

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.SC. HONOURS DATA SCIENCE

MAJOR (SCHEME)

w.e.f AY 2024-25 onwards

Year	Semester	Paper	Title of the Course	No. of Hrs./ Week	No. of Credits
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

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.SC. DATA SCIENCE HONOURS
III Semester w.e.f. 2024-25
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, Introduction to Python Language basics & and Jupyter Notebook.. Why Python for Data Analysis? Essential Python Libraries.

UNIT II:

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

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 Text Format, 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, Combining and Merging Datasets, Reshaping and Pivoting.

UNIT V:

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

Plotting and Visualization: A brief matplotlib API Primer, Plotting with Pandas and Sea born, 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 data analytics
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

B.Sc. Data Science Honours
III SEMESTER
COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

PRACTICALS

Credits: 1

2 hrs/week

List of Experiments

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.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III Semester
COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS
w.e.f. 2024-25

MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. **Short answer question from unit-I**
2. **Short answer question from unit-I**
3. **Short answer question from unit-II**
4. **Short answer question from unit-II**
5. **Short answer question from unit-III**
6. **Short answer question from unit-III**
7. **Short answer question from unit-IV**
8. **Short answer question from unit-IV**
9. **Short answer question from unit-V**
10. **Short answer question from unit-V**

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) **Essay question from Unit-I**
OR
b) **Essay question from Unit-I**
12. a) **Essay question from Unit-II**
OR
b) **Essay question from Unit-II**
13. a) **Essay question from Unit-III**
OR
b) **Essay question from Unit-III**
14. a) **Essay question from Unit-IV**
OR
b) **Essay question from Unit-IV**
15. a) **Essay question from Unit-V**
OR
b) **Essay question from Unit-V**

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. 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. 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.

UNIT III:

Sample tests: t-test for single mean, difference of means and paired t-test. 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.

UNIT IV:

Time Series: Time Series and its components with illustrations, additive, multiplicative models. Trends: Estimation of trend by 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. Life tables: construction and uses of life tables.

TEXT BOOKS:

- 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.
- K.V.S. Sarma: Statistics Made Simple: Do it yourself on PC.PHI.
- Fundamentals of applied statistics : VK Kapoor and SCGupta.
- 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

B.Sc. Data Science Honours
III Semester
Course 6: INFERENCE AND APPLIED STATISTICS

Practicals

Credits: 1

2 hrs/week

List of Experiments:

Sample Tests

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(paired samples).
6. Small sample test for single variance(χ^2 - test)
7. F-Test for variances.

Time Series

8. Measurement of trend by method of moving averages(odd and even period)
9. Measurement of trend by method of Least squares(linear and parabola)
10. Determination of seasonal indices by method simple averages
11. Determination of seasonal indices by method of Ratio to moving averages

Vital Statistics

12. Computation of various Mortality rates
13. Computation of various Fertility rates
14. Computation of various Reproduction rates.
15. Construction of Life Tables

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III Semester
Course 6: INFERENCE AND APPLIED STATISTICS
w.e.f. 2024-25
MODEL QUESTION PAPER

Time: 3 Hours

Max. Marks: 75

PART-A

Answer any FIVE of the following. Each Question Carries 5 marks.

(5X5=25)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any FIVE of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III Semester
COURSE 7: DATA MINING TECHNIQUES USING R
w.e.f. 2024-25

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 Item set Mining Methods: Apriority 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", 3rd Edition, Morgan Kaufmann Publishers, 2011.
2. Adelchi Azzalini, Bruno Scapa, "Data Analysis and Data mining" , 2nd 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”, 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 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.

B.Sc. Data Science Honours
III Semester
Course 7: DATA MINING TECHNIQUES USING R

Practicals

Credits: 1

2 hrs/week

List of Experiments:

1. Get and Clean data using dplyr exercises.
2. Visualize all Statistical measures(Mean ,Mode, Median, Range, Inter Quartile 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 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.
6. 'discrete' with Categorical variables and the class label.
7. Create a simple scatter plot using any dataset using 'dplyr' library. Use the same data to indicate distribution densities using box whiskers.
8. Write R Programs to implement k-means clustering, k-medoids clustering and density based clustering on any datasets.
9. Write a R Program to implement decision trees using 'reading Skills' dataset.
10. Implement decision trees using any dataset using package party and 'rpart'.
11. Generate top 5 association rules using apriori.
12. Generate top 5 association rules using ECLAT.
13. Write an R program to implement Naïve bayes Classification.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III Semester
Course 7: DATA MINING TECHNIQUES USING R
w.e.f. 2024-25
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. **Short answer question from unit-I**
2. **Short answer question from unit-I**
3. **Short answer question from unit-II**
4. **Short answer question from unit-II**
5. **Short answer question from unit-III**
6. **Short answer question from unit-III**
7. **Short answer question from unit-IV**
8. **Short answer question from unit-IV**
9. **Short answer question from unit-V**
10. **Short answer question from unit-V**

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) **Essay question from Unit-I**
OR
b) **Essay question from Unit-I**
12. a) **Essay question from Unit-II**
OR
b) **Essay question from Unit-II**
13. a) **Essay question from Unit-III**
OR
b) **Essay question from Unit-III**
14. a) **Essay question from Unit-IV**
OR
b) **Essay question from Unit-IV**
15. a) **Essay question from Unit-V**
OR
b) **Essay question from Unit-V**

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, WebTechnologies 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.

B.Sc. Data Science Honours
III Semester
Course 8: WEB TECHNOLOGIES

Practicals

Credits: 1

2 hrs/week

List of Experiments:

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 lefthand side of page (width=20%) and content page on right hand side of page (width
1. = 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

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
III Semester
Course 8: WEB TECHNOLOGIES
(w.e.f. 2024-25)
MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 5 marks.

(5X5=25)

1. **Short answer question from unit-I**
2. **Short answer question from unit-I**
3. **Short answer question from unit-II**
4. **Short answer question from unit-II**
5. **Short answer question from unit-III**
6. **Short answer question from unit-III**
7. **Short answer question from unit-IV**
8. **Short answer question from unit-IV**
9. **Short answer question from unit-V**
10. **Short answer question from unit-V**

PART-B

Answer any **FIVE** of the following. Each Question Carries 10 marks.

(5X10=50)

11. a) **Essay question from Unit-I**
OR
b) **Essay question from Unit-I**
12. a) **Essay question from Unit-II**
OR
b) **Essay question from Unit-II**
13. a) **Essay question from Unit-III**
OR
b) **Essay question from Unit-III**
14. a) **Essay question from Unit-IV**
OR
b) **Essay question from Unit-IV**
15. a) **Essay question from Unit-V**
OR
b) **Essay question from Unit-V**