ANDHRA PRADESH STATE COUNCIL OF HIGHER EDUCATION

Programme: B.Sc. Honours Data Science (Major)

w.e.f. AY 2023-24

COURSE STRUCTURE

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
I	I	1	Essentials and Applications of Mathematical, Physical and Chemical Sciences	3+2	4
		2	Advances in Mathematical, Physical and Chemical Sciences	3+2	4
	II	3	Introduction to Data Science and R Programming	3	3
			Introduction to Data Science and R Programming Practical Course	2	1
		4	Descriptive Statistics	3	3
			Descriptive Statistics Practical Course	2	1
II	III	5	Python Programming for Data Analysis	3	3
			Python Programming for Data Analysis Practical Course	2	1
		6	Inferential and applied statistics	3	3
			Inferential and applied statistics Practical Course	2	1
		7	Data mining techniques using R	3	3
			Data mining techniques using R Practical Course	2	1
		8	Web technologies	3	3
			Web technologies Practical Course	2	1
	IV	9	Data visualization using Tableau	3	3
			Data visualization using Tableau Practical Course	2	1
		10	Data visualization using python	3	3
			Data visualization using python Practical Course	2	1
		11	Introduction to SQL & Advanced Tableau	3	3
			Introduction to SQL & Advanced Tableau Practical Course	2	1
III	V	12	Supervised Machine Learning with Python	3	3
			Supervised Machine Learning with Python Practical Course	2	1

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
		13	Unsupervised Machine Learning with Python	3	3
			Unsupervised Machine Learning with Python Practical Course	2	1
		14 A	Web Scraping with Python	3	3
			Web Scraping with Python Practical Course	2	1
		OR			
		14 B	Predictive & Advanced Analytics using R	3	3
			Predictive & Advanced Analytics using R Practical Course	2	1
		15 A	Advanced Data Analysis Using Python	3	3
			Advanced Data Analysis Using Python Practical Course	2	1
		OR			
		15 B	Data Wrangling with Java Script	3	3
			Data Wrangling with Java Script Practical Course	2	1
	VI	Semester Internship/Apprenticeship with 12 Credits			
IV	VII	16 A	Big Data Analytics Using Spark & Hadoop	3	3
			Big Data Analytics Using Spark & Hadoop Practical Course	2	1
		OR			
		16 B	Big Data security	3	3
			Big Data security Practical Course	2	1
		17 A	Introduction to Deep Learning	3	3
			Introduction to Deep Learning Practical Course	2	1
		OR			
		17 B	Deep Learning with Pytorch	3	3
			Deep Learning with Pytorch Practical Course	2	1
		18 A	AI Concepts and Techniques With Python	3	3
			AI Concepts and Techniques With Python Practical Course	2	1
		OR			
		18 B	Data and Information Security	3	3
			Data and Information Security Practical Course	2	1
		SEC			
		19	Introduction to Neural Networks	3	3
			Introduction to Neural Networks Practical Course	2	1
		20	Natural Language Processing	3	3
			Natural Language Processing Practical	2	1

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
			Course		
	VIII	21 A	Research Exploration	3	3
			Research Exploration Practical Course	2	1
		OR			
		21 B	Computational Data Science	3	3
			Computational Data Science Practical Course	2	1
		22 A	Computer Vision with Python	3	3
			Computer Vision with Python Practical Course	2	1
		OR			
		22 B	Data Wrangling with Java Script	3	3
			Data Wrangling with Java Script Practical Course	2	1
		23 A	Social Media Analytics	3	3
			Social Media Analytics Practical Course	2	1
		OR			
		23 B	Pyspark Essentials For Data Science	3	3
			Pyspark Essentials For Data Science Practical Course	2	1
		SEC			
		24	Business Intelligence and Visualization	3	3
			Business Intelligence and Visualization Practical Course	2	1
		25	Data Visualization using JavaScript	3	3
			Data Visualization using JavaScript Practical Course	2	1

SEMESTER-I

COURSE 1: ESSENTIALS AND APPLICATIONS OF MATHEMATICAL, PHYSICAL ANDCHEMICAL SCIENCES

Theory Credits: 4 5 hrs/week

Course Objective:

The objective of this course is to provide students with a comprehensive understanding of the essential concepts and applications of mathematical, physical, and chemical sciences. The course aims to develop students' critical thinking, problem-solving, and analytical skills in these areas, enabling them to apply scientific principles to real-world situations.

Learning outcomes:

1. Apply critical thinking skills to solve complex problems involving complex numbers, trigonometric ratios, vectors, and statistical measures.

2. To Explain the basic principles and concepts underlying a broad range of fundamental areas of physics and to Connect their knowledge of physics to everyday situations

3. To Explain the basic principles and concepts underlying a broad range of fundamental areas of chemistry and to Connect their knowledge of chemistry to daily life.

4. Understand the interplay and connections between mathematics, physics, and chemistry in various applications. Recognize how mathematical models and physical and chemical principles can be used to explain and predict phenomena in different contexts.

5 To explore the history and evolution of the Internet and to gain an understanding of network security concepts, including threats, vulnerabilities, and countermeasures.

UNIT I: ESSENTIALS OF MATHEMATICS:

Complex Numbers: Introduction of the new symbol i – General form of a complex number – Modulus-Amplitude form and conversions

Trigonometric Ratios: Trigonometric Ratios and their relations – Problems on calculation of angles**Vectors:** Definition of vector addition – Cartesian form – Scalar and vector product and problems **Statistical Measures**: Mean, Median, Mode of a data and problems

UNIT II: ESSENTIALS OF PHYSICS:

Definition and Scope of Physics- Measurements and Units - Motion of objects: Newtonian Mechanics and relativistic mechanics perspective - Laws of Thermodynamics and Significance- Acoustic waves and electromagnetic waves- Electric and Magnetic fields and their interactions- Behaviour of atomic and nuclear particles- Wave-particle duality, the uncertainty principle- Theories and understanding of universe

UNIT III: ESSENTIALS OF CHEMISTRY: :

Definition and Scope of Chemistry- Importance of Chemistry in daily life -Branches of chemistry and significance- Periodic Table-

Electronic Configuration, chemical changes, classification of matter, Biomolecules- carbohydrates, proteins, fats and vitamins.

UNIT IV: APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY:

Applications of Mathematics in Physics & Chemistry: Calculus , Differential Equations & Complex Analysis

Application of Physics in Industry and Technology: Electronics and Semiconductor Industry, Robotics and Automation, Automotive and Aerospace Industries, Quality Control and Instrumentation, Environmental Monitoring and Sustainable Technologies.

Application of Chemistry in Industry and Technology: Chemical Manufacturing, Pharmaceuticals and Drug Discovery, Materials Science, Food and Beverage Industry.

UNIT V: ESSENTIALS OF COMPUTER SCIENCE:

Milestones of computer evolution - Internet, history, Internet Service Providers, Types of Networks, IP, Domain Name Services, applications.

Ethical and social implications: Network and security concepts- Information Assurance Fundamentals, Cryptography-Symmetric and Asymmetric, Malware, Firewalls, Fraud Techniques- Privacy and Data Protection

Recommended books:

1. Functions of one complex variable by John.B.Conway, Springer- Verlag.

2. Elementary Trigonometry by H.S.Hall and S.R.Knight

3.Vector Algebra by A.R.Vasishtha, Krishna Prakashan Media(P)Ltd.

4.Basic Statistics by B.L.Agarwal, New age international Publishers

5. University Physics with Modern Physics by Hugh D. Young and Roger A. Freedman

6. Fundamentals of Physics by David Halliday, Robert Resnick, and Jearl Walker

7. Physics for Scientists and Engineers with Modern Physics" by Raymond A. Serway and John W. Jewett Jr.

8. Physics for Technology and Engineering" by John Bird

9. Chemistry in daily life by Kirpal Singh

10. Chemistry of bio molecules by S. P. Bhutan

11. Fundamentals of Computers by V. Raja Raman

12. Cyber Security Essentials by James Graham, Richard Howard, Ryan Olson

STUDENT ACTIVITIES

UNIT I: ESSENTIALS OF MATHEMATICS:

1: Complex Number Exploration

Provide students with a set of complex numbers in both rectangular and polar forms.

They will plot the complex numbers on the complex plane and identify their properties

2: Trigonometric Ratios Problem Solving

Give students a set of problems that require the calculation of trigonometric ratios and their relations.

Students will solve the problems using the appropriate trigonometric functions (sine, cosine, tangent, etc.) and trigonometric identities.

3: Vector Operations and Applications

Provide students with a set of vectors in Cartesian form.

Students will perform vector addition and subtraction operations to find the resultant vectors.

They will also calculate the scalar and vector products of given vectors.

4: Statistical Measures and Data Analysis

Give students a dataset containing numerical values.

Students will calculate the mean, median, and mode of the data, as well as other statistical measures if appropriate (e.g., range, standard deviation).
They will interpret the results and analyze the central tendencies and distribution of the data.

UNIT II: ESSENTIALS OF PHYSICS:

1. Concept Mapping

Divide students into groups and assign each group one of the topics.

Students will create a concept map illustrating the key concepts, relationships, and applications related to their assigned topic.

Encourage students to use visual elements, arrows, and labels to represent connections and interdependencies between concepts.

2. Laboratory Experiment

Select a laboratory experiment related to one of the topics, such as motion of objects or electric and magnetic fields.
Provide the necessary materials, instructions, and safety guidelines for conducting the experiment.
Students will work in small groups to carry out the experiment, collect data, and analyze the results.

After the experiment, students will write a lab report summarizing their findings, observations, and conclusions.

UNIT III: ESSENTIALS OF CHEMISTRY

1: Chemistry in Daily Life Presentation

Divide students into groups and assign each group a specific aspect of daily life where chemistry plays a significant role, such as food and nutrition, household products, medicine,or environmental issues.
Students will research and create a presentation (e.g., PowerPoint, poster, or video) that showcases the importance of chemistry in their assigned aspect.
2: Periodic Table Exploration

Provide students with a copy of the periodic table.

Students will explore the periodic table and its significance in organizing elements based on their properties.
They will identify and analyze trends in atomic structure, such as electronic configuration, atomic size, and ionization energy.

3: Chemical Changes and Classification of Matter

Provide students with various substances and chemical reactions, such as mixing acids and bases or observing a combustion reaction.
Students will observe and describe the chemical changes that occur, including changes in color, temperature, or the formation of new substances.
4: Biomolecules Investigation

Assign each student or group a specific biomolecule category, such as carbohydrates,proteins, fats, or vitamins.
Students will research and gather information about their assigned biomolecule category, including its structure, functions, sources, and importance in the human body.

They can create informative posters or presentations to present their findings to the class.

UNIT IV: APPLICATIONS OF MATHEMATICS, PHYSICS & CHEMISTRY

1: Interdisciplinary Case Studies

Divide students into small groups and provide them with interdisciplinary case studies that involve the interdisciplinary application of mathematics, physics, and chemistry.

Each case study should present a real-world problem or scenario that requires the integration of concepts from all three disciplines.

2: Design and Innovation Project

Challenge students to design and develop a practical solution or innovation that integrates mathematics, physics, and chemistry principles.

Students can choose a specific problem or area of interest, such as renewable energy, environmental conservation, or materials science.
3: Laboratory Experiments

Assign students laboratory experiments that demonstrate the practical applications of mathematics, physics, and chemistry.

Examples include investigating the relationship between concentration and reaction rate, analyzing the behavior of electrical circuits, or measuring the properties of materials.

.4: Mathematical Modeling

Present students with real-world problems that require mathematical modeling and analysis.

UNIT V: ESSENTIALS OF COMPUTER SCIENCE:

1. Identifying the attributes of network (Topology, service provider, IP address and bandwidth of

2. your college network) and prepare a report covering network architecture.

3. Identify the types of malwares and required firewalls to provide security.

4. Latest Fraud techniques used by hackers.

SEMESTER-I

COURSE 2: ADVANCES IN MATHEMATICAL, PHYSICAL AND CHEMICAL SCIENCES

Theory | Credits: 4 | 5 hrs/week

Course Objective:

The objective of this course is to provide students with an in-depth understanding of the recent advances and cutting-edge research in mathematical, physical, and chemical sciences. The course aims to broaden students' knowledge beyond the foundational concepts and expose them to the latest developments in these disciplines, fostering critical thinking, research skills, and the ability to contribute to scientific advancements.

Learning outcomes:

1. Explore the applications of mathematics in various fields of physics and chemistry, to understand how mathematical concepts are used to model and solve real-world problems.

2. To Explain the basic principles and concepts underlying a broad range of fundamental areas of physics and to Connect their knowledge of physics to everyday situations.

3. Understand the different sources of renewable energy and their generation processes and advances in nanomaterials and their properties, with a focus on quantum dots. To study the emerging field of quantum communication and its potential applications. To gain an understanding of the principles of biophysics in studying biological systems. Explore the properties and applications of shape memory materials.

3. Understand the principles and techniques used in computer-aided drug design and drug delivery systems, to understand the fabrication techniques and working principles of nanosensors. Explore the effects of chemical pollutants on ecosystems and human health.

4. Understand the interplay and connections between mathematics, physics, and chemistry in various advanced applications. Recognize how mathematical models and physical and chemical principles can be used to explain and predict phenomena in different contexts.

5 Understand and convert between different number systems, such as binary, octal, decimal, and hexadecimal. Differentiate between analog and digital signals and understand their characteristics.Gain knowledge of different types of transmission media, such as wired (e.g., copper cables, fiber optics) and wireless (e.g., radio waves, microwave, satellite)..

UNIT I: ADVANCES IN BASICS MATHEMATICS

Straight Lines: Different forms – Reduction of general equation into various forms – Point ofintersection of two straight lines

Limits and Differentiation: Standard limits – Derivative of a function –Problems on product ruleand quotient rule

Integration: Integration as a reverse process of differentiation – Basic methods of integration

Matrices: Types of matrices – Scalar multiple of a matrix – Multiplication of matrices – Transpose ofa matrix and determinants

UNIT II: ADVANCES IN PHYSICS:

Renewable energy: Generation, energy storage, and energy-efficient materials and devices. **Recent advances in the field of nanotechnology**: Quantum dots, Quantum Communication- recent advances in biophysics- recent advances in medical physics- Shape Memory Materials.

UNIT III: ADVANCES IN CHEMISTRY:

Computer aided drug design and delivery, nano sensors, Chemical Biology, impact of chemical pollutants on ecosystems and human health, Dye removal - Catalysis method

UNIT IV: ADVANCED APPLICATIONS OF MATHEMATICS, PHYSICS &CHEMISTRY

Mathematical Modelling applications in physics and chemistry

Application of Renewable energy: Grid Integration and Smart Grids,

Application of nanotechnology: Nanomedicine,

Application of biophysics: Biophysical Imaging, Biomechanics, Neurophysics,

Application of medical physics: Radiation Therapy, Nuclear medicine

Solid waste management, Environmental remediation- Green Technology, Water treatment.

UNIT V: Advanced Applications of computer Science

Number System-Binary, Octal, decimal, and Hexadecimal, Signals-Analog, Digital, Modem, Codec, Multiplexing, Transmission media, error detection and correction- Parity check and CRC, Networking devices- Repeater, hub, bridge, switch, router, gateway.

Recommended books:

1. Coordinate Geometry by S.L.Lony, Arihant Publications
2. Calculus by Thomas and Finny, Pearson Publications
3. Matrices by A.R.Vasishtha and A.K.Vasishtha, Krishna Prakashan Media(P)Ltd.
4. "Renewable Energy: Power for a Sustainable Future" by Godfrey Boyle
5. "Energy Storage: A Nontechnical Guide" by Richard Baxter
6. "Nanotechnology: Principles and Applications" by Sulabha K. Kulkarni and Raghvendra A. Bohara
7. "Biophysics: An Introduction" by Rodney Cotterill
8. "Medical Physics: Imaging" by James G. Webster
9. "Shape Memory Alloys: Properties and Applications" by Dimitris C. Lagoudas
10. Nano materials and applications by M.N.Borah

11. Environmental Chemistry by Anil.K.D.E.

12. Digital Logic Design by Morris Mano

13. Data Communication & Networking by Bahrouz Forouzan.

STUDENT ACTIVITIES

UNIT I: ADVANCES IN BASIC MATHEMATICS

1: Straight Lines Exploration

Provide students with a set of equations representing straight lines in different forms, suchas slope-intercept form, point-slope form, or general form.

Students will explore the properties and characteristics of straight lines, including theirslopes, intercepts, and point of intersection.

2: Limits and Differentiation Problem Solving

Students will apply the concept of limits to solve various problems using standard limits.

Encourage students to interpret the results and make connections to real-world applications,such as analyzing rates of change or optimizing functions.

3: Integration Exploration

Students will explore the concept of integration as a reverse process of differentiation andapply basic methods of integration, such as the product rule, substitution method, or integration by parts.

Students can discuss the significance of integration in various fields, such as physics and chemistry

4: Matrices Manipulation

Students will perform operations on matrices, including scalar multiplication, matrix multiplication, and matrix transpose.

Students can apply their knowledge of matrices to real-world applications, such as solving systems of equations or representing transformations in geometry.

UNIT II: ADVANCES IN PHYSICS:

1: Case Studies

Provide students with real-world case studies related to renewable energy, nanotechnology, biophysics, medical physics, or shape memory materials.

Students will analyze the case studies, identify the challenges or problems presented, and propose innovative solutions based on the recent advances in the respective field.

They will consider factors such as energy generation, energy storage, efficiency, sustainability, materials design, biomedical applications, or technological advancements.

2: Experimental Design

Assign students to design and conduct experiments related to one of the topics: renewable

energy, nanotechnology, biophysics, medical physics, or shape memory materials.
They will identify a specific research question or problem to investigate and design an experiment accordingly.

Students will collect and analyze data, interpret the results, and draw conclusions based on their findings.

They will discuss the implications of their experimental results in the context of recentadvances in the field.

3: Group Discussion and Debate

Organize a group discussion or debate session where students will discuss the ethical, social, and environmental implications of the recent advances in renewable energy, nanotechnology, biophysics, medical physics, and shape memory materials.

Assign students specific roles, such as proponent, opponent, or moderator, and provide them with key points and arguments to support their positions.

UNIT III: ADVANCES IN CHEMISTRY:

1. Experimental Design and Simulation

In small groups, students will design experiments or simulations related to the assigned topic.

For example, in the context of computer-aided drug design, students could design a virtual screening experiment to identify potential drug candidates for a specific disease target.

For nano sensors, students could design an experiment to demonstrate the sensitivity and selectivity of nano sensors in detecting specific analytes.
Chemical biology-related activities could involve designing experiments to study enzyme-substrate interactions or molecular interactions in biological systems.
Students will perform their experiments or simulations, collect data, analyze the results, and draw conclusions based on their findings.

2. Case Studies and Discussion

Provide students with real-world case studies related to the impact of chemical pollutants on ecosystems and human health.
Students will analyze the case studies, identify the sources and effects of chemical pollutants, and propose mitigation strategies to minimize their impact.

Encourage discussions on the ethical and environmental considerations when dealing with chemical pollutants.
For the dye removal using the catalysis method, students can explore case studies where catalytic processes are used to degrade or remove dyes from wastewater.
Students will discuss the principles of catalysis, the advantages and limitations of the catalysis method, and its applications in environmental remediation.

3: Group Project

Assign students to work in groups to develop a project related to one of the topics.

The project could involve designing a computer-aided drug delivery system, developing anano sensor for a specific application, or proposing strategies to mitigate the impact of chemical pollutants on ecosystems.
Students will develop a detailed project plan, conduct experiments or simulations, analyze data, and present their findings and recommendations.
Encourage creativity, critical thinking, and collaboration throughout the project.

UNIT IV: ADVANCED APPLICATIONS OF MATHEMATICS, PHYSICS &CHEMISTRY

1: Mathematical Modelling Experiment

Provide students with a mathematical modelling experiment related to one of the topics. For example, in the context of renewable energy, students can develop a mathematical model to optimize the placement and configuration of solar panels in a solar farm.
Students will work in teams to design and conduct the experiment, collect data, and analyze the results using mathematical models and statistical techniques.
They will discuss the accuracy and limitations of their model, propose improvements, and interpret the implications of their findings in the context of renewable energy or the specific application area.

2: Case Studies and Group Discussions

Assign students to analyze case studies related to the applications of mathematical modellingin nanotechnology, biophysics, medical physics, solid waste management, environmental remediation, or water treatment.
Students will discuss the mathematical models and computational methods used in the case studies, analyze the outcomes, and evaluate the effectiveness of the modelling approach.
Encourage group discussions on the challenges, ethical considerations, and potential advancements in the field.

Students will present their findings and engage in critical discussions on the advantages and limitations of mathematical modelling in solving complex problems in these areas.

3. Group Project

Assign students to work in groups to develop a group project that integrates mathematical modelling with one of the application areas: renewable energy, nanotechnology, biophysics, medical physics, solid waste management, environmental remediation, or water treatment.

The project could involve developing a mathematical model to optimize the delivery of radiation therapy in medical physics or designing a mathematical model to optimize waste management practices.
Students will plan and execute their project, apply mathematical modelling techniques, analyze the results, and present their findings and recommendations.
Encourage creativity, critical thinking, and collaboration throughout the project.

UNIT V: Advanced Applications of computer Science

Students must be able to convert numbers from other number system to binary numbersystems

1. Identify the networking media used for your college network
2. Identify all the networking devices used in your college premises.

SEMESTER-II

COURSE 3: INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING

Theory Credits: 3 3 hrs/week

Aim and objectives of Course :

Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight. This course will introduce students to the collection. Preparation, analysis, modelling and visualization of data, covering both conceptual and practical issues. Examples and casestudies from diverse fields will be presented, and hands- on use of statistical and data manipulation software will be included.

Learning outcomes of Course:

- Recognize the various discipline that contribute to a successful data science effort.
- Understand the processes of data science identifying the problem to be solved, datacollection, preparation, modeling, evaluation and visualization.
- Be aware of the challenges that arise in Data Sciences.
- Be able to identify the application of the type of algorithm based on the type of the problem.
- Be comfortable using commercial and open source tools such as the R/Python languageandits associated libraries for data analytics and Visualization.

UNIT I:
Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process. Historyand Overview of R, Getting Started with R, R Nuts and Bolts

UNIT II:
The Data Science Process: Overview of the Data Science Process-Setting the research goal, Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. GettingData in and out of R, Using reader package, Interfaces to the outside world.

UNIT III:
Machine Learning: Understanding why data scientists use machine learning-What is machine learning and why we should care about, Applications of machine learning in data science, Where itis used in data science, The modeling process, Types of Machine Learning-Supervised and Unsupervised.

UNIT IV:
Handling large Data on a Single Computer: The problems we face when handling large data, General Techniques for handling large volumes of data, Generating programming tips fordealing with large datasets.

UNIT V:

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation. Case studies on preliminary data analysis.

TEXT BOOKS:

1. DavyCielen, Arno.D.B.Maysman, Mohamed Ali, "Introducing Data Science"ManningPublications, 2016.
2. Roger D. Peng, "R Programming for DataScience" Lean Publishing, 2015.

REFERENCE BOOKS:

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, AbhijitDasgupta, "PracticalData Science Cookbook", Packt Publishing Ltd., 2014.

WebReferences for case studies:

1. https://www.kaggle.com/datasets
2. https://github.com/

SEMESTER-II

COURSE 3: INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING

Practical | Credits: 1 | 2 hrs/week

Lab/Practical/Experiments/Tutorials syllabus:

1. Installing R and R studio, with proper notes on version management, cosmetic settings and different libraries.
2. Basic operations in r with arithmetic and statistics.
3. Getting data into R, Basic data manipulation, Loading Data into R
4. Basic plotting
5. Loops and functions
6. Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.
7. Demonstrate the visualization and graphics using visualization packages like ggplot2.
8. Implement Loop functions with lapply(), sapply(), tapply(), apply(), mapply().
9. Explore data using Single Variables: Unimodal, Bimodal, Histograms, Density Plots,Barcharts
10. Explore data using two Variables: Line plots, Scatter Plots, smoothing cures, Bar charts
11. Explore and implement commands using dplyr package
12. Download a dataset and work on basic data manipulation followed by inferential statistics.

RECOMMENDED TEXT BOOKS:

1. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley &Sons, Inc., 2012.
2. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.

Recommended Reference books:

3. The art of R Programming: A tour of Statistical Software design. Norman Matloff.KindleEdition
4. The book of R : The first course in Programming and Statistics by Tilman M. Davies.

Recommended Co-curricular activities: (Co-curricular Activities should not promote copyingfromtext book or from others' work and shall encourage self/independent and grouplearning)

A. Measurable:
1. Assignments on:
2. Student seminars (Individual presentation of papers) on topics relating to:
3. Quiz Programmes on:
4. Individual Field Studies/projects:
5. Group discussion on:
6. Group/Team Projects on:

B. General

1. Collection of news reports and maintaining a record of paper-cuttings relatingtotopics covered in syllabus
2. Group Discussions on:
3. Watching TV discussions and preparing summary points recording personalobservations etc., under guidance from the Lecturers
4. Any similar activities with imaginative thinking.
5. Recommended Continuous Assessment methods:

SEMESTER-II

COURSE 4: DESCRIPTIVE STATISTICS

Theory Credits: 3 3 hrs/week

Course Learning Outcomes: Students will acquire:

- knowledge of Statistics and its implementation through practical understanding for various domainsrelated to data science.
- knowledge of various types of data, their organization and evaluation of summary measures such asmeasures of central tendency and dispersion etc.
- knowledge of other types of data reflecting quality characteristics including concepts of independence andassociation between two attributes,
- insights into preliminary exploration of different types of data.
- Knowledge of correlation, regression analysis, regression diagnostics, partial and multiple correlations.

UNIT I:

Introduction to Statistics: Importance of Statistics. Scope of Statistics in different fields. Concepts of primary and secondary data. Diagrammatic and graphical representation of data: Histogram, frequency polygon, Pie. Measures of Central Tendency: Mean, Median, Mode, Geometric Mean and Harmonic Mean. Median and Mode through graph.

UNIT II:

Measures of Dispersion: Range, Quartile Deviation, Mean Deviation and Standard Deviation, Variance. Central and Non-Central moments and their interrelationship. Skewness and kurtosis.

UNIT III:

Curve fitting: Bi- variate data, Principle of least squares, fitting of degree polynomial. Fitting of straight line, Fitting of Second degree polynomial or parabola, Fitting of power curve and exponential curves.

Correlation: Meaning, Types of Correlation, Measures of Correlation: Scatter diagram, Karl Pearson's Coefficient of Correlation, Rank Correlation Coefficient (with and without ties), Bi-variate frequency distribution, correlation coefficient for bi-variate data and simple problems. Concept of multiple and partial correlation coefficients (three variables only) and properties

UNIT IV:

Regression : Concept of Regression, Linear Regression: Regression lines, Regression coefficients and it's properties, Regressions lines for bi-variate data and simple problems. Correlation vs regression, sigmoid curve, derivation from linear regression to logistic regression.

UNIT-V

Attributes : Notations, Class, Order of class frequencies, Ultimate class frequencies, Consistency of data,Conditions for consistency of data for 2 and 3 attributes only , Independence of attributes , Association of attributes and its measures, Relationship between association and colligation of attributes, Contingency table: Square contingency, Mean square contingency, Coefficient of mean square contingency,

TEXT BOOKS:

1. V.K.Kapoor and S.C.Gupta: Fundamentals of MathematicalStatistics,Sultan Chand &Sons,NewDelhi.
2. BA/BSc I year statistics - descriptive statistics, probability distribution - Telugu Academy - DrM.Jaganmohan Rao,Dr N.Srinivasa Rao, Dr P.Tirupathi Rao, Smt.D.Vijayalakshmi.
3. K.V.S. Sarma: Statistics Made Simple: Do it yourself on PC. PHI

REFERENCE BOOKS:

1. Willam Feller: Introduction to Probabilitytheory and its applications. Volume –I,Wiley
2. Goon AM, Gupta MK, Das Gupta B : Fundamentals of Statistics , Vol-I, the World Press Pvt.Ltd.,Kolakota.
3. Hoel P.G: Introduction to mathematical statistics, Asia Publishinghouse.
4. M. JaganMohan Rao and Papa Rao: A Text book of StatisticsPaper-I.
5. Sanjay Arora and Bansi Lal: New Mathematical Statistics: Satya Prakashan , NewDelhi

SEMESTER-II

COURSE 4: DESCRIPTIVE STATISTICS

Practical | Credits: 1 | 2 hrs/week

List of the experiments:

1. Graphical presentation of data (Histogram, frequency polygon).
2. Diagrammatic presentation of data (Bar andPie).
3. Computation of measures of central tendency(Mean, Median andMode)
4. Computation of measures of dispersion(Q.D, M.D andS.D)
5. Computation of non-central, central moments, □ l□and 2 for ungroupeddata.
6. Computation of Karl Pearson's coefficients of Skewness and Bowley's coefficients ofSkewness.
7. Fitting of straight line by the method of leastsquares
8. Fitting of parabola by the method of leastsquares
9. Fitting of power curve of the type by the method of leastsquares.
10. Fitting of exponential curve of the type and by the method of leastsquares.
11. Computation of correlation coefficient and regression lines for ungroupeddat.
12. Computation of correlation coefficient, forming regression lines for groupeddata

SEMESTER-III

COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

Theory | Credits: 3 | 3 hrs/week

Aim and objectives of Course:

- To be able to Program in Python
- To know and understand the data Analysis phases
- To know the usage of all libraries

Learning outcomes of Course:

- Understands and learn all basic concepts of
- Python Program Data Analysis methods in Python
- Get used with Python Programming environments

UNIT I:

What is Data Analysis? Differences between Data Analysis and Analytics, What is Python, Why Python for Data Analysis? What is Library, Essential Python Libraries. Python Language basics, I Python and Jupyter Notebook. Python Language Basics.

UNIT II:

Built-in Data Structures, Functions, Files and Operating System. **NumPy Basics:** Arrays and Vectorized Computation, The Numpy ndarray, Universal Functions, Array-Oriented Programming with Arrays, File Input and Output with Arrays, Linear Algebra, Pseudorandom Number Generation.

UNIT III:

Getting Started with Pandas: Introduction to Pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics

Data Loading, Storage and File Formats: Reading and Writing Data in TextFormat, Binary Data Formats, Interacting with Web APIs, Interacting with Databases.

UNIT IV:

Data Cleaning and Preparation: Handling Missing Data, Data Transformation,String Manipulation.

Data Wrangling: Join, Combine and Reshape: Hierarchical Indexing, Combiningand Merging Datasets, Reshaping and Pivoting.

UNIT V:

Introduction to Modeling Libraries in Python: Interfacing between pandas andModel code, Creating model descriptions with Patsy, Introduction to stats models.

Plotting and Visualization: A brief matplotlib API Primer, Plotting with Pandas and Seaborn, Other Python visualization tools.

TEXT BOOKS:

1. Wes McKinney "Python for Data Analysis" O'reilly Publications Second edition
2. Charles R Suverance "Python for Everybody" Exploring data using Python 3

REFERENCE BOOKS:

1. John Zelle Michael Smith Python Programming, second edition 2010

Co-curricular Activities
Take up any application which involves the python coding. Example Case studies/Simulators: (https://knightlab.northwestern.edu/2014/06/05/five-mini-programming-projects-for-the-python-beginner/)

- Dice Rolling Simulator
- Guess the number
- Text based adventure game
- Hangman

Continuous assessment:

Let the students be tested in the following questions from each unit

1. What is Data Analysis. List out the differences between data analysis and dataanalytics
2. What is Python? Explain Python basics
3. Explain NumPy Basics
4. What is Data Loading. Explain Pandas Data Structures
5. What is Data Cleaning. Explain different phases in it
6. Explain Plotting and Visualization in Python

SEMESTER-III

COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

Practical Credits: 1 2 hrs/week

1. Use matplotlib and plot an inline in Jupyter.
2. Implement commands of Python Language basics
3. Create Tuples, Lists and illustrate slicing conventions.
4. Create built-in sequence functions.
5. Clean the elements and transform them by using List, Set and DictComprehensions.
6. Create a functional pattern to modify the strings in a high level.
7. Write a Python Program to cast a string to a floating-point number but fails with Value Error on improper inputs using Errors and Exception handling.
8. Create an n array object and use operations on it.
9. Use arithmetic operations on Numpy Arrays
10. Using Numpy array perform Indexing and Slicing Boolean Indexing,FancyIndexing operations
11. Create an image plot from a two-dimensional array of function values.
12. Implement some basic array statistical methods (sum, mean, std, var, min,max, argmin, argmax, cumsum andcumprod) and sorting with sortmethod.
13. Implement numpy.random functions.
14. Plot the first 100 values on the values obtained from random walks.
15. Create a data frame using pandas and retrieve the rows and columns in itbyperforming some indexing options and transpose it.
16. Implement the methods of descriptive and summary statistics
17. Load and write the data from and to different file formats including WebAPIs.
18. Implement the data Cleaning and Filtering methods(Use NAhandlingmethods, fillna function arguments)
19. Transform the data using function or mapping
20. Rearrange the data using unstack method of hierarchical Indexing
21. Implement the methods that summarize the statistics by levels.
22. Use different Join types with how argument and merge data with keys andmultiple keys.

SEMESTER-III

COURSE 6: INFERENTIAL AND APPLIED STATISTICS

Theory Credits: 3 3 hrs/week

Course Learning Outcomes
After completion of this course, the students will know about

- Concept of law large numbers and their uses
- knowledge about important inferential aspects such as point estimation, test ofhypotheses and associated concepts,
- knowledge about inferences from Binomial, Poisson and Normal distributions as illustrations,
- concept about non-parametric method and some important non-parametric tests.

- Time series data, its applications to various fields and components of time series,
- Various data collection methods enabling to have a better insight in policy making, planning and systematic implementation, Construction and implementation of life tables, Population growth curves, population estimates and projections,
- Real data implementation of various demographic concepts as outlined above through practical assignments.

UNIT I:
Concepts: Population, Sample, Parameter, statistic, Sampling distribution, Standard error. convergence in probability and convergence in distribution, law of large numbers, central limit theorem (statements only).Student's t- distribution, F – Distribution, χ^2-Distribution: Definitions, properties and their applications.

UNIT II:
Theory of estimation and Hypothesis: Estimation of a parameter, criteria of a good estimator – unbiasedness, consistency, efficiency, &sufficiency and. Binomial, Poisson &Normal Population parameters estimate by MLE method. Confidence Intervals. Concepts of statistical hypotheses, null and alternative hypothesis, critical region, two types of errors, level of significance and power of a test. Examples in case ofBinomial, Poisson and Normal distributions.

UNIT III:
Sample tests: t-test for single mean, difference of means and paired t-test.□2. confidence intervalsfor mean(s). standard deviation(s) and correlation coefficient(s).Test for goodness of fit andindependence of attributes. F-test for equality of variances.
Non-parametric tests- their advantages and disadvantages, comparison with parametric tests. Measurementscale- nominal, ordinal, interval and ratio.

UNIT IV:
Time Series: Time Series and its components with illustrations, additive, multiplicative models. Trends:Estimation of trend by free hand curve method, method of semi averages. Determination of trend by least squares (Linear trend, parabolic trend only), moving averages method.

UNIT V:

Vital Statistics: Introduction, definition and uses of vital statistics, sources of vital statistics. measures ofdifferent Mortality and Fertility rates, Measurement of population growth. Life tables: construction and uses of life tables.

TEXT BOOKS:

1. BA/BSc II year statistics - statistical methods and inference - Telugu Academy by A.Mohanrao,N.Srinivasa Rao, Dr R.Sudhakar Reddy, Dr T.C. RavichandraKumar.
2. K.V.S. Sarma: Statistics Made Simple: Do it yourself on PC.PHI.
3. Fundamentals of applied statistics : VK Kapoor and SCGupta.
4. BA/BSc III year paper - III Statistics - applied statistics - Telugu academy by prof.K.SrinivasaRao,Dr D.Giri. Dr A.Anand, Dr V.PapaiahSastry.

REFERENCE BOOKS:

1. Brockwell, P.J. and Devis, R.A. (2003). Introduction to Time Series Analysis. Springer.
2. Chatfield, C. (2001). Time Series Forecasting., Chapman & Hall.
3. Srinivasan, K. (1998). Demographic Techniques and Applications. Sage Publications
4. Srivastava O.S. (1983). A Text Book of Demography. Vikas Publishing House
5. Fundamentals of Mathematics statistics : VK Kapoor and SCGuptha.
6. Outlines of statistics, Vol II : Goon Guptha, M.K.Guptha, Das GupthaB.
7. Introduction to Mathematical Statistics : HoelP.G.
8. Hogg Tanis Rao: Probability and Statistical Inference. 7th edition.Pearson.

CO-CURRICULAR ACTIVITIES:

- Quiz Competition
- Expert Lectures
- Seminars

EXTRA CURRICULAR ACTIVITIES:

- Formal Examination
- Lab Practical
- Presentation
- Simple Projects

SEMESTER-III

COURSE 6: INFERENTIAL AND APPLIED STATISTICS

Theory Credits: 3 3 hrs/week

List of Experiments:

1. Large sample test for difference of means.
2. Large sample test for single proportion
3. Large sample test for difference of proportions , standard deviations , correlation coefficient.
4. Small sample test for single mean, difference of means and correlation coefficient
5. Paired t-test(pairedsamples).
6. Small sample test for single variance(χ2 - test)

Time Series:

7. Measurement of trend by method of moving averages(odd and evenperiod)
8. Measurement of trend by method of Least squares(linear andparabola)
9. Determination of seasonal indices by method simpleaverages
10. Determination of seasonal indices by method of Ratio to movingaverages

Vital Statistics:

11. Computation of various Mortalityrates
12. Computation of various Fertilityrates
13. Computation of various Reproductionrates.
14. Construction of Life Tables

SEMESTER-III

COURSE 7: DATA MINING TECHNIQUES USING R

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:

- To understand Data mining techniques and algorithms.
- Comprehend the data mining environments and application.

Learning outcomes of Course:

Students who complete this course will be able to

- Compare various conceptions of data mining as evidenced in both research andapplication.
- Evaluate mathematical methods underlying the effective application of data mining.
- Should be able to apply the type of techniques based on the problems considered.
- Can find out the market patterns and association amongst different products.

UNIT I:

An idea on Data Warehouse, Data mining-KDD versus data mining, Stages of the Data MiningProcess-Task primitives., Data Mining Techniques – Data mining knowledge representation.

UNIT II

Data mining query languages- Integration of Data Mining System with a Data Warehouse-Issues, Data pre-processing – Data Cleaning, Data transformation – Feature selection – Dimensionality reduction

UNIT III

Concept Description: Characterization and comparison What is Concept Description,Data Generalization by Attribute-Oriented Induction(AOI), AOI for Data Characterization, Efficient Implementation of AOI.

Mining Frequent Patterns, Associations and Correlations: Basic Concepts, FrequentItemset Mining Methods: Apriori method, generating Association Rules, Improvingthe Efficiency of Apriori,Pattern-Growth Approach for mining Frequent Item sets.

UNIT-IV

Classification Basic Concepts: Basic Concepts, Decision Tree Induction: Decision TreeInduction Algorithm, Attribute Selection Measures, Tree Pruning. Bayes Classification Methods.

UNIT-V

Association rule mining: Antecedent, consequent , muti-relational association rules,

ECLAT.Case study on Market Basket Analysis.

Cluster Analysis: Cluster Analysis, Partitioning Methods, Hierarchal methods, Density basedmethods-DBSCAN.

TEXT BOOKS:

1. Jiawei Han, MichelineKamber, Jian Pei."Data Mining: Concepts and Techniques", 3rd Edition.Morgan Kaufmann Publishers, 2011.
2. AdelchiAzzalini, Bruno Scapa, "Data Analysis and Data mining" , 2ndEdiiton, Oxford Univeristy Press Inc., 2012.
3. Data Mining, The Textbook (2015) by Charu Aggarwal.

REFERENCES BOOKS:

1. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", 10th Edition, TataMcGraw Hill Edition , 2007.
2. G.K. Gupta, "Introduction to Data Mining with Case Studies", 1st Edition, EasterEconomy Edition, PHI, 2006.

Student Activities:

1. Students should be able to implement Data Mining algorithms provided the relevantdata
2. Given the data, students can visualize all statistical measures
3. Differentiate the types of mining problems and identify what type of algorithms are tobe implemented.

Continuous assessment:

Let the students be tested in the following questions from each unit

1. What is Data Mining and KDD? Where Data Mining fits in KDD Process
2. Describe all Preprocessing methods
3. Explain Data Description and AOI Algorithm
4. Explain Classification and Write any Decision tree induction algorithm
5. Explain the concept of clustering and write any algorithm to form clusters.

SEMESTER-III

COURSE 7: DATA MINING TECHNIQUES USING R

Practical | Credits: 1 | 3 hrs/week

1. Get and Clean data using dplyr exercises.
2. Visualize all Statistical measures(Mean ,Mode, Median, Range, InterQuartile Range etc.,using Histograms, Boxplots and Scatter Plots).
3. Create a data frame with atleast 10 entries of columns EMPID,EMPNAME,SALARY,STARTDATE

   a. Extract two column names using column name.
   b. Extract the first two rows and then all columns.
   c. Extract 3rd and 5th row with 2nd and 4th column.

4. Create a data frame with 10 observations and 3 variables and add new rows and columns to itusing 'rbind' and 'cbind' function.
5. Create a function to discretize a numeric variable into 3 quantiles and label them as low, medium, and high. Apply it on each attribute of any dataset to create a new data frame. 'discrete' with Categorical variables and the class label.
6. Create a simple scatter plot using any dataset using 'dplyr' library. Use the samedata to indicate distribution densities using box whiskers.
7. Write R Programs to implement k-means clustering, k-medoids clustering and density based clustering on any datasets.
8. Write a R Program to implement decision trees using 'reading Skills' dataset.
9. Implement decision trees using any dataset using package party and 'rpart'.
10. Generate top 5 association rules using apriori.
11. Generate top 5 association rules using ECLAT.
12. Write an R program to implement Naïve bayes Classification.

SEMESTER-III

COURSE 8: WEB TECHNOLOGIES

Theory | Credits: 3 | 3 hrs/week

COURSE OBJECTIVES: This subject enables the student to create flexible, attractive, user-friendlyweb sites comprised of both static and dynamic web pages.. Along with that students will also learn about interactions with web pages through JavaScript and host own web site on internet.

LEARNING OUTCOMES: After Studying this subject students would have capability to make theirown web site and host on internet. Also students would have enough knowledge about the technologies used in internet.

UNIT I : HTML: Basic HTML Tags and Attributes, Document body, Text, Hyper links, Adding moreFormatting, Lists, Tables, Grouping, Images. More HTML: Multimedia Objects, Frames, Forms, Headers.

UNIT II : Cascading Style Sheets: Introduction, Syntax, Selectors, Background Cursors, Text Fonts,Lists, Tables, Box Model, Using Styles, Simple Examples, Creation of Own Styles, Properties And Values In Styles, Formatting Blocks of Information, Layers.

UNIT III : Introduction to JavaScript: What is DHTML, JavaScript Basics, Variables, String Manipulations, Mathematical Functions, Statements, Operators, Arrays and Functions.

UNIT IV : DHTML with JavaScript: Data Validation, Opening A New Window, Messages andConfirmations, Status Bar, Different Frames, Rollover Buttons, Moving Images.

UNIT V : XML: Defining Data for Web Applications, Basic XML, Document Type Definition,Presenting XML, Document Object Model, Web Services.

TEXT BOOKS: 1. Harvey M. Deitel and Paul J. Deitel, "Internet & World Wide Web How to Program", 4/e, Pearson Education. 2. Uttam Kumar Roy,WebTechnologies from Oxford UniversityPress Student Activities.

Co-curricular Activities:

- We for Web – Students with right mix of skills are formed as groups to develop websites.
- Web Ninja- A platform to showcase creative websites developed by students to their peers.

Assessment Methods:

- Formal Examinations .
- Lab Practical Examination .
- Presentations .
- Simple Project.

SEMESTER-III

COURSE 8: WEB TECHNOLOGIES

Practical | Credits: 1 | 2 hrs/week

1. Design web pages for your college containing a description of the courses, departments, faculties, library etc, use href, list tags.
2. Create your class timetable using table tag.
3. Create a feedback form for your curriculum. Use textbox, text area, checkbox, radio buttonetc
4. Create a web page using frame. Divide the page into two parts with Navigation links on lefthand side of page (width=20%) and content page on right hand side of page (width = 80%). Onclicking the navigation Links corresponding content must be shown on the right hand side.
5. Write html code to develop a webpage having two frames that divide the webpage into two equal rows and then divide the row into equal columns fill each frame with a different background colour.
6. Create your resume using HTML tags. Experiment with colours, text, link, size and also other tags you studied.
7. Design a web page of your College Day Celebrations with an attractive background colour,text colour, images, font etc. Use CSS.
8. Use Inline CSS to format your resume that you created.
9. Use External CSS to format your class timetable as you created.
10. Use External, Internal, and Inline CSS to format web page of your start up.
11. Develop a JavaScript to display your admission details in the college.
12. Develop simple calculator for addition, subtraction, multiplication and division operationusing JavaScript.
13. Create HTML page with JavaScript which takes integer number as input and tells whetherthe number is odd or even.
14. Create HTML page that contains form for registration of your participation in a hackathon.Use relevant fields for input data. Write a JavaScript code to combine and display the input information when the button is clicked.
15. Create a login form with id and password. Perform input validation

SEMESTER-IV

COURSE 9: DATA VISUALIZATION

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:

- To know the importance of data Visualization in the world of DataAnalytics and Prediction
- To know the important libraries in Tableau
- To get equipped with Tableau Tool

Learning outcomes of Course:

- Students should be able to visualize data through seven stages of data analysisprocess
- Should be able to do explanatory and hybrid types of data visualization
- Should be able to understand various stages of visualizing data

UNIT I:

Creating Visual Analytics with tableau desktop, connecting to your data-How to Connect to your data, What are generated Values? Knowing when to use a direct connection, Joining tables with tableau, blending different data sources in a single worksheet.

UNIT II:

Building your first Visualization- How Me works- Chart types, Text Tables, Maps, bar chart, Line charts, Area Fill charts and Pie charts, scatter plot, Bullet graph, Gantt charts, Sorting data in tableau, Enhancing Views with filters, sets groups and hierarchies.

UNIT III:

Creating calculations to enhance your data- What is aggregation, what arecalculated values and table calculations, Using the calculation dialog box to create,Building formulas using table calculations, Using table calculation functions **UNIT IV:**

Using maps to improve insights-Create a Standard Map View, Plotting your ownlocations on a map, Replace Tableau's standard maps, Shaping data to enable Point-to-Point mapping.

UNIT V:

Developing an Adhoc analysis environment- generating new data with forecasts, providing self evidence adhoc analysis with parameters, Editing views in tableau Server.

TEXT BOOKS:

1. Tableau your data-Daniel G. Murray and the Inter works BI team, Wiley Publications
2. Tableau Data Visualizaton Cookbook, AshutoshNandeshwar, PACKT publishing.
3. Storytelling with Data: A Data Visualization Guide for BusinessProfessionals by Cole NussbaumerKnaflic (2014)
4. ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham (2009)

REFERENCE BOOKS:

1. Designing Data Visualizations: Representing Informational Relationshipsby Noah Iliinsky, Julie Steele (2011)
2. Alexandru C. Telea – "Data Visualization principles and practice" SecondEdition, CRC Publications
3. Joshua N. Millign–" Learning Tableau -2019" – Third Edition- Packt publications

Student Activity

Create a sample super store data set and visualize the following requirements

General Requirements

1. Dashboard size is 1250px wide by 750px tall.
2. Prefer using containers
3. The dashboard has a total of 5 containers (no more, no less)
4. The Filter Pane
5. Each filter has some padding

1. Charts Pane Requirement

1. All 3 charts must be in one vertical container
2. Do proper formatting
3. Each chart has some padding between them and other objects
4. Each chart has a grey border, slightly darker than the Pane background color.
5. The Pane under the Title has a border

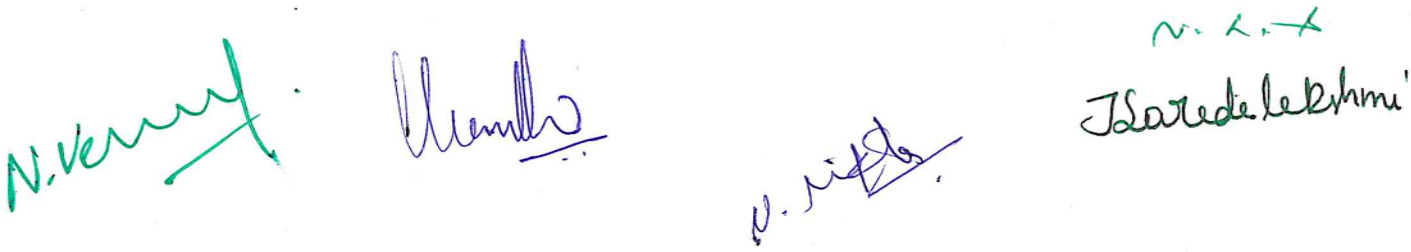
2. The second graph should have the title as "Sales" and should showmonthly sales per year. Make sure it is an area chart with proper formatting.

3. The third graph should the title as "Profit" and should show monthlyprofit per year. Make sure it is an area chart with proper formatting.

Continuous assessment:

Let the students be tested in the following questions from each unit

1. What are generated values? Join tables using Tableau
2. Create any visualization charts using Chart types, Text Tables, Maps, barchart, Line charts, Area Fill charts and Pie charts, scatter plot etc.,
3. What is aggregation, what are calculated values and table calculations?
4. Using Standard Map View, Plot your own locations on a map
5. Develop an Adhoc analysis environment.



SEMESTER-IV

COURSE 9: DATA VISUALIZATION

Practical | Credits: 1 | 2 hrs/week

1. Connect to data Sources
2. Create Univariate Charts
3. Create Bivariate and Multivariate charts
4. Create Maps
5. Calculate user-defined fields
6. Create a workbook data extract
7. Save a workbook on a Tableau server and web
8. Export images, data.

SEMESTER-IV

COURSE 10: DATA VISUALIZATION USING PYTHON

Theory | Credits: 3 | 3 hrs/week

Course Objective :

This course introduces students to data analysis and visualization inthe field of exploratory data science using Python.

Course Learning Outcomes : On successful completion of thecourse, the students will be able to

1. Use data analysis tools in the pandas library.
2. Load, clean, transform, merge and reshape data.
3. Create informative visualization and summarize data sets.
4. Analyze and manipulate time series data.
5. Solve real world data analysis problems.

Unit 1

Introduction: Introduction to Data Science, Exploratory Data Analysisand Data Science Process. Motivation for using Python for Data Analysis, Introduction of Python Jupyter Notebook. Essential Python Libraries: NumPy, pandas, matplotlib, SciPy, scikit-learn, statsmodels, seaborn.

Unit 2

Getting Started with Pandas: Arrays and vectorized conputation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. Data Loading, Storage and File Formats. Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, Interacting with Web APIs,

Interacting with Databases Data Cleaning and Preparation. HandlingMissing Data, Data Transformation, String ManipulationUnit 3

Data Wrangling: Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. Data Visualization matplotlib: Basics of matplotlib, plotting with pandas and seaborn, other pythonvisualization tools. Advanced categorical and numeric plots.

Unit 4

Data Aggregation and Group operations: Group by Mechanics, Dataaggregation, General split-apply-combine, Pivot tables and cross tabulation

Time Series Data Analysis: Date and Time Data Types and Tools,Time series Basics, date Ranges, Frequencies and Shifting, Time Zone Handling, Periods and Periods Arithmetic, Resampling and Frequency conversion, Moving Window Functions.

Unit 5 Advanced Pandas:

Categorical Data: cleaning data and visualization techniques, Advanced GroupBy methods .Use Techniques for Method Chaining.**Textbook:**

1. McKinney, W.(2017). Python for Data Analysis: Data Wranglingwith Pandas, NumPy and IPython. 2nd edition. O'Reilly Media.

Reference:
1. O'Neil, C., & Schutt, R. (2013). Doing Data Science: Straight Talkfrom the Frontline O'Reilly Media.

SEMESTER-IV

COURSE 10: DATA VISUALIZATION USING PYTHON

Practical | Credits: 1 | 2 hrs/week

1. Practicals based on NumPy ndarray
2. Practicals based on Pandas Data Structures
3. Practicals based on Data Loading, Storage and File Formats
4. Practicals based on Interacting with Web APIs
5. Practicals based on Data Cleaning and Preparation
6. Practicals based on Data Wrangling
7. Practicals based on Data Visualization using matplotlib
8. Practicals based on Data Aggregation
9. Practicals based on Time Series Data Analysis

SEMESTER-IV

COURSE 11: INTRODUCTION TO SQL & ADVANCED TABLEAU

Theory Credits: 3 3 hrs/week

Learning Objectives:

- ✓ Design a database using DBMS softwares.
- ✓ Perform SQL queries on database.
- ✓ Use Tableau's visualization tools to conduct data analysis,especially exploration of an unfamiliar dataset.

Course Outcomes:

- ✓ Design a database by their own and perform simple and adhocqueries.
- ✓ Employ best practices in data visualization to develop charts, maps,tables, and other visual representations of data.
- ✓ Employ best practices in data visualization to develop charts, maps,tables, and other visual representations of data.
- ✓ Create compelling, interactive dashboards to combine severalvisualizations into a cohesive and functional whole.
- ✓ Utilize advanced Tableau features including parameters, datablending, custom SQL, very
- ✓ large data

UNIT I:

Overview of Database Management System: Introduction to data,information, database, database management system , DBMS software's,

, keys in DBMS. the building blocks of an entity relationship diagram, classification of entity sets, attribute classification, relationship degree,relationship classification, reducing ER diagram to tables,

UNIT 2:

Structured Query Language: Introduction, History of SQL Standard, Commands in SQL, Data Types in SQL, Data Definition Language, DataManipulation Language, database constraints, Aggregate functions, Join Operation, Set Operations, Views. SQL queries, sub queries and corelated queries,

Unit 3 : Optimal visualization types – bar chart, pie chart, gantt chart, bubble chart, bullet chart, scatter plot, line chart, heat map, tree map Maps- geographical locational plotting, Binning values , Calculated fields , Tablecalculations , Level of Detail calculations.

Unit 4 : Dashboard development, Dashboard design principles, dashboard interactivity, Connected "drill-down" dashboardsBest Practices, Creating visualizations with Tableau.

Unit 5 : Advanced Tableau, Large datasets, Fiscal Year Calculations , Parameters, tableau scripting, tableau server, integration of tableau with Rprogramming.

Textbooks:

1. Show me the Numbers: Designing Tables and Graphs toEnlighten by Stephen Few
2. The Data Loom: Weaving Understanding by ThinkingCritically and Scientifically with Data by Stephen Few

Reference Books:

1. The Big Book of Dashboards: Visualizing your Data using Real-World Business Scenarios by Steve Wexler, Jeffrey Shaffer, andAndy Cotgreave

SEMESTER-IV

COURSE 11: INTRODUCTION TO SQL & ADVANCED TABLEAU

Practical | Credits: 1 | 2 hrs/week

DATABASE MANAGEMENT SYSTEM LAB

Consider following databases convert entities and relationships torelation table for a given scenario.

1. **COLLEGE DATABASE:**

   STUDENT (stno, SName, Address, Phone, Gender)course(courseid, Sem, Sec)

   CLASS (stno, courseid)

   SUBJECT (Subcode, Title, Sem, Credits)

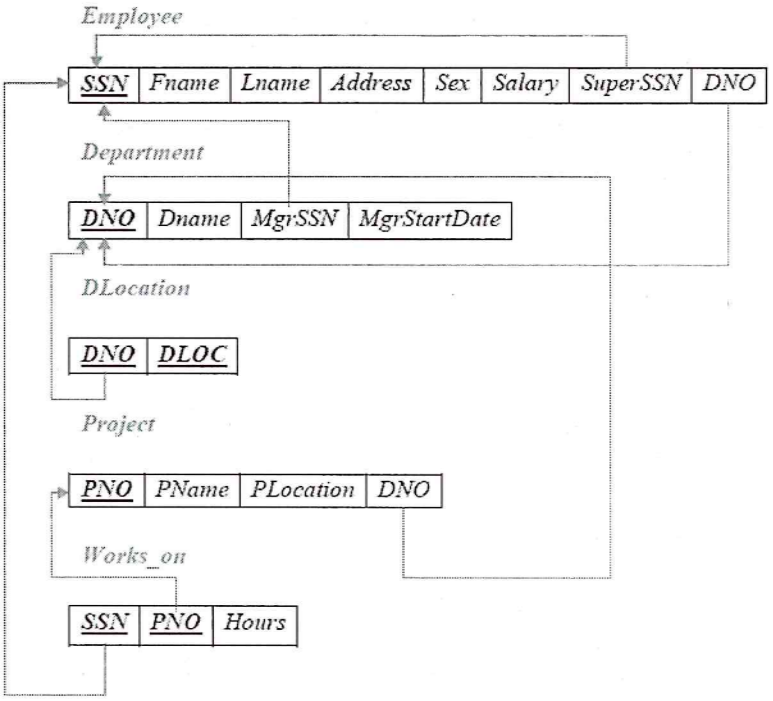
   MARKS (stno, Subcode, courseid, Test1, Test2, Test3, total)

2. **COMPANY DATABASE:**

   EMPLOYEE (SSN, Name, Address, Sex, Salary, SuperSSN, DNo)DEPARTMENT (DNo, DName, MgrSSN, MgrStartDate) DLOCATION (DNo,DLoc)

   PROJECT (PNo, PName, PLocation, DNo)WORKS_ON (SSN, PNo, Hours)

3. **Consider a college database schema**



a. Create above tables with relevant Primary Key, Foreign Key and otherconstraints

b. Populate the tables with data

4. Perform queries to generate outputs:

1. Display all the details of all employees working in the company.
2. Display ssn, lname, fname, address of employees who work indepartment no 7.
3. Retrieve the Birthdate and Address of the employee whose name is'Franklin T.Wong'
4. Retrieve the name and salary of every employee.
5. Retrieve all distinct salary values
6. Retrieve all employee names whose address is in 'Bellaire'
7. Retrieve all employees who were born during the 1950s
8. Retrieve all employees in department 5 whose salary is between50,000 and60,000 (inclusive)

5. Perform the following queries

1.Retrieve the names of all employees who do not have supervisors2.Retrieve SSN and department name for all employees

3. Retrieve the name and address of all employees who work for the 'Research'department

4. For every project located in 'Stafford', list the project number, thecontrolling department number, and the department manager's last name, address, and birth date.

5. For each employee, retrieve the employee's name, and the name of hisor her immediate supervisor.

6. Retrieve all combinations of Employee Name and Department Name7.Make a list of all project numbers for projects that involve an
employee whose last name is 'Narayan' either as a worker or as amanager of the department that controls the project.

8.Increase the salary of all employees working on the 'ProductX' projectby 15%. Retrieve employee name and increased salary of these employees. 9.Retrieve a list of employees and the project name each works in,ordered by the employee's department, and within each department ordered alphabetically byemployee first name.

10.Select the names of employees whose salary does not match withsalary of any employee in department.

6. Perform following queries :

1. Retrieve the employee numbers of all employees who workon project located in Bellaire, Houston, or Stafford.

2. Find the sum of the salaries of all employees, the maximumsalary, the minimum salary, and
the average salary. Display with proper headings

3. Find the sum of the salaries and number of employees of allemployees of the 'Marketing' department, as well as the maximum salary, the minimum salary, and the average salaryin this department.

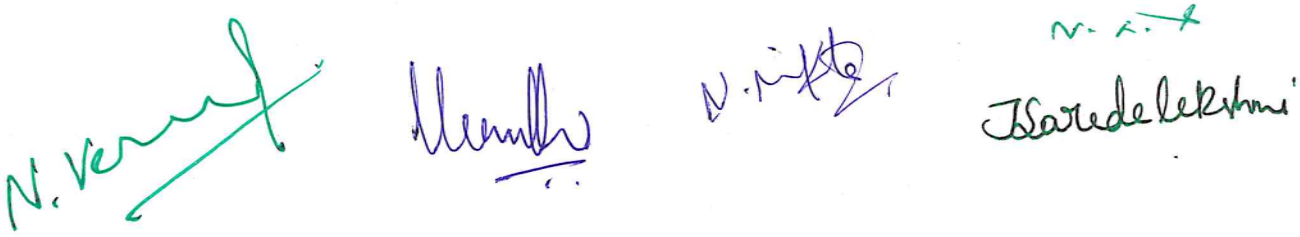
4. Select the names of employees whose salary is greater thanthe average salary of all employees in department 10.

5. Delete all dependents of employee whose ssn is'123456789'.

6. Perform a query using alter command to drop/add field and aconstraint in Employee table.

7. Format your data using filters with colors

8. create dashboards and stories.
9. Distribute and publish your visualization.
10. create advanced mapping –
    1. point-to-point map
    2. Dual axis map
11. Calculate distance between two points on a map.



KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours Data Science[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING

Time : 3 hrs *Max marks : 70*

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **FIVE** out of the following **EIGHT** questions*

1. What are the primary uses of R in data science? (Unit I)
2. Explain the facets of data . (Unit I)
3. Discuss the steps involved in setting research goals in the data science process. (Unit II)
4. Discuss the significance of data exploration and modeling in the Data Science process. (UNIT II)
5. Discuss the applications of machine learning in data science (Unit III)
6. Discuss the programming tips for dealing with large datasets in Data Science. (Unit iv)
7. Describe vectorized operations in R. (Unit V)
8. What are the scoping rules in R? How do they work?(Unit V)

[**Note:** *Question Paper setters are instructed to give one question from each unit compulsory.]*

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **FIVE** out of the following **TEN** questions*

8. (a). Define Data Science and Big Data. Discuss the benefits and uses of Big Data.(Unit I)

   OR

   (b) Explain the overview of the R programming language and its applications in data science. (Unit I)

9. (a) Discuss the steps involved in the data science process (Unit II)

   OR

   (b) Describe various techniques for retrieving data in R. How can data be imported and exported using R? (Unit II)

10. (a) What is machine learning? Explain the modeling process in machine learning. (Unit III)

OR

(b) Explain in detail supervised and unsupervised learning. (Unit III)

11. (a) Discuss the challenges faced when handling large datasets on a single computer. (Unit IV)

OR

(b) What are the general techniques for handling large volumes of data? (Unit IV)

12. (a) Describe the control structures and functions in R. (Unit V)

OR

(b) Discuss loop functions and debugging techniques in R. Provide examples to illustrate their usage. (Unit V)

[Note: *Question Paper setters are instructed to give two questions from each unit compulsory.]*

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours Data Science[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Introduction to Data Science and R Programming

Question Bank

Short Questions (20)

1. Define Data Science.
2. Explain the concept of Big Data.
3. What are the benefits of utilizing Big Data?
4. List and explain the facets of data.
5. What is the data science process? Briefly describe its steps.
6. Discuss the history and overview of the R programming language.
7. What are the primary uses of R in data science?
8. Define machine learning. Why is it important in data science?
9. Differentiate between supervised and unsupervised learning.
10. What challenges arise when handling large datasets on a single computer?
11. Describe the general techniques used for handling large volumes of data.
12. Explain vectorized operations in R.
13. How does the dplyr package help manage data frames in R?
14. What steps are involved in setting research goals in the data science process?
15. Discuss the control structures in R programming.
16. Explain the concept of functions in R.
17. What are the scoping rules in R? How do they work?
18. Define coding standards in R programming.
19. What are loop functions in R? Provide examples.
20. Explain the process of debugging in R.

Long Answers (20)

1. Define Data Science and Big Data. Discuss the benefits and uses of Big Data.
2. Explain the facets of data and the data science process.
3. Describe the history and overview of the R programming language. How is it used in data science?
4. Discuss the importance of machine learning in data science. Provide examples of its applications.
5. What are the challenges faced when handling large data on a single computer? How can these challenges be addressed?
6. Describe vectorized operations in R. How do they enhance data manipulation?
7. Explain the steps involved in setting research goals in the data science process.
8. Discuss control structures, functions, and scoping rules in R programming.
9. Explain the modeling process in machine learning. Differentiate between supervised and unsupervised learning.
10. Describe the techniques for retrieving data in R. How can data be imported and exported using R?

11. Discuss the challenges faced when handling large datasets on a single computer. What are some strategies for overcoming these challenges?
12. Describe the process of subsetting R objects. How can vectorized operations be performed in R?
13. Explain loop functions in R. Provide examples to illustrate their usage.
14. Discuss the benefits and limitations of using the dplyr package for managing data frames in R.
15. Explain the importance of data preparation in the data science process. What steps are involved?
16. Define coding standards in R programming. Why are they important?
17. Discuss the steps involved in the data science process. How does R facilitate these steps?
18. Explain the scoping rules in R programming. How do they impact variable accessibility?
19. Describe the process of debugging in R. What are some common debugging techniques?
20. Discuss the applications of machine learning in data science. How does it contribute to data analysis and decision-making?

KRISHNA UNIVERSITY :: MACHILIPATNAM - 521004

B.Sc. Honours Data Science (Major)

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

BLUE PRINT

INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	24
-III	1	2	28
IV	1	2	28
V	2	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours Data Science[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

Descriptive Statistics

Time : 3 hrs *Max marks : 70*

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **FIVE** out of the following **EIGHT** questions*

1. Explain the importance of statistics in various fields.
2. Describe the concepts of primary and secondary data.
3. Calculate the range and quartile deviation for the following dataset: {10, 12, 14, 16, 18, 20}.
4. Compute the standard deviation and variance for the dataset: {25, 30, 35, 40, 45}.
5. Explain the principle of least squares in curve fitting.
6. Define correlation and Rank Correlation Coefficien.
7. Discuss the concept of regression and its applications.
8. Explain Association of attributes and its measures,

SECTION-B

Answer ALL questions. Marks: 5X10=50M

9. (a) Explain the measures of central tendency with suitable examples.

OR

(b) Illustrate the process of representing data using histograms and frequency polygons.

10. a)Define skewness and kurtosis. Discuss their significance in statistics.

Or

b) Explain measures of dispersion.

11. (a) Discuss the fitting of a straight line and a second-degree polynomial (parabola) to a given dataset. Provide examples.

OR

(b) Calculate Karl Pearson's coefficient of correlation for the following data and interpret the result: X: {10, 20, 30, 40, 50} Y: {15, 25, 35, 45, 55}

12. (a) Explain linear regression and its properties. Provide an example demonstrating the calculation of regression coefficients.

OR

(b) Differentiate between correlation and regression. Discuss the applications of logistic regression.

13. (a) Define attributes and describe the conditions for the consistency of data for two and three attributes. Provide examples.

OR

(b) Explain the significance of the contingency table. Calculate the square contingency for the given data and interpret the result.

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

DESCRIPTIVE STATISTICS

Short Questions:

1. What is the scope of statistics in different fields? (Unit I)
2. Define primary and secondary data. (Unit I)
3. Explain the importance of histograms in data representation. (Unit I)
4. Compute the mean for the dataset: {5, 7, 9, 11, 13}. (Unit I)
5. What is the formula for calculating standard deviation? (Unit II)
6. Define skewness and provide an example of a skewed distribution. (Unit II)
7. Explain the principle of least squares in curve fitting. (Unit III)
8. Define correlation and its types. (Unit III)
9. Discuss the significance of regression analysis. (Unit IV)
10. What are the conditions for consistency of data in attributes? (Unit V)
11. Define independence of attributes. (Unit V)
12. Explain the concept of a contingency table. (Unit V)
13. What are the measures of central tendency? (Unit I)
14. Define quartile deviation. (Unit II)
15. Discuss the fitting of a second-degree polynomial to a given dataset. (Unit III)
16. Differentiate between correlation and regression. (Unit IV)
17. What are central and non-central moments? (Unit II)
18. Explain the concept of association of attributes. (Unit V)
19. Describe the relationship between association and colligation of attributes. (Unit V)
20. Define the coefficient of mean square contingency. (Unit V)

Long Questions:

1. Discuss the importance of statistics in various fields with examples. (Unit I)
2. Explain the process of representing data using histograms and frequency polygons. (Unit I)
3. Calculate the median and mode for the dataset: {8, 10, 12, 14, 16, 18}. (Unit I)
4. Compute the variance for the dataset: {20, 25, 30, 35, 40}. (Unit II)
5. Define kurtosis and discuss its implications in data analysis. (Unit II)
6. Discuss the fitting of a straight line to a given dataset. Provide an example. (Unit III)
7. Explain the scatter diagram method of correlation. (Unit III)
8. Discuss the concept of regression and its applications in real-life scenarios. (Unit IV)
9. Define regression coefficients and explain their significance. (Unit IV)
10. Describe the conditions for consistency of data for two and three attributes. Provide examples. (Unit V)
11. Explain the association of attributes and its measures. (Unit V)
12. Calculate the square contingency for a given contingency table and interpret the result. (Unit V)
13. Discuss the relationship between skewness and central tendency measures. (Unit II)
14. Explain the process of curve fitting using the principle of least squares. (Unit III)
15. Discuss the concept of rank correlation coefficient. (Unit III)
16. Explain the concept of attributes and their classification. (Unit V)
17. Discuss the significance of the correlation coefficient in statistical analysis. (Unit III)
18. Explain the importance of standard deviation as a measure of dispersion. (Unit II)
19. Discuss the significance of regression analysis in predictive modeling. (Unit IV)

20. Compare and contrast the different measures of central tendency. (Unit I)

KRISHNA UNIVERSITY :: MACHILIPATNAM - 521004

B.Sc. Honours Data Science (Major)

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 ADMITTED BATCH

BLUE PRINT

DESCRIPTIVE STATISTICS

Unit	S.A.Q **Section :A** (including choice) 4 Marks	E.Q **Section: B** (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	1	2	24
III	2	2	28
IV	1	2	28
V	2	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III PYTHON PROGRAMMING FOR DATA ANALYSIS

MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **five** out of the following **eight** questions*

1. What is Data Analysis? How does it differ from Data Analytics?
2. Explain the basic functionality of NumPy arrays
3. How is file input and output handled with arrays in NumPy?
4. How do you summarise and compute descriptive statistics in Pandas?
5. How do you get started with Pandas? Discuss the basic data structures in Pandas.
6. What are the different data transformations available in Pandas?
7. How is string manipulation handled in Pandas?
8. How do you create model descriptions with Patsy in Python?

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **five** out of the following **ten** questions*

9. Discuss the Python Language Basics in detail.
10. Explain the concept of a Library in Python. Why is it important in Data Analysis?
11. Discuss the built-in data structures in Python.
12. What are Universal Functions in NumPy? Give examples.
13. Discuss how Pandas interacts with Web APIs and Databases.
14. How do you read and write data in text format and binary data formats in Pandas?
15. What is hierarchical indexing in Pandas? How is it used?
16. Discuss combining and merging datasets, and reshaping and pivoting in Pandas.
17. Discuss the matplotlib API primer for plotting and visualisation.
18. How do you plot with Pandas and Seaborn in Python?

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III PYTHON PROGRAMMING FOR DATA ANALYSIS

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	1	2	24
II	2	2	28
III	2	2	28
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III INFERENTIAL AND APPLIED STATISTICS
MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **five** out of the following **eight** questions*

1. Define the terms: Population, Sample, Parameter, and Statistic.
2. What is Sampling Distribution? Discuss the concept of Standard Error.
3. What is the Theory of Estimation? Discuss the concept of Hypothesis.
4. Discuss Unbiasedness, Consistency, Efficiency, and Sufficiency.
5. Discuss the Test for Goodness of Fit and Independence of Attributes.
6. What is the F-test for Equality of Variances?
7. How is the Trend estimated by Free Hand Curve Method and Method of Semi Averages?
8. Discuss the measures of different Mortality and Fertility rates, and Measurement of Population Growth.

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **five** out of the following **ten** questions*

9. Discuss the Law of Large Numbers and the Central Limit Theorem.
10. Explain Student's t-distribution, F-Distribution, and $\chi 2$-Distribution. Discuss their definitions, properties, and applications.
11. How are Binomial, Poisson, and Normal Population parameters estimated by the MLE method? Explain these distributions with examples
12. Explain the concepts of Statistical Hypotheses, Null and Alternative Hypothesis, Critical Region, Two Types of Errors, Level of Significance, and Power of a Test.
13. Discuss Non-parametric tests, their advantages and disadvantages, and comparison with parametric tests.
14. Explain the Measurement Scale: Nominal, Ordinal, Interval, and Ratio
15. Discuss the Determination of Trend by Least Squares (Linear Trend, Parabolic Trend only), and Moving Averages Method.
16. Discuss the importance of Time Series in statistical analysis.
17. What are Life Tables? Discuss the construction and uses of Life Tables.
18. Discuss the importance of Vital Statistics in statistical analysis.

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III INFERENTIAL AND APPLIED STATISTICS

BLUEPRINT

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	2	2	28
IV	1	2	24
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III DATA MINING TECHNIQUES USING R
MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **five** out of the following **eight** questions*

1. Explain data mining task primitives.
2. What is a Data Warehouse? Discuss the concept of Data Mining.
3. Explain about Dimensionality Reduction
4. Explain about Data mining query language
5. Discuss about Attribute-Oriented Induction (AOI)
6. Explain the general approach to solve a classification problem
7. Discuss Tree Pruning
8. Explain about multi-relational association rules

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **five** out of the following **ten** questions*

9. Explain about stages of the data mining process/KDD.
10. Explain about data mining techniques.
11. Discuss about Data cleaning in detail
12. Discuss about Data Transformation and Feature Selection
13. Discuss about Apriori Frequent Itemset Mining Method with an example
14. Discuss about FP Growth approach for mining frequent item sets with an example
15. Discuss Decision Tree Induction and Decision Tree Induction Algorithm.
16. Discuss Bayes Classification Methods.
17. What is Cluster Analysis? Discuss Partitioning Methods
18. Discuss about Hierarchical Methods and Density-Based Methods

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III DATA MINING TECHNIQUES USING R

BLUEPRINT

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	1	2	24
IV	2	2	28
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III WEB TECHNOLOGIES
MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **five** out of the following **eight** questions*

1. What are the types of heading in HTML? Explain
2. Explain HTML frames with an example
3. How to create layers of various divisions using CSS
4. What is the CSS Box Model? Explain.
5. Explain the concept of functions in JavaScript with example
6. Explain about Java Script Arrays
7. What are the four components of DHTML? Explain.
8. What are web services?

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **five** out of the following **ten** questions*

9. Discuss the various HTML elements involved in creating tables and provide an example
10. Define Form. Write a HTML code for creating a registration form by covering all the input elements in a form.
11. What is CSS? Explain types of CSS with examples
12. Explain about id, class, universal and grouping selectors in CSS
13. Explain various operators and data types available in java script with examples.
14. Discuss about Mathematical and String methods available in Java Script
15. Explain about DHTML events
16. Write a program using CSS with JavaScript in DHTML
17. What is Document Type Definition (DTD) and its types? Explain how a DTD is created with an example.
18. What is DOM in XML? Explain.

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - III WEB TECHNOLOGIES

BLUEPRINT

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	2	2	28
IV	1	2	24
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - IV DATA VISUALIZATION MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

*Answer any **five** out of the following **eight** questions*

1. What is Tableau Desktop and why is it used for creating visual analytics?
2. How do you connect to your data in Tableau?
3. How do you build your first visualisation in Tableau?
4. How do you sort data in Tableau?
5. What is aggregation in Tableau?
6. How do you use the calculation dialog box to create calculations in Tableau?
7. How do you use maps to improve insights in Tableau?
8. Discuss how to generate new data with forecasts in Tableau.

SECTION-B

(Long Answer Type Questions) *Marks* : 5x10M = 50M

*Answer any **five** out of the following **ten** questions*

9. Discuss the process of joining tables in Tableau.
10. How do you blend different data sources in a single worksheet in Tableau?
11. Discuss the different chart types available in Tableau.
12. Discuss how to enhance views with filters, sets, groups, and hierarchies in Tableau.
13. Discuss how to build formulas using table calculations in Tableau.
14. How do you use table calculation functions in Tableau?
15. How do you replace Tableau's standard maps?
16. Discuss how to shape data to enable Point-to-Point mapping in Tableau.
17. How do you edit views in Tableau Server?
18. i) How do you develop an ad-hoc analysis environment in Tableau?
    ii) How do you provide self-evidence ad-hoc analysis with parameters in Tableau?

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - IV DATA VISUALIZATION BLUEPRINT

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	28
III	2	2	28
IV	1	2	24
V	1	2	24
Total	8	10	132

KRISHNA UNIVERSITY :MACHILIPATNAM - 521004

B.Sc. Honours DATA SCIENCE

[Major / Minor]

REVISED CBCS FRAMEWORK WITH EFFECT FROM 2023-2024 admitted batch

SEMESTER - IV DATA VISUALIZATION MODEL QUESTION PAPER

Time : 3 hrs **Max marks : 70**

SECTION-A

(Short Answer Type Questions) *Marks* : 5 x 4 M = 20M

Answer any five out of the following eight questions

1. Define Database Management System (DBMS) and explain its significance in modern computing.
2. What are the key components of an Entity-Relationship (ER) diagram?
3. Explain the concept of database constraints and provide examples of different types of constraints.
4. What are aggregate functions in SQL?
5. Explain heat map and tree map.
6. Discuss level of detailed calculations.
7. What are the dashboard design principles?
8. Discuss the tableau scripting.

SECTION-B

Answer ALL questions.

9. (a) Define the concept of entity sets and attributes in database management. Explain the classification of entity sets and attribute classification.

OR

(b) Discuss the different degrees of relationship in entity-relationship diagrams and provide examples of each.

10. a) Explain join operation in SQL and its types with examples.

OR

b) Explain DDL and DML commands with syntax and examples.

11. (a) Describe the process of binning values and how it is used in data visualization. Provide examples of situations where binning values are useful.

OR

(b) Explain the significance of calculated fields and table calculations in Tableau. Provide examples of scenarios where each is used.

12. (a) Discuss the importance of dashboard interactivity in Tableau and explain how it enhances data analysis.

OR

(b) Explain the concept of connected "drill-down" dashboards and their benefits in data exploration.

13. (a) What are fiscal year calculations in Tableau? How are they implemented, and what are their applications?

OR

(b) Discuss the integration of Tableau with R programming and the advantages of combining these two tools for data analysis.

Blue print

INTRODUCTION TO SQL & ADVANCED TABLEAU

Unit	S.A.Q Section :A (including choice) 4 Marks	E.Q Section: B (including choice) 10 Marks	Marks Allotted
I	2	2	28
II	2	2	24
III	2	2	28
IV	1	2	28
V	1	2	24
Total	8	10	132