

## ANDHRA KESARI UNIVERSITY



### 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

## 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, Fancy Indexing 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 and cumprod) and sorting with sort method.
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 it by performing 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 Web APIs.
18. Implement the data Cleaning and Filtering methods (Use NA handling methods, 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 and multiple keys.

## SEMESTER-III

### COURSE 6: INFERENCE 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 of hypotheses 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,  $\chi^2$ -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 of Binomial, Poisson and Normal distributions.

#### UNIT III:

**Sample tests:** t-test for single mean, difference of means and paired t- test. 2. confidence intervals for mean(s). standard deviation(s) and correlation coefficient(s). Test for goodness of fit and independence of attributes. F-test for equality of variances.

**Non-parametric tests-** their advantages and disadvantages, comparison with parametric tests. Measurement scale- 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 of different 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. 7<sup>th</sup> edition.Pearson.

#### CO-CURRICULAR ACTIVITIES:

- Quiz Competition
- Expert Lectures
- Seminars

#### EXTRA CURRICULAR ACTIVITIES:

- Formal Examination
- Lab Practical
- Presentation
- Simple Projects

## SEMESTER-III

### COURSE 6: INFERENCE AND APPLIED STATISTICS

Practical

Credits: 1

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( $\chi^2$  - 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 and application.
- 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 Mining Process-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, Frequent Itemset Mining Methods: Apriori method, generating Association Rules, Improving the Efficiency of Apriori, Pattern-Growth Approach for mining Frequent Item sets.

UNIT-IV

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

UNIT-V

**Association rule mining: Antecedent, consequent , multi-relational association rules, ECLAT. Case study on Market Basket Analysis.**

**Cluster Analysis:** Cluster Analysis, Partitioning Methods, Hierarchical methods, Density based methods-DBSCAN.

#### TEXT BOOKS:

1. Jiawei Han, Micheline Kamber, Jian Pei. "Data Mining: Concepts and Techniques", 3<sup>rd</sup> Edition, Morgan Kaufmann Publishers, 2011.
2. Adelchi Azzalini, Bruno Scapa, "Data Analysis and Data mining", 2<sup>nd</sup> Edition, Oxford University 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", 10<sup>th</sup> Edition, Tata McGraw Hill Edition, 2007.
2. G.K. Gupta, "Introduction to Data Mining with Case Studies", 1<sup>st</sup> Edition, Eastern Economy Edition, PHI, 2006.

#### Student Activities:

1. Students should be able to implement Data Mining algorithms provided the relevant data
2. Given the data, students can visualize all statistical measures
3. Differentiate the types of mining problems and identify what type of algorithms are to be 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 3<sup>rd</sup> and 5<sup>th</sup> row with 2<sup>nd</sup> and 4<sup>th</sup> column.
4. Create a data frame with 10 observations and 3 variables and add new rows and columns to it using '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 same data 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-friendly web 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 their own 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 more Formatting, 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 and Confirmations, 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, Web Technologies from Oxford University Press 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 button etc
4. Create a web page using frame. Divide the page into two parts with Navigation links on left hand side of page (width=20%) and content page on right hand side of page (width = 80%). On clicking 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 operation using JavaScript.
13. Create HTML page with JavaScript which takes integer number as input and tells whether the 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 show monthly 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 monthly profit 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, bar chart, 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 in the field of exploratory data science using Python.

Course Learning Outcomes : On successful completion of the course, 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 Analysis and 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 computation, 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. Handling Missing Data, Data Transformation, String Manipulation

Unit 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 python visualization 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 several visualizations 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, Data Manipulation Language, database constraints, Aggregate functions, Join Operation, Set Operations, Views. SQL queries, sub queries and correlated 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 to relation 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)

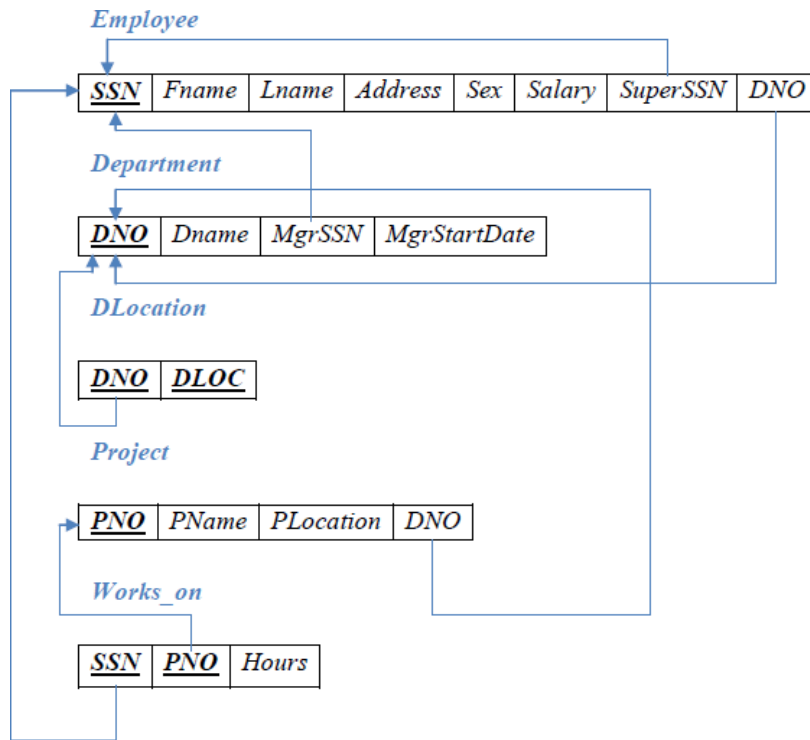
MARKS (stno, Subcode, courseid, Test1, Test2, Test3, total)

##### 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 other constraints

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 in department 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 between 50,000 and 60,000 (inclusive)

**5. Perform the following queries**

1. Retrieve the names of all employees who do not have supervisors  
 2. 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, the controlling 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 his or her immediate supervisor.
6. Retrieve all combinations of Employee Name and Department Name. Make a list of all project numbers for projects that involve an employee whose last name is 'Narayan' either as a worker or as a manager of the department that controls the project.
8. Increase the salary of all employees working on the 'ProductX' project by 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 by employee first name.
10. Select the names of employees whose salary does not match with salary of any employee in department.

## **6. Perform following queries :**

1. Retrieve the employee numbers of all employees who work on project located in Bellaire, Houston, or Stafford.
  2. Find the sum of the salaries of all employees, the maximum salary, the minimum salary, and the average salary. Display with proper headings
  3. Find the sum of the salaries and number of employees of all employees of the 'Marketing' department, as well as the maximum salary, the minimum salary, and the average salary in this department.
  4. Select the names of employees whose salary is greater than the 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 a constraint 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.

ANDHRA KESARI UNIVERSITY-ONGOLE, PRAKASAM DISTRICT  
Single Major Programme from the Year 2023-24 Onwards  
Programme-B.Sc. Honours Data Science - Question Paper model,  
Second Year-Semester III & IV

Time: 3 Hours

Total Marks: 75

PART –A

**Answer any Five of the following**

**5X5=25 Marks**

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.
- 9.
- 10

PART –B

**Answer the following**

**5x10=50 Marks**

- 11.
- 12.
- 13.
- 14.
- 15.
- 16.
- 17.
- 18.
- 19.
- 20.