PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE

Autonomous

Siddhartha Nagar, Vijayawada–520010

Re-accredited at 'A+' by the NAAC

Course Code				23AIMIP231			
Title of the Course				Python for Data Analysis Lab			
Offered to:				BCA Honours / B.Sc. Honours (CSCS)			
L	0	T	0	P	2	C	1
Year of Introduction:		2024-25		Semester:			3
Course Category:		Minor Lab		Course Relates to:		Global	
Year of Revision:		-		Percentage:		--	
Type of the Course:				Skill Development			
Crosscutting Issues of the Course :				-			
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. N O	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

This lab list covers the key areas of a ….(title of the course) course, providing hands-on practice with ….(technology/software)

Unit 1: [Introduction to Python and Its Libraries]   (6Hrs)

Lab 1:

1. Write a Program to print "Hello World!" using Jupyter Notebook.
2. Write a program to access the elements in a list using Jupyter Notebook.
- **Dataset** (web link) / **Experiment**: Learn basic operations in Python using Jupyter Notebook.
- **Tasks:**

1. **Hello World Program:**Write a Python program to print "Hello, World!" in Jupyter Notebook.
2. **List Element Access:** Write a Python program to access elements in a list using Jupyter Notebook.

Lab 2:

1. Write a program toDisplay Keys and values in a Dictionary using Jupyter Notebook.
- **Dataset**(web link) **/ Experiment**: Explore data structures in Python using Jupyter Notebook.
- **Tasks:**
  1. **Dictionary Keys and Values:**
     Write a Python program to display keys and values in a dictionary using Jupyter Notebook.

Unit 2: [**Numpy**]   (6Hrs)

Lab 3:

1. Create a panda's series from a dictionary of values and a ndarray.
2. Give an example to create a DataFrame from a single ndarray.
- **Dataset** (web link) **/ Experiment**: Utilize pandas to handle data structures in Python.
- **Tasks:**
  1. **Create pandas Series:**
     Write a Python program to create a pandas Series from a dictionary and anndarray.
2. **Create DataFrame:**
   Write a Python program to create a DataFrame from a single ndarray using pandas.

Lab 4:

1. Write a Pandas program to select the rows where the score is missing, i.e. NaN.
2. Write a program to generate a series of float numbers from 21.0 to 30.0 with an increment of 1.5 each.
- **Dataset**(web link) **/ Experiment**: Practice data manipulation and series creation in pandas.
- **Tasks:**
  1. **Select Rows with NaN:**
     Write a pandas program to select rows where the score is missing (NaN).
  2. **Generate Float Series:**
     Write a Python program to generate a series of float numbers from 21.0 to 30.0 with a 1.5 increment.

Unit 3: [Pandas]   (6Hrs)

Lab 5:

1. Write a program to generate a series of the first 10 numbers.
2. Write a Pandas program to count the number of rows and columns of a Data Frame.
- **Dataset** (web link) **/ Experiment**: Explore basic operations in pandas with series and DataFrames.
- **Tasks:**

1. **Generate Number Series:**
   Write a Python program to generate a series of the first 10 numbers.
2. **Count Rows and Columns:**
   Write a pandas program to count the number of rows and columns in a DataFrame.

Lab 6:

1. Write a program to generate a series and print the top 3 elements using the head function.
2. Write a program in Python to create a Series in Python from the given dictionary.
   D= {"Jan" : 31, "Feb" : 28, "Mar" : 31}.

- **Dataset**(web link) / **Experiment**: Perform basic operations with pandas Series.
- **Tasks:**
  1. **Top 3 Elements with head():**
     Write a Python program to generate a series and print the top 3 elements using the head() function.
  2. **Create Series from Dictionary:**
     Write a Python program to create a Series from the dictionary D = {"Jan": 31, "Feb": 28, "Mar": 31}.

Unit 4: [Data Wrangling and Data Visualization]   (6Hrs)

Lab 7:

1. 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.
2. Design a program to visualize data with a histogram.

- **Dataset** (web link) / **Experiment**: Create visual representations of data using matplotlib.
- **Tasks:**
  1. **Line Chart:**
     Plot a line chart of date versus temperature, adding labels on the X and Y axis, a title, and grids.
2. **Histogram Visualization:**
   Design a program to visualize data using a histogram.

Lab 8:

1. To plot a bar chart, we will specify kind='bar'. We can also specify 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".
2. 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.

- **Dataset**(web link) / **Experiment**: Enhance data visualization skills using Matplotlib and Seaborn.
- **Tasks:**
  1. **Bar Chart with Days Column:**
     Plot a bar chart specifying kind='bar', adding a "Days" column to MelaSales.csv for the X-axis.
  2. **Multiple Visualizations:**
     Use Matplotlib and Seaborn to create line charts, bar charts, and scatter plots, focusing on effective data representation.

Unit 5: [Data Aggregation and Time Series Data Analysis]   (6Hrs)
Lab 9:
1. How does the "pd.Series" function utilize the dates variable to create the time series data.
2. 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?
*   **Dataset** (web link) / **Experiment**:Explore time series creation and data grouping in pandas.
*   **Tasks:**
    1. **Time Series with pd.Series:**
        Utilize the dates variable with pd.Series to create time series data.
    2. **Groupby and Sum Function:**
        Understand the purpose of the groupby function, how it groups data, and the result of applying the sum function to grouped data.
Lab 10:
1. 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.
*   **Dataset** (web link) / **Experiment**:Learn to use pivot tables for data transformation in pandas.
*   **Tasks:**
    1. **Pivot Table Function:**
        Explain the pivot_table function in pandas, detailing how it transforms DataFramedf using the index, columns, and aggfunc parameters.

Lab Manual:

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

References:

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

Lab - Question Paper Pattern

23AIMIP231: Python for Data Analysis Lab

Offered to: BCA/ B.Sc. Honours (CSCS)

Max. Marks: 50   **Max. Time: 3Hrs**

Pass. Min: 20

(A)	**Evaluation Procedure**	**35 Marks**
I	Experiments (Exam & Execution)	**30 Marks**
II	Viva	**3 Marks**
III	Record	**2 Marks**
(B)	**CONTINUOUS ASSESMENT(Internal)**	**15 MARKS**

15 marks for the continuous assessment (Day to day work in the laboratory shall be evaluated for 15 marks by the concerned laboratory teacher based on the regularity/ record/viva). Laboratory teachers are mandated to ensure that every student completes 80%-90% of the lab assessments.

TOTAL: (A)+(B) =   **50 MARKS**