PARVATHANENI BRAHMAYYA
SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada–520010
Re-accredited at 'A+' by the NAAC

Course Code				**23DSMIL231**			
Title of the Course				Python for Data Analysis			
Offered to: (Programme/s)				**B.Sc. Honours (Statistics/ Computer Science)**			
L	**4**	**T**	**0**	**P**	**0**	**C**	**3**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**Minor**			**Course Relates to:**	**Global**		
Year of Revision:	**--**			**Percentage:**	**--**		
Type of the Course:				**Skill Development**			
Crosscutting Issues of the Course :				**Professional Ethics**			
Pre-requisites, if any				**Programming Logic**			

Course Description:

This course offers a detailed introduction to Data Science, emphasizing practical applications in Python. It covers key areas such as data exploration, cleaning, and visualization using libraries like NumPy, pandas, and matplotlib. Students will learn to handle and analyze data with pandas, create and manipulate arrays with NumPy, and visualize data with matplotlib and seaborn. The course also includes advanced topics such as data aggregation, group operations, and time series analysis, equipping students with the skills to manage and interpret complex datasets effectively. Ideal for those seeking a solid foundation in data science with hands-on Python experience.

Course Aims and Objectives:

S. NO	COURSE OBJECTIVES
1	Understand the basics of Data Science, including the data analysis process, and how to use Python tools like iPython and Jupyter Notebook for data analysis.
2	Apply NumPy to handle arrays and matrices, including creating, reshaping, and performing operations on them.
3	Use pandas to analyze data by cleaning, organizing, and exploring datasets, and create visualizations to gain insights.
4	Perform data wrangling by combining and reshaping datasets, and use visualization tools like matplotlib and seaborn to present data effectively.
5	Conduct advanced data analysis by aggregating and grouping data, and analyze time series data with techniques such as resampling and frequency conversion.

Course Outcomes

At the end of the course, the student will be able to…

CO NO	COURSE OUTCOME	BTL	PO	PSO
CO1	**Understand** the basics of Data Science, the data analysis process, and how Python tools like iPython and Jupyter Notebook can help.	**K2**	**PO1, PO5, PO6, PO7**	**PSO1, PSO2**
CO2	**Apply** NumPy to work with arrays and matrices, including creating, reshaping, and performing basic operations on them.	**K3**	**PO1, PO5, PO6, PO7**	**PSO1, PSO1**
CO3	**Analyze** data using pandas by cleaning, organizing, and exploring datasets, and create simple visualizations to understand the data better.	**K4**	**PO1, PO5, PO6, PO7**	**PSO1, PSO1**
CO4	**Evaluate** and combine data from different sources, reshape it, and use visualization tools like matplotlib and seaborn to gain insights.	**K5**	**PO1, PO5, PO6, PO7**	**PSO1, PSO1**
CO5	**Create** advanced data analysis techniques by working with grouped data, time series, and more complex operations to uncover deeper insights.	**K6**	**PO1, PO5, PO6, PO7**	**PSO1, PSO1**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1	**1**	**-**	**-**	**-**	**3**	**2**	**3**	**3**	**2**
CO2	**1**	**-**	**-**	**-**	**3**	**3**	**3**	**3**	**2**
CO3	**1**	**-**	**-**	**-**	**3**	**3**	**3**	**3**	**3**
CO4	**1**	**-**	**-**	**-**	**3**	**3**	**3**	**3**	**3**
CO5	**1**	**-**	**-**	**-**	**3**	**3**	**3**	**3**	**3**

Use the codes 3,2,1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure:

Unit – 1 : [Introduction to Python and Its Libraries] (12Hrs)

Introduction: Introduction to Data Science, Exploratory Data Analysis and Data Science Process. Motivation for using Python for Data Analysis, Introduction of Python shell iPython and Jupyter Notebook.

Essential Python Libraries: NumPy, pandas, matplotlib, SciPy, scikit-learn, statsmodels.

Examples/Applications/Case Studies:

- Write a Program to print “Hello World!” using Jupyter Notebook.
- Write a program to access the elements in a list using Jupyter Notebook.

Exercises/Projects:

- Write a program toDisplay Keys and values in a Dictionary using Jupyter Notebook.

Specific Resources: (web)

- W3Schools

Unit – 2 : [Numpy] (12Hrs)

Numpy: NumPy Arrays - difference between python lists and NumPy array, What is NumPy array, creating basic array, adding, removing and sorting elements, reshaping array, converting 1d array to 2d array, indexing and slicing, creating array from existing data, creating matrices, getting random numbers getting count and unique numbers, transposing and reshaping a matrix, reverse an array, reshaping multidimensional arrays.

Examples/Applications/Case Studies:

- Create a panda's series from a dictionary of values and andarray.
- Give an example to create a DataFrame from a singlendarray.

Exercises/Projects:

- Write a Pandas program to select the rows where thescore is missing, i.e. NaN.
- Write a program to generate a series of float numbersfrom 21.0 to 30.0 with an increment of 1.5 each.

Specific Resources: (web)

- W3Schools

Unit – 3 : [Pandas] (12Hrs)

Pandas: Introduction, Getting Started, Series, Data Frame**,** Read CSV, Read JSON **-** Analyzing DataFrames, Cleaning Data, Cleaning Empty Cell, Cleaning Wrong Format, Cleaning Wrong Data,Removing Duplicates, Correlations, Plotting.

Examples/Applications/Case Studies:

- Write a program to generate a series of the first 10 numbers.
- Write a Pandas program to count the number of rows and columns of a Data Frame.

Exercises/Projects:

- Write a program to generate a series andprint the top 3 elements using the headfunction.
- Write a program in Python to create aSeries in Python from the given dictionary.
  D= {"Jan" : 31, "Feb" : 28, "Mar" : 31}.

Specific Resources: (web)

- W3Schools

Unit – 4: [Data Wrangling and Data Visualization] (12Hrs)

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.

Examples/Applications/Case Studies:

- Plotting a line chart of date versus temperature by adding Label on X and Y axis, and adding a Title and Grids to the chart.
- Design a Plotting Histogram

Exercises/Projects:

- To plot a bar chart, we will specify kind='bar'. We can alsospecify the DataFrame columns to be used as x and y axes.Let us now add a column "Days" consisting of day names to"MelaSales.csv" .
- Use Matplotlib and Seaborn to create line charts, bar charts, and scatter plots from a given dataset. Focus on different types of visualizations to represent data effectively.

Specific Resources: (web)

- Matplotlib Documentation

- Seaborn Documentation
- Data Wrangling with Pandas

Unit – 5 : [Data Aggregation and Time Series Data Analysis] (12Hrs)

Data Aggregation and Group operations: Group by Mechanics, Data aggregation, 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.
Examples/Applications/Case Studies:

- How does the "pd.Series" function utilize the dates variable to create the time series data.
- What is the purpose of using the groupby function? How does it group the data and what is the result of applying the sum function to the grouped data?

Exercises/Projects:

- What does the pivot_table function do in this code? Explain how it transforms the DataFrame df and the purpose of the index, columns, and aggfunc parameters.

SpecificResources: (web)

- Data Aggregation Guide
- Time Series Analysis Tutorial
- Matplotlib Basics

Text Books/References:

- McKinney, W.(2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython. 2nd edition. O'Reilly Media
- O'Neil, C., & Schutt, R. (2013). Doing Data Science: Straight Talk from the Frontline O'Reilly Media

***

SEMESTER -END QUESTION PAPER STRUCTURE

Course Code & Title of the Course:	**23DSMIL231** Python for Data Analysis
Offered to:	**B.Sc. Hons(Statistics/ Computer Science)**
Category:	**SEMESTER: 3**
Max. Marks	**70**
Max.Time	**3 Hrs**

Section A: Short Answer Questions (20 Marks)

Answer All questions. Each question carries 4 Marks.

Q1 (a) Explain the motivation for using Python in Data Analysis.(K2)

OR

(b) Describe the steps in the Data Science process.(K2)

Q2 (a) How do you create a basic NumPy array? Provide an example.(K1)

OR

(b) What is the difference between Python lists and NumPy arrays? Explain with an example.(K1)

Q3 (a) How can you read a CSV file into a Pandas DataFrame?(K2)

OR

(b)Explain how to clean empty cells in a Pandas DataFrame.(K2)

Q4 (a) What is Hierarchical Indexing in Pandas? Explain its importance.(K2)

OR

(b) Describe how to plot data using Matplotlib.(K2)

Q5 (a) What is the purpose of the groupby function in Pandas?(K3)

OR

(b) Explain the basics of Time Series data analysis in Python.(K3)

Section B: Long Answer Questions (50 Marks)

Answer All questions. Each question carries 10 Marks.

Q6 (a) How is Python used in Data Science? Give examples of Jupyter Notebook features.(K3)

OR

(b) What are the main steps in the Data Science process? How do Python libraries help?(K3)

Q7 (a) What is the difference between Python lists and NumPy arrays? Show examples.(K3)

OR

(b) How do you perform basic operations like reshaping with NumPy arrays?(K3)

Q8 (a) Create and clean a DataFrame in Pandas? Explain with a CSV example.(K4)

OR

(b) Esplain how to manage missing values and duplicates in Pandas? (K4)

Q9 (a) What is Hierarchical Indexing in Pandas? How do you use Matplotlib for plotting?(K3)

OR

(b) How do you merge and reshape data in Pandas? Give examples.(K3)

Q10 (a) How does the groupby function work in Pandas? Provide an example. (K3)

OR

(b) What are the key techniques for analyzing time series data in Python?(K3))