models
	B	, Statistical Inference in Linear Regression, Prediction of New Observations, Model Adequacy Checking
	C	, Model Adequacy Checking, Generalized and Weighted Least Squares, Regression Models for General Time Series Data.
	Unit 3	
	A	Exponential Smoothing Methods: First-Order Exponential Smoothing, Modeling Time Series Data
	B	, Second-Order Exponential Smoothing, Higher-Order Exponential Smoothing,
	C	Forecasting: Constant Process, Linear Trend Process, Estimation of σ_e^2, Adaptive Updating of the Discount Factor, Model Assessment.
	Unit 4	
	A	Autoregressive Integrated Moving Average (ARIMA) Models : Linear Models for Stationary Time Series, Stationary Time Series, 3 Finite Order Moving Average (MA) Processes.
	B	The First-Order Moving Average Process, MA(l), The Second-Order Moving Average Process, MA(2),Finite Order Autoregressive Processes,1 First -Order Autoregressive Process, AR(l), Second-Order Autoregressive Process, AR(2),
	C	General Autoregressive Process, AR(p), Partial Autocorrelation Function, PACF, Mixed Autoregressive-Moving Average CARMA) Processes, Time Series Model Building, Model Identification, Parameter Estimation, Examples of Building ARIMA Models, Forecasting ARIMA Processes.
	Unit 5	
	A	Index Numbers: Definition, construction of index numbers and problems thereof for weighted and unweighted index numbers including
	B	Laspeyre's, Paasche's, Edgeworth-Marshall and Fisher's. Chain index numbers,
	C	Conversion of fixed based to chain based index numbers and vice-versa. Consumer price index numbers.
	Mode of examination	Theory
	Weightage Distribution	CA: 25 / MTE: 25 / ETE: 50
	Text book/s*	1.Business Statistics: For Contemporary Decision Making, 7th Edition by Ken Black
	Other References	1.Daniel,WayneW.,"Biostatistics": Basic concept and Methodology for Health Science. 2.Grewal,B.S, "Higher Engineering Mathematics".

Machine learning & Deep Learning (BDA 314)

School: SBSR		**Batch: 2021-2022**
Program: B.Sc.(H)		**Academic Year: 2023-24**
Branch: **Data Science & Analytics**		**Semester: V**
1	Course Code	BDA 314
2	**Course Title**	**Machine learning & Deep Learning**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	The objective of this course is to introduce machine learning fundamentals to students.
6	Course Outcomes	CO1: Recognize the characteristics of machine learning that make it useful to real-world problems. Characterize machine learning algorithms as supervised, semi-supervised, and unsupervised (K2,K3). CO2: Design and implement machine learning solutions to classification, regression, and clustering problems (K3, K6). CO3: Be able to evaluate and interpret the results of the algorithms .(Effectively use machine learning toolboxes (K4, C5, (K6). CO4: Ability to identify the deep learning techniques. Ability to select and implement Machine learning and deep learning. Ability to Train machine and solve problems associated with batch learning and online learning (K2, K3, K4). (K2,K3,K4). CO5: Ability to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques. Ability to integrate deep learning libraries and mathematical and statistical tools (K4, K5). (K3, K4,K5). CO6:Ability to apply Deep learning Techniques to various engineering and social applications (K4, K6).
7	Course Description	This course provides introductory concepts of various machine learning techniques to students which will help to build foundation for further understanding. This course also aims to provide details of various steps involved in machine learning pipeline such as data collection, pre-processing, feature engineering etc. This course also introduces popular tools used in the area of machine learning. This course mainly focused on Regression and Neural network based Machine learning algorithms. This aim to make students aware of various recent developments in the field of Neural network such as deep learning.
8	Outline syllabus :	
	Unit 1	
	A	Machine Learning Fundamentals –Types of Machine Learning - Supervised, Unsupervised, Reinforcement- The Machine Learning process.Terminologies in ML- Testing ML algorithms:

		Over fitting, Training, Testing and Validation Sets-Confusion matrix -Accuracy metrics- ROC Curve.Basic Statistics: Averages, Variance and Covariance, The Gaussian- The Bias-Variance trade off- Applications of Machine Learning.
	B	Regression: Linear Regression – Multivariate Regression analysis, Linear Basis Function Models, The Bias-Variance Decomposition, Bayesian Linear Regression
	C	Classification: Linear Discriminant Analysis, Logistic Regression- K-Nearest Neighbor classifier.Decision Tree based methods for classification and Regression- Ensemble methods.
	Unit 2	
	A	Clustering- K-Means clustering, Hierarchical clustering.The Curse of Dimensionality – Dimensionality Reduction - Principal Component Analysis - Probabilistic PCA- Independent Components analysis.The Internet, Business and Retail, Law Enforcement, Computing, Clustering models: How the K-means and PCA works, Calculating the number of clusters in a dataset.
	B	Perceptron- Multilayer perceptron- Back Propagation- Initialization, Training and Validation Support.Vector Machines(SVM) as a linear and non-linear classifier - Limitations of SVM.
	C	Recognition of MNIST handwritten digits using Artificial Neural Network. Build an email spam classifier using SVM.
	Unit 3	
	A	Bayesian Networks - Learning Naive Bayes classifiers-Markov Models – Hidden Markov Models.
	B	Sampling – Basic sampling methods – Monte Carlo -Reinforcement Learning.Classify the given text segment as ‘Positive’ or ‘Negative’ statement using the Naive Bayes Classifier.
	C	Predict future stock price of a company using Monte Carlo Simulation.
	Unit 4	
	A	History of Deep Learning, Mc Culloch Pitts Neuron.Multilayer Perceptron’s (MLPs), Representation Power of MLPs, Sigmoid Neurons, Gradient Descent.Feed Forward Neural Networks, Back propagation.
	B	Gradient Descent (GD), Momentum Based GD, Nesterov Accelerated GD, Stochastic GD.Principal Component Analysis and its interpretations, Singular Value Decomposition.Auto encoders and relation to PCA, Regularization in auto encoders, Denoising auto encoders, Sparse auto encoders.
	C	Regularization: Bias Variance Tradeoff, L2 regularization, Early stopping, Data set augmentation.Greedy Layer wise Pre-training, Better activation functions, Better weight initialization methods, Batch Normalization.Learning Vectorial Representations Of Words.
	Unit 5	
	A	Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet.Recurrent Neural Networks, Back propagation through time (BPTT), Vanishing and Exploding Gradients, Truncated BPTT, GRU, LSTMs.Encoder Decoder Models, Attention Mechanism, Attention over images.
	B	Advanced Deep architectures: Recurrent Neural networks (RNNs), Generative Adversarial Networks (GANs).In-depth discussion of DL examples.
	C	Advanced topics, Recent papers, Influential papers: Visual Question Answering, Visual Dialog, Novel deep methods (Deep internal learning, Deep image prior).

	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Mitchell Tom, Machine Learning. McGraw Hill, 1997. 2. Introduction to machine learning, EthemAlpaydin. —2nd ed., The MIT Press, Cambridge, Massachusetts, London, England. Dr. Nilesh Shelke, Dr. Gopal Sakarkar, Dr N V Choudhari, Introduction to Machine Learning, Ganu Prakashan.		
	Other References	1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction 2. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern classification, Wiley, New York, 2001. 3. Machine Learning: The Art and Science of Algorithms that Make Sense of Data (1st Edition) – Peter Falch. 4. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012. 5. Stephen Marsland, "Machine Learning –An Algorithmic Perspective", CRC Press, 2009. 6. Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, "Machine Learning", Pearson Education, 2018. Andreas C. Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", O'Reilly,2016.		

MULTIVARIATE ANALYSIS (BDA 315)

School: SBSR		**Batch: 2021- 2024**
Program: B. Sc. (H)		**Academic Year: 2023-24**
Branch: **Data Science & Analytics**		**Semester: V**
1	Course Code.	BDA 315
2	Course Title	Multivariate Analysis
3	Credits	5
4	Contact Hours (L-T-P)	5-0-0
	Course status	Compulsory
5	Course Objectives	Familiarise students with the multivariate normal distribution, estimation of the mean vector and the covariance matrix, the distributions and uses of sample correlation coefficients, classification of observations, the distribution of the sample covariance matrix and the sample generalized variance.
6	Course Outcomes	CO1: Demonstrate knowledge and understanding of the multivariate normal distribution. (K2, K3) CO2: Demonstrate knowledge and understanding the concept of estimation of the mean vector and the covariance matrix. (K2, K3) CO3: Demonstrate advanced understanding of the concepts of dimension reduction technique. (K2, K3) CO4: Describe the concepts of how to use and apply dependence techniques in multivariate data analysis. (K2, K3)

		CO5: Describe the concepts of analysis of variance and covariance in multivariate data analysis. (K3, K4, K5) CO6: Apply the statistical tool and software in multivariate data analysis. (K2, K6)
7	Course Description	The aim of this module is to provide an understanding of the multivariate normal distribution, estimation of the mean vector and the covariance matrix, the distributions and uses of sample correlation coefficients, classification of observations, the distribution of the sample covariance matrix and the sample generalized variance.
8	Outline syllabus:	
UNIT 1		
A	Brief review of Univariate and Bivariate distribution with their properties.	
B	Basic Multivariate Distribution: mean, variance, Covariance, correlation, linear combination of variables.	
C	The multivariate normal distribution, Mean Vectors and Covariance Matrices.	
UNIT 2		
A	Multivariate normal distribution; maximum likelihood estimation, Wishart's distribution	
B	Hotelling's T2 and hypothesis testing for multivariate normal data. Inference from a single sample, Inference from two dependent samples Inference from two independent samples.	
C	Simple, Multiple, Partial and Canonical correlation with their properties.	
UNIT 3		
A	Principal Components Analysis and derivation of principal components; PCA structural model; PCA on normal populations; bi-plots.	
B	Factor Analysis, Factor extraction Factor rotation, Factor scores Validation of factor analysis, Higher order factor analysis Q-type factor analysis	
C	Cluster Analysis, Types of clustering, Correlation and distance, Partitioning methods, hierarchical clustering, K-means clustering and their interpretation.	
UNIT 4		
A	Simple, Multiple, Multivariate regression with their properties.	
B	Binary and multidimensional Logistic regression.	
C	Linear discriminant function analysis. Estimating linear discriminant functions and their properties.	
UNIT 5		
A	Analysis of variance and covariance.	
B	Multivariate analysis of variance and Covariance.	
C	Concepts of correspondence analysis, chi-square distance and inertia, multiple correspondence analysis.	

	Mode of Examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50

	Text books	1. Johnson, R.A. and Wichern, D.W.: (2015). Applied Multivariate Statistical Analysis, Sixth Edition, Pearson Education India. 2. Hardle, W.K. and Hlavka, Z. (2015): Multivariate Statistics, Springer.
	Other references	1. Anderson, T.W. (2003): An Introduction to Multivariate Statistical Analysis, Third Edition, Wiley. 2. Härdle, W.K. and Simar, L. (2015): Applied Multivariate Statistical Analysis, Springer.

Statistical Analysis (Count Data and survival Analysis): BDA 316

School: SBSR		**Batch: 2021-2022**
Program: B.Sc.(H)		**Academic Year: 2023-24**
Branch: Data Science& Analytics		**Semester: V**
1	Course Code	BDA 316
2	Course Title	**Statistical Analysis & Simulation**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
6	Course Status	**Compulsory**
7	Course Objective	To demonstrate and intended to verse students in the techniques necessary to understand and carry out methods of research in survival analysis.
6	Course Outcomes	CO1: Explain the concept of survival data, and the roles played by censoring, and survival and hazard functions. Format data appropriately for analysis, and understanding (K1, K2, K3). CO2:Apply and drew the graph of survival data, and the Kaplan – Meier curve (K1,K2,K4) CO3: Explain the concept of Kernel smoothed distribution estimator and kernel smoothed hazard rate estimator .Describe how to fit the Cox Proportional Hazards model (K3, K4, K5). CO4: Apply models to the data analysis using the Cox proportional hazards model (K3, K4, K5). CO5: Recognize the concepts of probability and statistics that are relevant to modeling and simulation. How to generate random numbers by the different methods (K3, K4, K5). CO6:Design and implement Bootstrapping; jackknife resampling. How simulation may be used to understand the behavior of real world systems by utilizing mathematical models with an emphasis on simulation (K3, K4, K5).
7	Course Description	A UG-level course in survival analysis, intended to verse students in the techniques necessary to understand and carry out methods of research in survival analysis. Lectures study the large-sample properties of estimators based on one-sample, k-sample and partial likelihood inference, with proofs based on counting process and Martingale theory. The theory of competing risks is studied from several angles. Many extensions of the Cox model to more complex data structures are considered.
8	Outline syllabus	
	Unit 1	
	A	Basic quantities. The survival function. The hazard function. The mean residual life time function and median life. Common parametric models for survival data. Models for competing risks.
	B	Right censoring. Left or interval censoring. Truncation. Likelihood construction for censored and truncated data. Basic ideas for counting processes and martingales.
	C	Nonparametric estimators of the survival and cumulative hazard functions. Kaplan-Meier estimator and Nelson-Allen estimator.
	Unit 2	
	A	Point wise confidence intervals for the survival and cumulative hazard functions. Confidence bands for the survival function. Point and interval estimates of the mean and median survival time, and quintiles.
	B	Estimators of the survival function for left-truncated and right-censored data. Summary curves for competing risks.
	C	Estimating the survival function for left, double and interval censoring. Estimation of the survival functions for right-truncated data. Estimation in the cohort life table or grouped data.
	Unit 3	
	A	Kernel smoothed distribution estimator and kernel smoothed hazard rate estimator. Hypothesis testing. One-sample tests. Tests for two samples and more than two samples. Tests for trend. Stratified log-rank test.
	B	Parametric models with covariates. The accelerated failure time (AFT) model. Some popular AFT models. Diagnostic methods for parametric models.
	C	Additional materials: Model building and high-dimensional data analysis using the Cox proportional hazards model.
	Unit 4	

	A	Review of R/Python.Random number generation: Inverse-transform; acceptance-rejection; transformations.		
	B	Statistic simulations: generating random variables, simulating normal, gamma and beta random variables.		
	C	Simulating multivariate distributions, MCMC methods.Gibbs sampler, simulating random fields, simulating stochastic process.		
	Unit 5			
	A	Bootstrapping; jackknife resampling.Bootstrapping for estimation of sampling distribution. Confidence intervals, variance stabilizing transformation.Bootstrapping in regression and sampling from finite populations.		
	B	Simulating a non-homogeneous Poisson process.Optimization using Monte Carlo methods simulated annealing for optimization.Simulating a non-homogeneous Poisson process.		
	C	Optimization using Monte Carlo methods simulated annealing for optimization.Solving differential equations by Monte Carlo methods. Univariate density estimation; kernel smoothing multivariate density estimation.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	Lee, E. T. and Wang, J. W. (2003).Statistical Methods for Survival Data Analysis, 3rdEdition. John Wiley.		
	Other References	1. Liu, X. (2012).Survival Analysis: Models and Applications, Wiley, New York. 2. Kleinbaum, D. G. andKlein, M.(2012). SurvivalAnalysis:A Self-Learning Text, 3rdEd, Springer, New York. Hosmer, D. and Lemeshow, S. (1999).Applied Survival Analysis: Regression Modeling of Time to Event Data, Wiley, New York		

Data Scientist Toolbox (BDA 302)

School: SBSR		**Batch: 2021-2022**
Program: B.Sc.(H)		**Academic Year: 2023-24**
Branch: Data Science& Analytics		**Semester: V**
1	Course Code	BDA 302
2	**Course Title**	**Data Scientist Toolbox**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	**Compulsory**
5	Course Objective	To make students familiar with the main tool and idea of the Data Scientist Toolbox.
6	Course Outcomes	CO1: Explain how to set up R, R-studio, Github and other useful tools (K2, K3) CO2: Understand the data, problem and tools that data analysts use (K2, K3) CO3: Explain essential study design concepts (K3) CO4: Explain concept for identification of tools for applying on particular problem (K3, K4). CO5: Create a Github repository (K5, K6). CO6: Explain how to analysis big data (K3, K4).

7	Course Description	In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, R, and R Studio.
8	**Outline syllabus**	
	Unit 1	
	A	Data Science Fundamentals: Why Automated Videos? What is Data Science?
	B	What isData? Getting Help.
	C	The Data Science Process.
	Unit 2	
	A	R and R Studio. Installing R. Installing R Studio.
	B	R Studio Tour. R Packages.
	C	Projects in R.
	Unit 3	
	A	Version Control and GitHub.
	B	Version Control.
	C	Github and Git.
	Unit 4	
	A	Linking Github and RStudio.
	B	Projects under Version Control.
	C	R Markdown. Scientific Thinking.
	Unit 5	
	A	R Markdown with Big Data.
	B	Types of Data ScienceQuestions.
	C	Experimental Design with Big Data.
	Mode of examination	Theory

	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Data Science Mindset. Methodologies. and Misconceptions By Zacharias Voulgaris Donein-Specific Languages in R. Advanced Statistical Programming By Thomas Mailund.		
	Other References	1. R programming for data science by roger d. Peng The analytics lifecycle toolkit a practical guide for an effective analytic s capability, by Greg Nelson.		

STATISTICAL INFERENCE (NON- PARAMETRIC) (BDA 317)

School: SBSR		**Batch: 2020- 2023**
Program: B.Sc. (H)		**Current Academic Year: 2021-22**
Branch: **Data Science & Analytics**		**Semester: VI**
1	Course Code	**BDA 317**
2	Course Title	**STATISTICAL INFERENCE (NON- PARAMETRIC)**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	Familiarise students with basic concepts of order statistics, nonparametric estimation, interval estimation and tolerance limits, permutation tests, ordered least squares estimators.
6	Course Outcomes	CO1: Explain the concept of order statistics and large sample properties of sample quintiles. (K2, K4) CO2: Apply the concept of nonparametric estimation and explain completeness of the order statistic. (K3) CO3: Explain and use ordered least squares estimators. (K2, K3, K4) CO4: Explain optimum properties of ordered least squares estimates.(K2, K4) CO5: Describe the interval estimation and tolerance limits. (K1, K2) CO6: Understand and evaluate permutation tests and modified permutation tests. (K2, K6)
7	Course Description	This course will cover the basic concepts of order statistics, nonparametric estimation, interval estimation and tolerance limits, permutation tests, ordered least squares estimators.
8	Outline syllabus	
	Unit 1	
	A	Order Statistics: Domain of Nonparametric Statistics, Order Statistics, Distribution Theory of Order Statistics, Distribution of Sample Range and Mid Range,
	B	The Distribution of the Median, Sampling Distribution of the Coverages, Moments of Order Statistics, Order Statistics for Discrete Populations, Representation of

		Exponential Order Statistics as a Sum of Independent Random Variables,		
	C	Representation of General Order Statistics, Angel and Demons' Problems, Large Sample Properties of Order Statistics, Large Sample Properties of Sample Quintiles.		
	Unit 2			
	A	Nonparametric Estimation: Problems in Non-parametric Estimation, One-sided Confidence Interval for p,		
	B	Two-sided Confidence Interval for p, Estimation of Distribution Function,		
	C	Characterization of Distribution-free Statistics, Completeness of the Order Statistic.		
	Unit 3			
	A	Ordered Least Squares Estimators: Explicit Formulae for Estimators,		
	B	Estimation for Symmetric Populations, Estimation in a Single Parameter Family,		
	C	Optimum Properties of Ordered Least Squares Estimates.		
	Unit 4			
	A	Interval Estimation and Tolerance Limits: Confidence Intervals for Quantiles,		
	B	Large Sample Confidence Intervals: Wilks' (1962) Method, Tolerance Limits,		
	C	Distribution-free Tolerance Limits, Other Tolerance Limit Problems, Tolerance Regions .		
	Unit 5			
	A	Permutation Tests: Bivariate Independence, Two-sample Problems, Critical Regions Having Structures,		
	B	Most Powerful Permutation Tests, One-sample Problems, Tests in Randomized Blocks,		
	C	Large-sample Power, Modified Permutation Tests.		
	Mode of examination	**Theory**		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text books	1. Gibbons, J.D. & Chakraborti, S. (2010). Nonparametric Statistical Inference, 5th Edition. CRC Press. Hollander, M., Wolfe, D. & Chicken, E. (2013). Nonparametric Statistical Methods, 3rd Edition. Wiley.		
	Other references	1. Bonnini, S., Corain, L., Marozzi, M. & Salmaso, L. (2014). Nonparametric Hypothesis Testing Rank and Permutation Methods with Applications in R. Wiley. Sprent, P. & Smeeton, N.C. (2013). Applied Nonparametric Statistical Methods, 4th Edition. CRC Press.		

DATA VISUALIZATION (BDA 318)

School: SBSR		**Batch: 2021- 2024**
Program: B.Sc. (H)		**Current Academic Year: 2023-24**
Branch: **Data Science & Analytics**		**Semester: VI**
1	Course Code	**BDA318**
2	Course Title	DATA VISUALIZATION
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	Familiarise students with basic concepts of data visualization. Give an idea of data-analytic thinking, storytelling with data, data visualization using tableau 1. Given an understanding of a decision analytic thinking, fitting a model to data. Discuss the concept of visualizing model performance, data visualization using tableau 2, similarity, neighbors, and clusters.
6	Course Outcomes	CO1: Explain the concept of data-analytic thinking. (K2, K4) CO2: Discuss the concept of data understanding; data preparation; modelling; evaluation; deployment. Analytic techniques and technologies. (K3) CO3: Explain the use of storytelling with data and support vector machines, decision trees.(K2, K3, K4) CO4: Explain the data visualization using tableau 1 and decision analytic thinking. (K2, K4,K5) CO5: Describe the fitting a model to data and visualizing model performance. (K1, K2, K4) CO6: Explain and evaluate data visualization using tableau 2 and similarity, neighbors, and clusters. (K2, K6)
7	Course Description	This course will cover the basic concepts of data visualization. Give an idea of data-analytic thinking, storytelling with data, data visualization using tableau 1. Given an understanding of a decision analytic thinking, fitting a model to data. Discuss the concept of visualizing model performance, data visualization using tableau 2, similarity, neighbors, and clusters
8	Outline syllabus	
	Unit 1	
	A	**Data-Analytic Thinking**: The Ubiquity of Data Opportunities,*f* Data Processing and "Big Data" *f* From Big Data 1.0 to Big Data 2.0,
	B	Data and Data Science Capability as a Strategic Asset. From Business Problems to Data Mining Tasks: Business Understanding;
	C	Data Understanding; Data Preparation; Modeling; Evaluation; Deployment. Analytic techniques and technologies.
	Unit 2	
	A	**Story Telling with Data**: Importance of context ; Choosing an effective visual ; Focus audience's attention ;Thinking like designer ;
	B	Dissecting model visuals ;Lessons in story telling ;Putting it all together ; Case studies. Introduction to
	C	Predictive Modeling: Linear Regression; *f*Classification: Logistic, Regression, Support Vector Machines, Decision Trees.
	Unit 3	
	A	**Data Visualization Using Tableau 1**: *f* Introduction to Tableau; Data Import and Management: Data import, Extract and live , Data management – Join, Data management – Relationship, Data Management – Replace; Data Type and Operation: Data type, Pivot and separate , Change type, Set and group, Hierarchy.
	B	**Decision Analytic Thinking**: Targeting the Best Prospects for a Charity Mailing -The Expected Value Framework: Decomposing the Business Problem and Recomposing the

		Solution Pieces , A Brief Digression on Selection Bias;		
	C	Churn Example Revisited with Even More Sophistication - The Expected Value Framework: Structuring a More Complicated Business Problem , Assessing the Influence of the Incentive; From an Expected Value Decomposition to a Data Science Solution.		
	Unit 4			
	A	**Fitting a Model to Data:** What is a good model? -Overfitting , Generalization *ƒ* Evaluating Classifiers , Plain Accuracy and Its Problems , Confusion Matrix , Problems with Unbalanced Classes, Problems with Unequal Costs and Benefits ;		
	B	*ƒ* Generalizing Beyond Classification - Using Expected Value to Frame Classifier Evaluation; *ƒ* Evaluation, Baseline Performance, and Implications for Investments in Data.		
	C	**Visualizing Model Performance**: *ƒ* Ranking Instead of Classifying; Profit Curves; ROC Graphs and Curves; The Area Under the ROC Curve (AUC); Cumulative Response and Lift Curves; Example: Performance Analytics for Churn Modeling.		
	Unit 5			
	A	**Data Visualization Using Tableau 2**: *ƒ* Different types of data visualizations - Visual encoding , Bar chart and pie chart , Line chart ,Multiple chart and distribution , Highlight tables , Scatter plot and trend lines, Heatmap , Geographic mapping ,Bullet graph , Gnatt chart , Data calendar , Circle view.		
	B	**Similarity, Neighbors, and Clusters** : Similarity and Distance; Nearest-Neighbor Reasoning o Example: Whiskey Analytics , How Many Neighbors and How Much Influence? , Issues with Nearest-Neighbor Methods;		
	C	Clustering - Hierarchical clustering *ƒ* Example: Whiskey Analytics, Nearest Neighbors Revisited: Clustering Around Centroids; *ƒ* Example: Clustering Business News Stories - Understanding the Results of Clustering; Stepping Back: Solving a Business Problem Versus Data Exploration.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text books	1) Information Dashboard Design: Displaying Data for At-a-glance Monitoring" by Stephen Few Beautiful Visualization, Looking at Data Through the Eyes of Experts by Julie Steele, Noah Iliinsky		
	Other references	1. The Accidental Analyst: Show Your Data Who's Boss" by Eileen and Stephen McDaniel		

Digital Marketing (BDA 308)

1	Course Code	BDA 308
2	**Course Title**	Digital Marketing
3	Credits	5
4	Contact Hours (L-T-P)	5-0-0
	Course Status	Elective
5	Course Objective	The aim of the Digital Marketing Course is to provide students with the knowledge about business advantages of the digital marketing and its importance for marketing success.
6	Course Outcomes	CO1: Students will be able to identify the importance of the digital marketing for marketing success(K2, K3) . CO2: To manage customer relationships across all digital channels and build better customer relationships(K2, K3). CO3: To create a digital marketing plan, starting from the SWOT analysis and defining a target group(K3, K4). CO4: Describe and identifying digital channels for their advantages and limitations(K3,K4). CO5: Describe how to do business with different digital platform and also cost optimization through this platform (K2, K3, K4). CO6: Illustrate how to make a sample Business model of digital marketing in different platform. (K3,K6)
7	Course Description	In this course to provide students with the knowledge about business advantages of the digital marketing and its importance for marketing success; to develop a digital marketing plan; to make SWOT analysis; to define a target group; to get introduced to various digital channels, their advantages and ways of integration; to get basic knowledge of Google Analytics for measuring effects of digital marketing and getting insight of future trends that will affect the future development of the digital marketing.
8	Outline syllabus:	
	Unit 1	
	A	Introduction of the digital marketing, Digital vs. Real Marketing, Digital Marketing Channels.
	B	Creating initial digital marketing plan, content management, SWOT analysis, target group analysis.
	C	.Web design, Optimization of Web sites, MS Expression Web.
	Unit 2	
	A	SEO Optimization, Writing the SEO content.
	B	Google Ad Words- creating accounts, Google Ad Words- types.

	C	Introduction to CRM, CRM platform, CRM models.		
	Unit 3			
	A	Introduction to Web analytics, Web analytics – levels, Introduction of Social Media Marketing.		
	B	Creating a Facebook page, Visual identity of a Facebook page, Types of publications.		
	C	Business opportunities and Instagram options, Optimization of Instagram profiles, Integrating Instagram with a Web Site and other social networks, Keeping up with posts.		
	Unit 4			
	A	Business tools on LinkedIn, Creating campaigns on LinkedIn, Analyzing visitation on LinkedIn		
	B	Creating business accounts on YouTube, YouTube Advertising, YouTube Analytics,		
	C	Facebook Ads, Creating Facebook Ads, Ads Visibility.		
	Unit 5			
	A	E-mail marketing, E-mail marketing plan, E-mail marketing campaign analysis, Keeping up with conversions.		
	B	Digital Marketing Budgeting- resource planning- cost estimating- cost budgeting- cost control.		
	C	Recapitulation:- lessons learned- student satisfaction survey- closing.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	Digital Marketer.Pulizzi,J.(2014) Epic Content Marketing, Mcgraw Hill Education.		
	Other References	1. Ryan, D. (2014). Understanding Digital Marketing: Marketing Strategies for Engaging the Digital Generation, Kogan Page Limited. The Beginner's Guide to Digital Marketing (2015).		

Business Intelligence(BDA 321)

1	Course Code	BDA 321
2	**Course Title**	Business Intelligence
3	Credits	5
4	Contact Hours (L-T-P)	5-0-0
	Course Status	Elective
5	Course Objective	The objectives of this course are to provide UG students to Information Systems with comprehensive and in-depth knowledge of Business Intelligence (BI) principles and techniques by introducing the relationship between managerial and technological perspectives.
6	Course Outcomes	CO1: Identify the major frameworks of computerized decision support: decision support systems (DSS), data analytics and business intelligence (BI) (K1,K3,). CO2: Demonstrate the impact of business reporting, information visualization, and dashboards (K1,K2,K3). CO3: Explain data mining, neural networks, support vector machines, text analytics, text mining, sentiment analysis, web mining, web analytics, social analytics, social network analysis. (K3, K4, K5) CO4: Outline the definitions, concepts, and enabling technologies of big data analytics.(K2,K4) CO5: Apply big data technologies in business intelligence using geospatial data, location-based analytics, social networking, Web 2.0, reality mining, and cloud computing. (K3, K5) CO6: Describe how analytics are powering consumer applications and creating a new opportunity for entrepreneurship for analytics. (K3, K4, K5)
7	Course Description	This course will examine Business Intelligence (BI) technologies that help a company to improve its business. It discusses BI topics from both managerial and technical perspectives. Managerial perspectives discuss how BI affects the organization's decision-making process, while technical perspectives discuss the foundation for an intelligent system.
8	Outline syllabus:	
	Unit 1	
	A	Business Intelligence, Analytics and Decision Support: Overview.
	B	Foundations and Technologies for Decision Making.
	C	Making the Business Case for BI, Barriers to B, BI in the Marketplace.
	Unit 2	
	A	Data Warehouses and Star Schemas, Course Case Study.

	B	SQL Server Integration Services.		
	C	Data extraction and upload. Data integration models. Usage of metadata.		
	Unit 3			
	A	Deploying an enterprise wide reporting solution.		
	B	Building up multidimensional cubes. Non-relational and denormalized databases physical design.		
	C	Defining measures and dimensions. Introducing ad-hoc reporting.		
	Unit 4			
	A	KDD (Knowledge discovery from databases) process definition.		
	B	Types of interesting and potentially useful output patterns, common algorithms.		
	C	Use cases in different industries and knowledge domains.		
	Unit 5			
	A	A modern paradigm for strategic management. A key to long term success and business development. Common steps for implementing a Balanced Scorecards.		
	B	Simple toolkit for data engineer and business analyst: take the most of BI at your enterprise and make it simple and convincing.		
	C	Introducing Reporting, Authoring Simple Reports, Other Report Types and Reports Based on Analysis Services Cubes.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Dieby Eric Siegel and Thomas H. Davenport (2013) Precision Marketing: Maximizing Revenue Through Relevanceby Sandra Zoratti and Lee Gallagher (2012).		

	Other References	1. Data Science for Business: What you need to know about data mining and data-analytic thinkingby Foster Provost and Tom Fawcett (2013) 2. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analyticsby Bill Franks (2012) Digital Marketing Analytics: Making Sense of Consumer Data in a Digital Worldby Chuck Hemann and Ken Burbary (2013).

Internet of Things (BDA 324)

1	Course Code	BDA 324
2	**Course Title**	**Internet of Things** (IOT)
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Elective
5	Course Objective	Objective of this course to enable students to understand scope of Internet of things in Industry and students will get to understand the working of different techniques, protocol and algorithms which are commonly used in IoT applications.
6	Course Outcomes	CO1: Able to understanding the concept of Internet of things. (K1,K2, K3). CO2: Analyze various IoT devices and its technology. (K2, K3, K4). CO3: Know about the selection and use of appropriate IoT technologies & Gateways protocols for application development. (K1,K2, K3) CO4: Design and development of IoT application with the use of different cloud technology. (K3, K4, K5) CO5: develop and apply Advance method for Implementation of Internet of Things. (K5, K6). CO6: Enable students to understand scope of Internet of things in Industry. (K4, K6)
7	Course Description	This course aims to give a basic understanding about Internet of Things (IoT). Students will get to understand the working of different techniques, protocol and algorithms which are commonly used in IoT applications. During course students will also get expose the various architecture for developing IoT applications such OSI reference model etc.
8	Outline syllabus:	
	Unit 1	

	A	The Internet of Things Today, Time for Convergence, Towards the IoT Universe, Internet of Things Vision, IoT Strategic Research and Innovation Directions.
	B	IoT Applications, Future Internet Technologies, Infrastructure, Networks and Communication, Processes, Data Management, Security, Privacy & Trust.
	C	Device Level Energy Issues, IoT Related Standardization, Recommendations on Research Topics.
	Unit 2	
	A	M2M to IoT-The Vision-Introduction, From M2M to IoT, M2M towards IoT-the global context, A use case example, Differing Characteristics.
	B	A Market Perspective– Introduction, Some Definitions, M2M Value Chains, IoT Value Chains, An emerging industrial structure for IoT, The international driven global value chain and global information monopolies.
	C	M2M to IoT-An Architectural Overview– Building an architecture, Main design principles and needed capabilities, An IoT architecture outline, standards considerations.
	Unit 3	
	A	IoT Architecture-State of the Art –Introduction, State of the art.
	B	Architecture Reference Model-Introduction, Reference Model and architecture, IoT reference Model.
	C	IoT Reference Architecture-Introduction, Functional View, Information View, Deployment and Operational View, Other Relevant architectural views.
	Unit 4	
	A	IoT Applications for Value Creations, IoT applications for industry: Future Factory Concepts, Brownfield IoT, Smart Objects, Smart Applications, Four Aspects in your Business to Master.
	B	IoT Value Creation from Big Data and Serialization, IoT for Retailing Industry.
	C	IoT For Oil and Gas Industry, Opinions on IoT Application and Value for Industry, Home Management, eHealth.
	Unit 5	
	A	Overview of Governance, Privacy and Security Issues.
	B	Contribution from FP7 Projects, Security, Privacy and Trust in IoT-Data-Platforms for Smart Cities.
	C	First Steps Towards a Secure Platform, Smartie Approach. Data Aggregation for the IoT in Smart Cities, Security.
	Mode of examination	Theory

	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Samuel Greengard ,The Internet of Things" by Samuel Greengard. CunoPfister Author: CunoPfister ,Getting started with Internet of Things.		
	Other References	1. Vijay Madisetti and Arshdeep Bahga,"Internet of Things (A Hands-on-Approach)",1stEdition, VPT, 2014. 2. Francis daCosta,"Rethinking the Internet of Things: A Scalable Approach to Connecting Everything",1stEdition, Apress Publications, 2013. Cuno Pfister, Getting Started with the Internet of Things, O"Reilly Media, 2011, ISBN: 978-1-4493-9357-1		

Econometrics (BDA 327)

1	Course Code	BDA 327
2	**Course Title**	Econometrics
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Elective
5	Course Objective	Objective of this course is to introduce regression analysis to students so that they are able to understand its applications in different fields in economics.
6	Course Outcomes	CO1: Able to concise knowledge of basic regression analysis of economic data and interpret and critically evaluate outcomes of empirical analysis. (K1, K2, K3). CO2: Analyze the theoretical background for standard methods used in empirical analyses, like properties of least squares estimators and statistical testing of hypotheses. (K2, K3, K4). CO3: Able to apply modern computer programs in regression analyses of empirical data, including statistical testing to investigate whether the classical assumptions in regression analysis are satisfied. (K2, K3, K4). CO4: Design and development a real life model based on econometric methods. (K4, K5, K6) CO5: Develop and apply Advance method for Implementation of econometric techniques also various functions for economic analysis and future forecasting. (K5, K6). CO6: Enable students to make use of econometric models in your own academic work. (K4,K5)

7	Course Description	The purpose of this course is to give students a solid foundation in econometric techniques, various functions for economic analysis and future forecasting. Many of the methods introduced in this course are also useful in business, finance and many other disciplines.
8	Outline syllabus:	
	Unit 1	
	A	Introduction to econometrics. A review of least squares and maximum likelihood estimation methods of parameters in classical linear regression model and their properties.
	B	Generalized least squares estimation and prediction, construction of confidence regions.
	C	Tests of hypotheses, use of dummy variables and seasonal adjustment.
	Unit 2	
	A	Regression analysis under linear restrictions, restricted least squares estimation method and its properties.
	B	Problem of Multicollinearity, its implications and tools for handling the problem.
	C	Ridge regression. Hetroscedasticity, consequences and tests for it.
	Unit 3	
	A	Estimation procedures under hetroscedastic disturbances, Bartlett's test, Breusch Pagan test and Goldfelf Quandt test.
	B	Autocorrelation, sources and consequences.
	C	Autoregressive process tests for autocorrelation.
	Unit 4	
	A	Durbin Watson test. Asymptotic theory and regressors.
	B	Instrumental variable estimation, errors in variables.
	C	Simultaneous equations model, problem of identification, necessary and sufficient condition for the identifiability of parameters in a structural equation.
	Unit 5	
	A	Ordinary least squares, indirect least squares.
	B	Two-stage least square.
	C	Limited information maximum likelihood method.
	Mode of	Theory

	examination			
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	1. Gujrati, D.N.&Porter, D.C.(2017).Basic Econometrics, 6thEdition.McGraw Hill. Maddala, G.S. &Lahiri, K. (2010). Introduction to Econometrics, 4thEdition.Wiley.		
	Other References	1. Greene, W.H. (2012).Econometric Analysis, 7thEdition.Pearson. Studenmund, A.H. &Johnson, B.K. (2017).Using Econometrics:A Practical Guide, 7thEdition.Pearson.		

Artificial Intelligence (BDA 344)

1	Course Code	BDA 344
2	**Course Title**	Artificial Intelligence
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Elective
5	Course Objective	Objective of this course is to help students to learn the application of machine learning /A.I algorithms in the different fields of science, medicine, finance etc.
6	Course Outcomes	CO1: Understand basic concepts and applications of machine learning (K1, K3, K4). CO2: Able to predicate logic and transform the real life information in different representation. (K3, K6). CO3: Analyze the state space and its searching strategies. (K2, K5). CO4: Able to apply machine learning concepts and range of problems that can be handled by machine learning. (K2, K3, K4). CO5: Analyze problem specifications and derive appropriate solution techniques for them and also design and implement appropriate solutions for search problems and for planning problems. (K4, K6). CO6: Enable students to apply the machine learning concepts in real life problems. (K5,K6)

7	Course Description	The aim of this course is to introduce the fundamental concepts of Artificial Intelligence to students. The course will explain various important concepts such as searching techniques, Knowledge representation, Uncertainty and Natural Language Processing.
8	Outline syllabus:	
	Unit 1	
	A	Overview of AI problems, AI problems as NP, NP-Complete and NP-Hard problems.
	B	Strong and weak, neat and scruffy, symbolic and sub-symbolic, knowledge-based and data-driven AI.
	C	Search Strategies: Problem spaces (states, goals and operators), problem solving by search, Heuristics and informed search, Minmax Search, Alpha-beta pruning.
	Unit 2	
	A	Constraint satisfaction (backtracking and local search methods).
	B	Knowledge representation and reasoning: propositional and predicate logic, Resolution and theorem proving, Temporal and spatial reasoning.
	C	Probabilistic reasoning, Bayes theorem.
	Unit 3	
	A	Totally-ordered and partially-ordered Planning.
	B	Goal stack planning, Nonlinear planning, Hierarchical planning.
	C	Learning: Learning from example, Learning by advice, Explanation based learning, Learning in problem solving, Classification, Inductive learning, Naive Bayesian Classifier, decision trees.
	Unit 4	
	A	Natural Language Processing: Language models, n-grams, Vector space models, Bag of words, Text classification. Information retrieval.
	B	Agents: Definition of agents, Agent architectures (e.g., reactive, layered, cognitive).
	C	Multi-agent systems-Collaborating agents, Competitive agents, Swarm systems and biologically inspired models.
	Unit 5	
	A	Intelligent Systems: Representing and Using Domain Knowledge, Expert System Shells, Explanation, Knowledge Acquisition.
	B	Key Application Areas: Expert system, decision support systems.

	C	Speech and vision, Natural language processing, Information Retrieval, Semantic Web.		
	Mode of examination	Theory		
	Weightage Distribution	CA	MTE	ETE
		25	25	50
	Text book/s*	Artificial Intelligenceby Elaine Rich, Kevin Knight and Shivashankar B Nair, Tata McGraw Hill.		
	Other References	1. Introduction to Artificial Intelligence and Expert Systemsby Dan W. Patterson, Pearson Education. Artificial Intelligence: A Modern Approachby S. Russell and P. Norvig, Prentice Hall.		

Big Data Analytics (BDA 306)

1	Course Code	**BDA 306**
2	**Course Title**	**Big Data Analytics**
3	Credits	4
4	Contact Hours (L-T-P)	4-0-0
	Course Status	Compulsory
5	Course Objective	This course is aimed to provide an advance understanding to the big data overview, model building, clustering and advance analytics.
6	Course Outcomes	CO1: Discuss the concept big data analysis and data preparation (K3). CO2: Describe the concept model building, communicating results and check the basic data analysis.(K1, K2, K3). CO 3: Explain the concept how using R to look at data introduction to R , Analyzing and Exploring the Data, Statistics for Model Building and Evaluation Advanced Analytics. (K3, K4) CO 4: Illustrate the concept of K Means Clustering, association rules, linear regression, logistic regression, Naïve Bayesian Classifier and evaluate decision trees, time series analysis, text analysis. (K2, K3, K4). CO 5: Discuss the concept of unstructured data – Map Reduce and Hadoop, The Hadoop Ecosystem In-database Analytics and illustrate SQL Essentials, Advanced SQL and MADlib for In-database Analytics (K3, K4, K5). CO6: Demonstrate the understanding of the Endgame, or putting it all together: operationalizing an analytics project, creating the final deliverables, data visualization techniques, final lab exercise on big data analytics (K2,K4, K6).

7	Course Description	This course is given the deep knowledge of big data, model building, clustering and advance analytics.		
8	Outline syllabus :			
	Unit 1			
	A	State of the Practice in Analytics, the Data Scientist,		
	B	Big Data Analytics in Industry Verticals		
	C	Data Analytics Life cycle: Discovery, Data Preparation, Model Planning.		
	Unit 2			
	A	Model Building, Communicating Results, Operationalizing Review of Basic Data Analytic Methods Using R:		
	B	Using R to Look at Data Introduction to R ,		
	C	Analyzing and Exploring the Data, Statistics for Model Building and Evaluation Advanced Analytics.		
	Unit 3			
	A	K Means Clustering, Association Rules, Linear Regression,		
	B	Logistic Regression, Naïve Bayesian Classifier,		
	C	Decision Trees Time Series Analysis, Text Analysis.		
	Unit 4			
	A	Technologies and Tools : Analytics for Unstructured Data – Map Reduce and Hadoop ,		
	B	The Hadoop Ecosystem In-database Analytics – SQL Essentials		
	C	Advanced SQL and MADlib for In-database Analytics		
	Unit 5			
	A	The Endgame, or Putting it All Together: Operationalizing an Analytics Project,		
	B	Creating the Final Deliverables, Data Visualization Techniques,		
	C	Final Lab Exercise on Big Data Analytics.		
	Mode of examination	Theory		
	Weightage	CA	MTE	ETE

	Distribution	25	25	50
	Text book/s*	1. Big Data, Big Dupe, 2016 2. Tom White, "Hadoop: The Definitive Guide", 3rd edition, O'Reilly Media. Big Data (Black Book), Wiley Publications.		
	Other References	1. V. Prajapati, "Big Data Analytics with R and Hadoop", Packt Pub. 2. MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, Adam Shook and Donald Miner N. Dasgupta, Practical Big Data Analytics, Packt Publication Ltd.		

LAB COURSES
(Semester- I)

Mathematics Lab - 1 (BHM 151)

School: SBSR		**Batch: 2021- 24**
Program: B.Sc.(H).		**Academic Year: 2021- 22**
Branch: Data Science & Analytics		**Semester: I**
1	**Course Code**	**MSM 151**
2	**Course Title**	**Mathematics Lab - 1**
3	**Credits**	**2**
4	**Contact Hours (L-T-P)**	**0-0-4**
	Course Status	**Compulsory**
5	Course Objective	
6	Course Outcomes	**CO1:** The main objective of the course is to equip the student to plot the different graph and solve the different types of equations by plotting the graph using different computer software such as Mathematica /MATLAB /Maple /Scilab/Maxima etc. (K1,K2,K3) **CO2.** After completion of this course student would be able to know the convergence of sequences through plotting, verify Bolzano-Weierstrass theorem through plotting the sequence, Cauchy's root test by plotting *n th* roots and Ratio test by plotting the ratio of *n th* and (*n* + 1)th term. (K2,K3) **CO3.** Student would be able to plot Complex numbers and their representations, Operations like addition, substraction, Multiplication, Division, Modulus and Graphical representation of polar form. (K2,K3,K4) **CO4**: Student would be able to perform following task of matrix as Addition, Multiplication, Inverse, Transpose, Determinant, Rank, Eigenvectors, Eigenvalues, Characteristic equation and verification of the Cayley-Hamilton theorem, Solving the systems of linear equations. (K2,K3,K4) **CO5**: Develop program scripts and functions using the Mathematica /MATLAB /Maple /Scilab/Maxima development environment. (K3,K4,K5) **CO6**: Write the program for evaluates linear system of equations, ordinary differential equations in Mathematica /MATLAB /Maple /Scilab/Maxima. (K4,K5,K6).
7	Course Description	

8	Outline syllabus			
	Unit 1	**List of the practicals to be done using Mathematica /MATLAB /Maple /Scilab/Maxima etc.**		
	a, b, c	1. Plotting the graphs of the following functions: (i) ax (ii) [x] (greatest integer function) (iii) x 2n ; n ∈ N (iv) x 2n−1 ; n ∈ N (v) 1 ;n ∈ N X 2n−1 (vi) 1 ;n ∈ N X 2n		
	Unit 2	**List of the practicals to be done using Mathematica /MATLAB /Maple /Scilab/Maxima etc.**		
	a, b, c	(vii) √ax + b, |ax + b|, c ± |ax + b| (ix) |X|, sin (1 , x sin 1 , eX, e−X for x ≠ 0.) ( ) X X X (x) e ax+b, log(ax + b) , 1 , sin(ax + b), cos(ax + b) , |sin(ax+b)|,|cos(ax + b)|. ax+b Observe and discuss the effect of changes in the real constants a and b on the graphs.		
	Unit 3	**List of the practicals to be done using Mathematica /MATLAB /Maple /Scilab/Maxima etc.**		
	a, b, c	2. By plotting the graph find the solution of the equation x = ex, x2 + 1 = ex, 1 − x2 = ex, x = log10(x) , cos(etc)		
	Unit 4	**List of the practicals to be done using Mathematica /MATLAB /Maple /Scilab/Maxima etc.**		
	a, b, c	Plotting the graphs of polynomial of degree 2,3, 4 and 5, and their first and second derivatives.		
	Unit 5	**List of the practicals to be done using Mathematica /MATLAB /Maple /Scilab/Maxima etc.**		
	a, b, c	3. Sketching parametric curves, e.g., Trochoid, Cycloid, Epicycloid and Hypocycloid etc. 4. Tracing of conic in Cartesian coordinates. 5. Graph of circular and hyperbolic functions. Obtaining surface of revolution of curves.		
	Mode of examination	Jury+Practical+Viva		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	**MAT LAB Differential and Integral Calculus,** Apress901 Grayson Street Suite 204 Berkely, CAUnited States		
	Other References	SOLVING APPLIED MATHEMATICAL PROBLEMS WITH MATLAB, CRC Press.		

Statistics Lab- I (BDA151)

School: SBSR		**Batch: 2021-2024**
Program: BSc. (H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: 1**
1	Course number	**BDA-151**
2	Course Title	**Statistics Lab - 1**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory

5	Course Objective	To familiarize the student in introducing and exploring MS excel. To enable the student on how to approach for solving statistical problems using excel tools. To prepare the students to use excel in their project works. To provide a foundation in use of this MS office for real time applications.		
6	Course Outcomes	CO1: Understand the procedures,Analyzing and Visualizing Data with Excel. (K2) CO2: Discuss and develop the basic understanding of creating formulas and how cells are referenced by rows and columns within Excel. (K2, K5, K6) CO3: Discuss and construct table and graph of data with excel. (K2, K5, K6) CO4: Discuss and calculate basic statistical parameters (mean, measures of dispersion, correlation coefficient, indexes). (K2, K5, K6) CO5: Discuss and calculate /computeconditional Bayes Theorem .. (K2, K5, K6) CO6: Discuss, predict and estimate the variable by regression analysis with excel. (K2, K5, K6)		
7	Course Description	Enable students for using the computer program MS Excel, apply basic statistical techniques and methods for grouping, tabular and graphical display, analysis and interpretation of Statistical data.		
8	Outline syllabus			
	Unit 1	**Lab. Experiment 1:**		
		Exploring Data in Excel		
	Unit 2	**Lab. Experiment 2:**		
		Create Charts		
	Unit 3	**Lab. Experiment 3:**		
		Calculate Descriptive Statistics		
	Unit 4	**Lab. Experiment 4:**		
		Problems based on calculation of Moments, Measures of Skewness andKurtosis.		
	Unit 5	**Lab. Experiment 5:**		
		Computation of conditional probabilities based on Bayestheorem.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1. Sukhatme, P.V., Sukhatme, B.V. Sukhatme, S. (1984). Sampling Theories of Survey with Application, IOWA State University Press and Indian Society of Agricultural Statistics. 2. Goon A.M., Gupta M.K. and Dasgupta B. (2008): Fundamentals of Statistics, World Press.		
	Other References	1. Murthy, M.N. (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta. 2. Des Raj and Chandhok P. (1998): Sample Survey Theory, Narosa Publishing House.3. Cochran, W.G. (1984): Sampling Techniques (3rd Ed.), Wiley Eastern.		

C Programming Lab (BDA152)

School: SBSR		**Batch: 2021-2024**		
Program: BSc. (H)		**Academic Year: 2021-22**		
Branch: Data Science & Analytics		**Semester: 1**		
1	Course number	**BDA-152**		
2	Course Title	**C Programming Lab**		
3	Credits	2		
4	Contact Hours (L-T-P)	0-0-4		
	Course Status	Compulsory		
5	Course Objective	To understand and demonstrate how to solve logical and scientific problems using programming C.		
6	Course Outcomes	CO1: How to read, understand and trace the execution of programs written in C language. (K2,K3, K4). CO2: Apply c programming knowledge to convert algorithm into program in C(K2, K3, K4). CO3: Maximize the knowledge of Array and String concepts of C programming language (K1, K2). CO4: Demonstrate the concept of function, pointers and structure. (K3, K4, K5 (K2, K3,K4). CO5: Develop the uses of computers in engineering industry. (K4,K5,K6) CO6: Discuss about the more advanced features of the C language (K3,K4,K6).		
7	Course Description	To understand and demonstrate how to solve logical and scientific problems using programming C.		
8	Outline syllabus			
	Unit 1	**Lab. Experiment 1:**		
		Write a c program to swap two numbers with temporary variable. Write a c program to swap two numbers without temporary variable.		
	Unit 2	**Lab. Experiment 2:**		
		Write a c Program to Add Two Integers. Write a program to check given year is leap year.		
	Unit 3	**Lab. Experiment 3:**		
		Write a c program to calculate the average using arrays. Write a c program to find the largest element of the array.		
	Unit 4	**Lab. Experiment 4:**		
		Write a function to calculate factorial of a number. Write a c program to store information of a student using structure		
	Unit 5	**Lab. Experiment 5:**		
		Write a c program to store information of a student using union. Write a c program to swap two values using pointers.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks

	Text book/s*	1. Sukhatme, P.V., Sukhatme, B.V. Sukhatme, S. (1984). Sampling Theories of Survey with Application, IOWA State University Press and Indian Society of Agricultural Statistics. 2. Goon A.M., Gupta M.K. and Dasgupta B. (2008): Fundamentals of Statistics, World Press.
	Other References	1. Murthy, M.N. (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta. 2. Des Raj and Chandhok P. (1998): Sample Survey Theory, Narosa Publishing House.3. Cochran, W.G. (1984): Sampling Techniques (3rd Ed.), Wiley Eastern.

Statistics Lab- II (BDA 153)

School: SBSR		**Batch: 2021- 2024**
Program: B.Sc.(H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: II**
1	Course Code	**BDA 153**
2	Course Title	**Statistics Lab II**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	Introduce basic statistical concepts, logics and analytical tools MS excel. Provide students with a general understanding of descriptive and inferential statistics and the opportunity to apply them to examine business and economic data. Equip students with the skills to apply statistical concepts and analytical tools to analyze and handle real-world business issues. Train students for presenting and exchanging statistical findings and views.
6	Course Outcomes	CO1: Understand the procedures, Analyzing and Visualizing Data with Excel. (K2) CO2: Discuss and develop the basic understanding of SRSWOR, estimate mean, standard error, the sample size. (K2, K5, K6) CO3: Discuss and calculate Ratio and Regression estimation. (K2,K5, K6) CO4: Discuss and calculate systematic with stratified sampling and SRS in the presence of a linear trend. (K2, K5, K6) CO5: Discuss and calculate correlationbetween two variables with excel. (K2, K5, K6) CO6: Discuss, predict and estimate the variable by regression analysis with excel. (K2, K5, K6)
7	Course Description	This course provides students with basic statistical concepts and analytical tools, and the opportunity to apply them to analyze real-world problem data. Main topics include sampling methods, descriptive statistics, probability & probability distributions, sampling distributions andconfidence interval estimation, hypothesis testing, simple linear regression and correlation, time series analysis and applications in quality and production management.
8	Outline syllabus	
	Unit 1	**Lab. Experiment 1:**
		To select a SRS with and without replacement. For a population of size 5, estimate population mean, population mean square and population variance.
	Unit 2	**Lab. Experiment 2:**
		Enumerate all possible samples of size 2 by WR and WOR and establish all properties relative to SRS.
	Unit 3	**Lab. Experiment 3:**

		For SRSWOR, estimate mean, standard error, the sample size. Stratified Sampling: allocation of sample to strata by proportional and Neyman's methods. Compare the efficiencies of above two methods relative toSRS.		
	Unit 4	**Lab. Experiment 4:**		
		Estimation of gain in precision in stratified sampling. Comparison of systematic with stratified sampling and SRS in the presence of a linear trend. Ratio and Regression estimation: Calculate the population mean or total of the population.		
	Unit 5	**Lab. Experiment 5:**		
		Calculate mean squares. Compare the efficiencies of ratio and regression estimators relative to SRS. Cluster sampling: estimation of mean or total, variance of the estimate, estimate of intra-class correlation coefficient, efficiency as compared to SRS. Practical based on two stage sampling. Practical based on multistage sampling		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1. Sukhatme, P.V., Sukhatme, B.V. Sukhatme, S. (1984). Sampling Theories of Survey with Application, IOWA State University Press and Indian Society of Agricultural Statistics. 2. Goon A.M., Gupta M.K. and Dasgupta B. (2008): Fundamentals of Statistics, World Press.		
	Other References	1. Murthy, M.N. (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta. 2. Des Raj and Chandhok P. (1998): Sample Survey Theory, Narosa Publishing House.3. Cochran, W.G. (1984): Sampling Techniques (3rd Ed.), Wiley Eastern.		

R Programming Lab (BDA 154)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: **Data Science & Analytics**		**Semester: II**
1	Course number	**BDA 154**
2	Course Title	**R Programming Lab**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	To familiarise students with basics programming in R, and its applications in data analysis.
6	Course Outcomes	CO1: Explain the R Windows Environment and describe various data types. (K1, K2, K3, K4). CO2: Explain and describe Outliers, Combining Datasets. (K2, K3) CO3: Explain and illustrate R Functions and loops, Summary Statistics –Summarizing data with R. (K2,K3, K4). CO4: Discuss how to load data, plot a graph and illustrate different types of graphs with

		graphical summaries of data. (K2, K3, K4) CO5: Discuss how to generate automated reports giving detailed basic statistics using R and evaluate measures of central tendency and dispersion. Covariance, correlation and lines of regression in R.(K2, K3, K4) CO6: Explain fitting of polynomials and exponential curves and illustrate Normal probability plot. (K 4, K6)		
7	Course Description	This course is an introduce basics programming in R, and its applications in data analysis.		
8	**Outline syllabus:**	**R Programming Lab**		
	Unit 1	**Lab. Experiment 1**		
		Problem based on data types and data Structure		
	Unit 2	**Lab. Experiment 2**		
		Problem based on loops and Function in R		
	Unit 3	**Lab. Experiment 3**		
		Problem Based on Mathematics in R		
	Unit 4	**Lab. Experiment 4**		
		Problem based on Graphs		
	Unit 5	**Lab. Experiment 5**		
		Problem based on Fitting of Polynomial and Distributions		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1. Gardener, M (2012): Beginning R: The Statistical Programming Language, Wiley Publications. 2. Braun W J, Murdoch D J (2007): A First Course in Statistical Programming with R. Cambridge University Press. New York		
	Other References	1. Crawley, M.J. (2015): Statistics: An Introduction Using R, 2nd Edition. Wiley. 2. Crawley, M.J. (2012): The R Book, 2nd Edition. Wiley.		

Statistics Lab-III (BDA 251)

School: SBSR		**Batch: 2021-2024**		
Program: B.Sc. (H)		**Academic Year: 2021-22**		
Branch: Data Science & Analytics		**Semester: III**		
1	Course number	**BDA 251**		
2	Course Title	**Statistics Lab-III**		
3	Credits	2		
4	Contact Hours (L-T-P)	0-0-4		
	Course Status	Compulsory		
5	Course Objective	To introduce concepts of statistical analysis of descriptive statistics, logics and analytical tools, analyze and communicate quantitative data verbally, graphically, symbolically and numerically. To make students familiar with the concept of Probability and Statistics and hypothesis.		
6	Course Outcomes	CO1: Describe the process Statistical analysis of descriptive statistics, principle of least square, lines of regression, simple linear regression and evaluate multiple linear regression, coefficient of multiple determination.(K2, K5) CO2: Describe the process of fitting of polynomials and exponential curves. (K2) CO3: Explain the criteria for obtaining a good estimator. (K2, K3) CO4: Calculate and interpret the point estimation, confidence interval, construction of confidence intervals using pivotal, shortest expected length confidence interval. (K2, K3) CO5: Understand the null hypothesis, alternative hypothesis, type I error, type II error, level of significance, p-value and power of test, develop the ability to use one sample t-test, two-sample t-test, paired-sample t-test. Tests for variance based on normal distribution – one sample and two-sample problem. (K2, K5) CO6: Develop the skills to interpret the results of statistical analysis by using Z-test, F-test, Chi-square test for goodness of fit. One-way and Two-way analysis of variance (ANOVA) techniques. (K2, K5)		
7	Course Description	This is an advances course in statistics. Students are introduced to the f concepts involved in using sample data to make inferences about populations. Included are the study of measures of central tendency and dispersion, finite probability, statistical inferences from large and small samples, linear regression, and correlation and hypothesis.		
8	**Outline syllabus:**	**Statistics Lab-III**		
	Unit 1	**Lab. Experiment 1**		
		Problem based on principle of least square, Simple linear regression, Multiple linear regression		
	Unit 2	**Lab. Experiment 2**		
		Problem based on obtaining a good estimator: Unbiasedness, Consistency, Efficiency, Sufficiency.		
	Unit 3	**Lab. Experiment 3**		
		Problem based on Point and Interval Estimation.		
	Unit 4	**Lab. Experiment 4**		
		Problem based on Hypothesis Testing.		
	Unit 5	**Lab. Experiment 5**		
		Problem based on One-way and Two-way analysis of variance (ANOVA) techniques.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE

		25 Marks	25 Marks	50 Marks
	Text book/s*	Gupta,S.C and Kapoor,V.K, "Fundamental of Mathematical Statistics".		
	Other References	Daniel,WayneW.,"Biostatistics": Basic concept and Methodology for Health Science. 1. Grewal,B.S, "Higher Engineering Mathematics".		

Data Structure Lab (BDA 252)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: **Data Science & Analytics**		**Semester: III**
1	Course number	**BDA-252**
2	Course Title	**Data Structure Lab**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	To make students familiar with the data structure & algorithms. The concept of data organizations, data structure operations; analysis of an algorithm; Stacks and Queues; Linked Lists; Sorting and Hashing; Graph.
6	Course Outcomes	CO1: Explain and illustrate the concepts basic terminologies: elementary data organizations, data structure operations: insertion, deletion, traversal etc. (K2, K3, K4) CO2: Describe the analysis of an algorithm, asymptotic; notations, time-space trade off. (K1, K2, K3) CO3: Describe Linear Search and Binary Search Techniques and explain their complexity analysis. (K2, K3, K4) CO4: Describe ADT Stack and its operations: Algorithms and their complexity analysis, Applications of Stacks; Types of Queue; Algorithms and their analysis. (K2, K3, K4) CO5: Describe the Singly linked lists; trees; algorithms and analysis. (K2, K3, K6) CO6: Describe and analyze the basic concepts of Sorting and Hashing; Graphs. (K1,K2, K4)
7	Course Description	This course an introduce data structure & algorithms. The concept of data organizations, data structure operations; analysis of an algorithm; Stacks and Queues; Linked Lists; Sorting and Hashing; Graph.
8	**Outline syllabus:**	**Data Structure Lab**
	Unit 1	Problem based on uses functions to perform the following operations on singly linked list i) Creation ii) Insertion iii) Deletion iv) Traversal. Problem based on uses functions to perform the following operations on doubly linked list i) Creation ii) Insertion iii) Deletion iv) Traversal.
	Unit 2	Problem based on uses functions to perform the following operations on circular linked List i) Creation ii) Insertion iii) Deletion iv) Traversal. Problem based on implement stack (its operations) using i) Arrays ii) Linked list(Pointers).
	Unit 3	Problem based on implement Queue (its operations) using i) Arrays ii) Linked list (Pointers). Problem based on implement Circular Queue using arrays. Problem based on both recursive and non recursive functions to perform the following searching operations for a Key value in a given list of integers: a) Linear search b) Binary search.
	Unit 4	Problem based on implements the following sorting i) Bubble sort ii) Selection sort iii) Quick sort. Problem based on implements the following i) Insertion sort ii) Merge sort iii) Heap sort. Problem based on implement all the functions of a dictionary (ADT) using Linked List.

	Unit 5	Problem based on to perform the following operations: a) Insert an element into a binary search tree. b) Delete an element from a binary search tree. c) Search for a key element in a binary search tree. Problem based on to implement the tree traversal methods. Problem based on to perform the following operations: a) Insert an element into a AVL tree. b) Delete an element from a AVL tree. c) Search for a key element in a AVL tree.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1.Fundamentals of Data Structures", Illustrated Edition by Ellis Horowitz, SartajSahni, Computer Science Press.		
	Other References	1.Algorithms, Data Structures, and Problem Solving with C++", Illustrated Edition by Mark Allen Weiss, Addison-Wesley Publishing Company. 2. How to Solve it by Computer", 2nd Impression by R. G. Dromey, Pearson Education.		

Data Cleaning Lab (BDA 253)

School: SBSR		**Batch: 2021-2024**		
Program: B.Sc. (H)		**Academic Year: 2021-22**		
Branch: **Data Science & Analytics**		**Semester: IV**		
1	Course number	**BDA 253**		
2	Course Title	**Data Cleaning Lab**		
3	Credits	2		
4	Contact Hours(L-T-P)	0-0-4		
	Course Status	Compulsory		
5	Course Objective	To make students familiar with the concepts of preparing your data; Working with dates and times, Data Cleaning, Data Structure, Cleaning Text Data.		
6	Course Outcomes	CO1: Describe preparing data: Rearranging and removing variables, Renaming variables, Variable classes, Calculating new numeric variables and explain how to Dividing a continuous variable into categories, Working with factor variables. (K1, K3) CO2: Discuss how to working with dates and times, adding and removing observations and explain about removing duplicate observations, selecting a subset of the data, selecting a random sample from a dataset, sorting a dataset. (K2, K3, K4) CO3: Explain the data cleaning and technical representation of data. (K2,K3, K4) CO4: Discuss about the data structure. (K2, K6) CO5: Describe Character Normalization, Encoding Conversion and Unicode Normalization, Character Conversion and Transliteration. (K1, K2) CO6: Discuss and evaluate Generating Regular Expressions in R, Common String Processing Tasks in R, Approximate Text Matching, String Metrics, String Metrics and Approximate Text Matching in R.		
7	Course Description	This course is an introduces preparing your data; Working with dates and times, Data Cleaning, Data Structure, Cleaning Text Data.		
8	**Outline syllabus:**	**Data Cleaning Lab**		
	Unit 1	**Lab. Experiment 1**		
		Problem based on data collection and source of error.		
	Unit 2	**Lab. Experiment 2**		
		Problem based on screening, diagnosis, treatment of data.		
	Unit 3	**Lab. Experiment 3**		
		Problem based on missing value and record value.		
	Unit 4	**Lab. Experiment 4**		
		Problem based on quality control procedure, data Integration.		
	Unit 5	**Lab. Experiment 5**		
		Problem based on tools and technique for data cleaning.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	2. Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work by Q. Ethan McCallum 3. Best Practices in Data Cleaning: A Complete Guide to Everything You Need		

		to Do Before and After Collecting Your Data by Jason W Osborne
	Other References	2) Data Wrangling with Python by Jacqueline Kazil 3) Principles of Data Wrangling: Practical Techniques for Data Preparation by Tye Rattenbury

DBMS Lab (BDA 254)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: IV**
1	Course number	**BDA 254**
2	Course Title	**DBMS Lab**
3	Credits	2
4	Contact Hours(L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	To make students familiar with the basic concepts of Databases and Transactions and Data Models, Database Design ,ER-Diagram and Unified Modeling Language, Relational Algebra and Calculus, Constraints, Views and SQL, Transaction management and Concurrency control.
6	Course Outcomes	CO1: Discuss the basics of Databases and Transactions and Data Models. (K1, K2, K3) CO2: Discuss about Database Design ,ER-Diagram and Unified Modeling Language. (K1, K3) CO3: Explainrelational algebra and calculus, describe Domain relational Calculus, calculus vs algebra, computational capabilities. (K3, K4) CO4: Explain and illustrate Constraints, Views and SQL. (K3,K6) CO5: Evaluate different types of transaction management. (K4,K5) CO6: Explain concurrency control, time stamping methods, optimistic methods, database recovery management. (K2, K4, K5)
7	Course Description	This course is an introduce the basic concepts of Databases and Transactions and Data Models, Database Design ,ER-Diagram and Unified Modeling Language, Relational Algebra and Calculus, Constraints, Views and SQL, Transaction management and Concurrency control.
8	**Outline syllabus: DBMS Lab**	
	Unit 1	**Lab. Experiment 1**
		Problem based on E-R diagram and convert entities and relationships to relation table for a given scenario. Write relational algebra queries for a given set of relations.
	Unit 2	**Lab. Experiment 2**
		Problem based on Viewing all databases, Creating a Database, Viewing all Tables in a Database, Creating Tables (With and Without Constraints), Inserting/Updating/Deleting Records in a Table, Saving (Commit) and Undoing (rollback)
	Unit 3	**Lab. Experiment 3**
		Problem based on Altering a Table, Dropping/Truncating/Renaming Tables, Backing up / Restoring a Database
	Unit 4	**Lab. Experiment 4**
		Problem based on set of relation schemes, create tables and perform the following Simple Queries, Simple Queries with Aggregate functions, Queries with Aggregate functions (group by and having clause), Queries involving- Date Functions, String Functions , Math Functions Join Queries- Inner Join, Outer Join Subqueries- With IN clause, With EXISTS clause

	Unit 5	**Lab. Experiment 5**		
		Problem based on set of relation tables perform the following a. Creating Views (with and without check option), Dropping views, Selecting from a view		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1."Database System Concepts", 6th Edition by Abraham Silberschatz, Henry F. Korth, S. Sudarshan, McGraw-Hill		
	Other References	1 "Principles of Database and Knowledge – Base Systems", Vol 1 by J. D. Ullman,Computer science Press. 2 "Fundamentals of Database Systems", 5th Edition by R. Elmasri and S. Navathe, Pearson Education 3 "Foundations of Databases", Reprint by Serge Abiteboul, Richard Hull, VictorVianu,Addison-Wesley		

Operating System Lab (BDA 255)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: IV**
1	Course number	**BDA 255**
2	Course Title	**Operating System Lab**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	To familiarise students with basic concepts of Operating Systems, Process Management Processes, Interprocess Communication Race Conditions, Deadlocks, Memory Management, I/O Management Principles of I/O Hardware, File Management.
6	Course Outcomes	CO1: Describe the concept of operating systems and process management processes. (K2) CO2: Explain the concept of interprocess communication race conditions, deadlocks (K2, K4) CO3: Recognize and decide basic memory management and virtual memory. (K1, K6) CO4: Define and discriminate I/O Management Principles of I/O Hardware and I/O Software. (K1, K6) CO5: Discuss about file management and directory implementation efficiency & performance. (K1,K2,K5) CO6:Explain Unix/Linux operating system and development of Unix/Linux. (K2,K4, K6)
7	Course Description	This course will cover basic concepts of Operating Systems, Process Management Processes, Inter process Communication Race Conditions, Deadlocks, Memory Management, I/O Management Principles of I/O Hardware, File Management.
8	**Outline syllabus:**	**Operating System Lab**
	Unit 1	**Lab. Experiment 1**
		Working with Linux

	Unit 2	**Lab. Experiment 2**		
		Basic concepts in OS: Through Linux tools for Process, Memory and I/O Management		
	Unit 3	**Lab. Experiment 3**		
		Introducing xv6 : Booting xv6		
	Unit 4	**Lab. Experiment 4**		
		Tracing and Creating System Calls in xv6		
	Unit 5	**Lab. Experiment 5**		
		Add history command to the kernel and obtain process statistics		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1."Operating System Concepts" by Avi Silberschatz and Peter Galvin 2."Operating Systems: Internals and Design Principles" by William Stallings		
	Other References	1."Operating Systems: A Concept-Based Approach" by D M Dhamdhere 2."Operating System: A Design-oriented Approach" by Charles Crowley		

Statistical Simulation Lab (BDA 351)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: VI**
1	Course number	**BDA-351**
2	Course Title	**Statistical Simulation Lab**
3	Credits	2
4	Contact Hours(L-T-P)	0-0-4
	Course Status	Compulsory
5	Course Objective	The learning objectives include: Concept of simulation and simulation modeling, Generation of Pseudo random number generators as well as from standard statistical distributions, Monte-Carlo simulation technique and application of simulation techniques.
6	Course Outcomes	CO1: Recognize the concepts of probability and statistics that are relevant to modeling and simulation (K2, K3). CO2: How to generate random numbers by the different methods (K2, K3). CO3: Design and implement Bootstrapping; jackknife resampling(K3, K4). CO4: Be able to evaluate and interpret the Markov-Chain Monte Carlo (MCMC) simulations (K3, K4). CO5: Hands-on experience in using simulation software packages/structured programming languages (K3, K4, K5) CO6: How simulation may be used to understand the behavior of real world systems by utilizing mathematical models with an emphasis on simulation (K4, K6).

7	Course Description	The course topics will include a review of concepts from probability and statistics that are relevant to modeling and simulation, algorithms for random-variable sampling, modeling and analysis of basic queueing systems, variance-reduction techniques, statistical-validation techniques, Independent Monte Carlo (IMC) and Markov-Chain Monte Carlo (MCMC) simulations, and discrete-event modeling and simulation.		
8	**Outline syllabus:**	**Statistical Simulation Lab**		
	Unit 1	**Lab. Experiment 1**		
		Review of probability, Independent Monte Carlo method		
	Unit 2	**Lab. Experiment 2**		
		Sampling discrete and continuous random variables, Queueing theory		
	Unit 3	**Lab. Experiment 3**		
		Discrete-event simulation, Variance reduction techniques		
	Unit 4	**Lab. Experiment 4**		
		Markov-Chain Monte Carlo methods, Simulated annealing		
	Unit 5	**Lab. Experiment 5**		
		Statistical analysis of simulated data, Statistical validation techniques		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1. Fishman, G.S. (1996) Mante Carlo: Concepts, Algorithms and Applications. (Springer). 2. Rubinstein, R.Y. (1981); Simulation and the Monte Carlo Method. (Wiley). 3. Ripley, B.D. (1987) Stochastic Simulations (Wiley).		
	Other References	1. Ross, S. M. (2002) Simulation (Third Edition) (Academic). 2. Efron,B. and Tibshirani. R.J. (1993); An introduction to the Bootstrap. 3. Davison, A.C. and Hinkley,D.V. (1997) Bootstrap methods and their applications (Chapman and Hall). 4. Sho.J and Tu,D (1995); The Jackknife and the Bootstrap. Springer Verlag.		

Data Visualization Lab (BDA 352)

School: SBSR		**Batch: 2021-2024**
Program: B.Sc. (H)		**Academic Year: 2021-22**
Branch: Data Science & Analytics		**Semester: VI**
1	Course number	**BDA 352**
2	Course Title	**Data Visualization Lab**
3	Credits	2
4	Contact Hours (L-T-P)	0-0-4
	Course Status	Compulsory

5	Course Objective	To introduce concepts of data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. To make students familiar with the concept of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.		
6	Course Outcomes	CO1: Describe the process Statistical analysis of descriptive statistics in graphically .(K2, K5) CO2: Describe the process of fitting of curves. (K2, K3) CO3: Explain the criteria for obtaining a good graph. (K2, K3) CO4: Describe data very easily. (K2, K3) CO5: Tests for variance based on normal distribution – one sample and two-sample problem. (K2, K5) CO6: Develop different- different graphs. (K2, K5)		
7	Course Description	This is an advances course in statistics. Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets.		
8	**Outline syllabus:**	**Data Visualization Lab**		
	Unit 1	**Lab. Experiment 1**		
		Introduction to Tableau and Aggregation Methods in Tableau.		
	Unit 2	**Lab. Experiment 2**		
		Visual Encodings and Basic Dashboards in Tableau.		
	Unit 3	**Lab. Experiment 3**		
		Hierarchical and Topographical Data Visualizations in Tableau.		
	Unit 4	**Lab. Experiment 4**		
		Time Series Data Visualization in Tubule		
	Unit 5	**Lab. Experiment 5**		
		Dashboards, Actions and Story Telling in Tableau.		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25 Marks	25 Marks	50 Marks
	Text book/s*	1. Tableau tutorial for beginners		
	Other References	Tableau Tutorial for Beginners - Learn Tableau Step By Step		

Vocational Courses

(Semester-I)

1	Course Code	BHM 152
2	**Course Title**	**Basic Statistical Analysis Using Excel**
3	Credits	3
4	Contact Hours (L-T-P)	0-0-5
	Course Status	Compulsory
5	Course Objective	Learn to use Effective and Efficient Statistical Analysis with Microsoft Excel to make Business and Economic decisions.
6	Course Outcomes	CO1: Identify the major frameworks of Excel to take large data sets and perform analysis to assist in business decision making. (K1,K2, K3) CO2: Demonstrate the impact of business reporting, take raw data and create descriptive tabular and visual reports to help in business decision making. (K2,K3) CO3: Explain, calculate and apply probability rules to assist in business decision making. (K2. K3, K4) CO4: Take samples and use sampling distributions to help business make decisions. (K2, K3, K4) CO5: Make inferences based on sample data using methods such as interval estimation and hypothesis testing. (K3,K4,K5) CO6: Describe how to perform regression analysis on x and y data sets to assist in business decision making. (K5,K6)
7	Course Description	The objective of the class is to have the student leave the class with an introductory knowledge of statistics and how to use the tool Excel to take the raw data they are given and convert it into useful information to help make decisions. At its essence, this class teaches how to do Data Analysis (convert raw data into useful information) using Statistical Methods and Excel.
8	Outline syllabus :	
	Unit 1	
	A	Formulas & Functions, Cell References, Number Format as Façade

	B	Effective and efficient spreadsheet design, including Excel's Golden Rule.		
	C	Data Analysis features, such as PivotTables, Sorting, Filtering and Importing Data. Charting in Excel.		
	Unit 2			
	A	What is Statistics (Descriptive and Inferential).		
	B	Descriptive Statistics: Tabular & Graphical Presentation.		
	C	Descriptive Statistics: Numerical Measures.		
	Unit 3			
	A	Introduction to Probability.		
	B	Discrete Probability Distributions.		
	C	Continuous Probability Distributions.		
	Unit 4			
	A	Sampling and Sampling Distributions.		
	B	Interval Estimation.		
	C	Hypothesis Testing (1 and 2 means, 1 and 2 proportions).		
	Unit 5			
	A	ANOVA.		
	B	Test of Independence.		
	C	Simple Linear Regression.		
	Mode of examination	Practical		
	Weightage	CA	Viva	ETE

	Distribution	25 Marks	25 Marks	50 Marks
	Text book/s*	1. Gupta. S.C. & Kapoor,V.K. (2002) . Fundamentals of Mathematical Statistics , Sultan Chand & Sons Pvt. Ltd. New Delhi.		
	Other References	1. Goon A.M. Gupta. A.K. & Das Gupta, B (1987). Fundamentals of Statistics, Vol.2, World Press Pvt. Ltd., Calcutta. 2. Hogg, R.V. and Craig, A.T. (2002). Introduction to Mathematical Statistics. Pearson Education India. Statistical Analysis with Excel for Dummies, 4th Edition 4th Edition by Joseph Schmuller.		

Statistical Thinking with R/Python/SPSS (BHM 153)

1	Course Code	**BHM 153**
2	**Course Title**	**Statistical Thinking with R/Python/SPSS**
3	Credits	3
4	Contact Hours (L-T-P)	0-0-5
	Course Status	Compulsory
5	Course Objective	Objective of this course to introduces how to present, analyze and interpret data using the statistical analysis software package SPSS /R/Python.
6	Course Outcomes	CO1: Identify the major frameworks of **R/Python/SPSS** to take large data sets and perform analysis to assist in business decision making. (K1,K2,K3) CO2: Demonstrate the impact of business reporting, take raw data and create descriptive tabular and visual reports to help in business decision making. (K1,K2,K3) CO3: Explain, calculate and apply probability rules to assist in business decision making. (K2,K3,K4) CO4: Take samples and use sampling distributions to help business make decisions. (K2,K3,K4) CO5: Make inferences based on sample data using methods such as interval estimation and hypothesis testing. (K3,K4,K5) CO6: Describe how to perform regression analysis on x and y data sets to assist in business decision making. (K2,K3,K6)
7	Course Description	This course introduces how to present, analyze and interpret data using the statistical analysis software package SPSS /R/Python. This course will help you develop those skills using SPSS, which is a statistical package widely used in business, industry,

		government, commerce and the education and health sectors.		
8	Outline syllabus :			
	Unit 1			
	A	Introduction to with R/Python/SPSS.		
	B	**Entering and Editing Data** with R/Python/SPSS.		
	C	Data Analysis features, such as PivotTables, Sorting, Filtering and Importing Data with R/Python/SPSS.		
	Unit 2			
	A	What is Statistics (Descriptive and Inferential) with R/Python/SPSS.		
	B	Descriptive Statistics: Tabular & Graphical Presentation with R/Python/SPSS.		
	C	Descriptive Statistics: Numerical Measures with R/Python/SPSS.		
	Unit 3			
	A	Introduction to Probability with R/Python/SPSS.		
	B	Discrete Probability Distributions with R/Python/SPSS.		
	C	Continuous Probability Distributions with R/Python/SPSS.		
	Unit 4			
	A	Sampling and Sampling Distributions with R/Python/SPSS.		
	B	Interval Estimation with R/Python/SPSS.		
	C	Hypothesis Testing (1 and 2 means, 1 and 2 proportions) with R/Python/SPSS.		
	Unit 5			
	A	ANOVA with R/Python/SPSS.		
	B	Test of Independence with R/Python/SPSS.		
	C	Simple Linear Regression with R/Python/SPSS.		
	Mode of examination	Practical		
	Weightage	CA	Viva	ETE

	Distribution	25 Marks	25 Marks	50 Marks
	Text book/s*	1. Gupta. S.C. & Kapoor,V.K. (2002) . Fundamentals of Mathematical Statistics , Sultan Chand & Sons Pvt. Ltd. New Delhi. 2. An Introduction to R, Notes on R: A Programming Environment for Data Analysis and Graphics. W. N. Venables, D.M. Smith and the R Development Core Team. Version 3.0.1 (2013-05-16). Field, A. P. 2009. Discovering Statistics using SPSS (Introducing Statistical Method). Oriental Press, Chennai, India.		
	Other References	1. Goon A.M. Gupta. A.K. & Das Gupta, B (1987). Fundamentals of Statistics, Vol.2, World Press Pvt. Ltd., Calcutta. 2. Dunlop, Dorothy D., and Ajit C. Tamhane. Statistics and data analysis: from elementary to intermediate. Prentice Hall, 2000. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".		

Semester- III

Prediction and Forecasting Management

1	Course Code	**BHM 252**
2	**Course Title**	**Prediction and Forecasting Management**
3	Credits	3
4	Contact Hours (L-T-P)	0-0-5
	Course Status	Compulsory
5	Course Objective	After completing this course, students are expected to grasp good qualitative and quantitative skills of developing forecasts using averaging and regression-based models and evaluating the forecasts for accuracy and parsimony. More importantly, students are expected to provide analytical solution to a business forecasting problem using appropriately selected model and data and discover meaningful business knowledge from the solution.
6	Course Outcomes	CO1: Describe how to perform regression analysis on x and y data sets to assist in business decision making. (K1,K2,K3) CO2: Understand the fundamental advantage and necessity of forecasting in various situations. (K2,K3) CO3: Know how to choose an appropriate forecasting method in a particular

		environment. (K2,K3,K4) CO4: Demonstrate knowledge and understanding of index numbers theory and methods and be able to provide practical solutions to general aggregation problems. (K2,K3,K4) CO5: Know how to apply various forecasting methods, which includes obtaining the relevant data and carrying out the necessary computation. (K3,K4,K5) CO6: Improve forecast with better statistical models based on statistical analysis. (K3,K4,K6)
7	Course Description	This course presents a set of topics in developing analytical methodologies that make prediction and forecasting about future events of interest to individual business and industry in general. Students are introduced to managerial techniques and analytical models that reveal valuable relationships in economic and business data for supporting short-term and long-term planning. Students will learn how to build the models, how to interpret the predictions and forecasts produced from the models, and how to evaluate the reliability of the model results.
8	Outline syllabus :	
	Unit 1	
	A	Simple Linear Regression Forecasting Model.
	B	Multiple Linear Regression Forecasting Model .
	C	Applications in Business.
	Unit 2	
	A	Introduction to Business Forecasting.
	B	Qualitative vs. Quantitative Methods.
	C	Characteristics of Time Series Data.
	Unit 3	
	A	Moving Averaging Models for Trend Identification.
	B	Naive Average Forecasting, Moving Average Forecasting Model.
	C	Smoothing Forecasting Model, Applications in Business.
	Unit 4	
	A	Time Series Models for Observation Forecast.
	B	Auto regressive Forecasting (AR) Model, Auto regressive Moving Average (ARMA) Model.

	C	Auto regressive Integrated Moving Average (ARIMA) Model, Dealing with periodic fluctuation.		
	Unit 5			
	A	Introduction, Basic Problems in the construction of Index Numbers.		
	B	Construction of Index Numbers, Uses and Limitations of Index Numbers.		
	C	Chain Index, Base Shifting, Splicing and Deflating, Cost of Living Index.		
	Mode of examination	Practical		
	Weightage	CA	Viva	ETE
	Distribution	25	25	50
	Text book/s*	**1.** Gupta, S.C., Kapoor, V. K. (2007): Fundamentals of Applied Statistics, 4thEdition, Sultan Chand & Sons**.** Gun, A.M., Gupta, M.K. and Dasgupta, B. (2008): Fundamentals of Statistics, Vol. II, 9th Edition, World Press.		
	Other References	1. "Forecasting and Time Series", 4thEdition, by Bowerman and O'Connell, Duxbury. 2. Croxton, Fredrick E., Cowden, Dudley J. and Klein, S. (1973): Applied General Statistics, 3rd Edition. Prentice Hall of India Pvt. Ltd. 3. Karmel, P.H. and Polasek, M. (2012): Applied Statistics for Economists, 4th edition. Khosla Publishing House by arrangement with Pitman. 4. Mukhopadhyay P. (1999): Applied Statistics, Books and Allied (P) Ltd. Montogomery, D. C. (2009): Introduction to Statistical Quality Control, 6thEdition, Wiley India Pvt. Ltd.		

Semester- IV

Advanced Statistical Analysis (BHM 254)

1	Course Code	BHM 254
2	**Course Title**	**Advanced Statistical Analysis**
3	Credits	3
4	Contact Hours (L-T-P)	0-0-5
	Course Status	Compulsory
5	Course Objective	After completing this course, students are expected to become a specialist to analyze the observed phenomena in advanced statistical level. More importantly, students are expected to provide analytical solution to a problem using appropriately selected model and data and discover meaningful knowledge from the solution.
6	Course Outcomes	CO1: Describe how to Differentiate various probability distributions. (K1,K2)
		CO2: Understand the concept of estimation. (K2,K3)
		CO3: Know how to recognize the sampling distributions. (K2,K3)
		CO4: Learn non-parametric test such as the Chi- Square test for Independence as well as Goodness of Fit. (K3,K4)
		CO5: Know how to apply various statistical and analysis. (K3,K4,K5)
		CO6: Able to know statistical technique implantation in practical situation. (K3,K4,K5)
7	Course Description	This course provides students with statistical foundation of the various problems of real life. Students will learn to recognize the main features of the processes under investigation that could be analyzed in terms of advanced statistical approaches. Grading this course will help the future specialist to analyze the observed phenomena in advanced statistical level.
8	Outline syllabus :	
	Unit 1	
	A	Use of discrete distribution (Uniform, Binomial and Poisson) in real life problem.

	B	Use of continuous distribution (Normal, Exponential and Gamma) in real life problem.		
	C	Its applications in Industrial work.		
	Unit 2			
	A	Sampling Distributions.		
	B	$\chi 2$ distribution properties and Interrelationships.		
	C	t distribution properties and Interrelationships.		
	Unit 3			
	A	F distribution properties.		
	B	Interrelationship of $\chi 2$, t, F distributions.		
	C	Point Estimation, Interval estimation for mean, variance of normal population and proportion of binomial population.		
	Unit 4			
	A	Type I and Type II errors, Critical Region, Size of the test, P value, Power.		
	B	Large Sample test -Z test.		
	C	Large Sample test – Chi-Square test-goodness of fit, test of independence.		
	Unit 5			
	A	ANOVA, Randomization and randomization-based		
		analysis.		
	B	Factor Analysis.		
	C	Cluster and Principal Components Analysis (PCM).		
	Mode of examination	Practical		
	Weightage Distribution	CA	Viva	ETE
		25	25	50

	Text book/s*	1. Gupta. S.C. & Kapoor,V.K. (2002) . Fundamentals of Mathematical Statistics , Sultan Chand & Sons Pvt. Ltd. New Delhi. Westfall, P., & Henning, K. S. (2013). Understanding advanced statistical methods. CRC Press.
	Other References	1. Croxton, Fredrick E., Cowden, Dudley J. and Klein, S. (1973): Applied General Statistics, 3rd Edition. Prentice Hall of India Pvt. Ltd. 2. Mukhopadhyay P. (1999): Applied Statistics, Books and Allied (P)Ltd. 3.Triola, M. M., & Triola, M. F. (2006). Biostatistics for the biological and health sciences (pp. 47-48). Boston: Pearson Addison-Wesley.

These courses (Core: Major/ Minor) can be opted as an elective by the students of following subjects: Engg. and Tech. (UG), Economics(UG/PG), Commerce(UG), BBA/BCA, B.Sc.(C.S.) and others.

Suggested Continuous Evaluation Methods: Max. Marks: 25

SN	Assessment Type	Max. Marks
1	**Class Tests**	**10**
2	**Online Quizzes/ Objective Tests**	**5**
3	**Presentation/** Research Orientation assignment	**5**
4	**Assignment (Indian Ancient Mathematics/ Statistics and Mathematicians/ Statisticians).**	**5**

---------- **THE END** ----------