ANDHRA PRADESH STATE COUNCIL OF HIGHER EDUCATION

MINOR

Subject: Data Science

w.e.f. AY 2023-

24 COURSE

STRUCTURE

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
II	II	1	Introduction to Data Science and R Programming	3	3
			Introduction to Data Science and R Programming Practical Course	2	1
	III	2	Python Programming for Data Analysis	3	3
			Python Programming for Data Analysis Practical Course	2	1
	IV	3	Data visualization using Tableau	3	3
			Data visualization using Tableau Practical Course	2	1
		4	Data visualization using python	3	3
			Data visualization using python Practical Course	2	1

COURSE 2: 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

COURSE 2: 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.

COURSE 3: 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.

COURSE 3: 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.

COURSE 4: 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.

COURSE 4: 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

ANDHRA KESARI UNIVERSITY-ONGOLE, PRAKASAM DISTRICT
Minor 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.