COURSE TITLE	**DATA ANALYTICS PROGRAMMING**
COURSE CODE	**05FN0502**
COURSE CREDITS	**2**

Objective:

1 To enable students to apply Python programming techniques to data analysis problems.
2 To train students in the use of libraries such as Pandas, NumPy and Matplotlib.
3 To understand and perform data cleaning, visualization, and statistical analysis.
4 To enhance students' ability to interpret data and derive meaningful insights.

Course Outcomes: After completion of this course, student will be able to:

1 Use Python libraries such as NumPy and Pandas to perform data manipulation tasks on structured and unstructured datasets.
2 Perform data cleaning and preprocessing techniques to prepare raw datasets for analysis.
3 Create effective data visualizations using Matplotlib and Seaborn to represent data insights.
4 Analyze datasets using statistical methods to identify patterns, trends, and meaningful insights.
5 Evaluate and implement complete data analytics solutions using real-world datasets to support decision-making.

Pre-requisite of course: Fundamental understanding of Python programming Basic knowledge of statistics and data formats Familiarity with spreadsheets or basic data analysis tools

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
0	0	4	0	30	20	50	50

Contents : Unit	Topics	Contact Hours
Total Hours		

DR. SUNIL LALCHAND BAJEJA
Digitally signed by (Name of HOD)

DR. RIJWAN KHAN
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Unit 1** Install Python and set up Jupyter Notebook or Google Colab. Write a program that uses different data types and prints their types, Write a Python program using loops to calculate the factorial of a given number using a user-defined function, Create a Python script to read a .csv file and display its content. Perform file I/O operations with .json and .xlsx formats as well, Develop a program that accepts a user input, writes it to a text file, and reads it back.	15
2	**Unit 2** Create NumPy arrays (1D, 2D) and perform basic operations like addition, subtraction, and element-wise multiplication, Demonstrate slicing and reshaping of NumPy arrays. Convert a 1D array into a 3x3 matrix, Use broadcasting to perform matrix addition with different shapes. Show how broadcasting works using examples, Apply NumPy aggregation functions (mean, max, min, std) and perform matrix multiplication using np.dot().	15
3	**Unit 3** Load a dataset (CSV) using Pandas and display basic information such as shape, head, and data types, Perform filtering on a dataset where age > 30, and select only specific columns like Name and Salary, Handle missing data using fillna() and dropna() functions. Show how to detect missing values, Group data by a specific column and compute aggregate values (like average salary by department).	15
4	**Unit 4** Write a Python program to detect and remove outliers from a dataset using the IQR method, Normalize and standardize a numeric column using Min-Max Scaling and Z-score techniques, Encode categorical variables using Label Encoding with sklearn.preprocessing.LabelEncoder, Convert categorical columns into dummy/indicator variables using One-Hot Encoding with Pandas.	15
Total Hours		**60**

Textbook :

1 Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, Wes McKinney, O'Reilly Media, 2022

References:

1 Python for Data Analysis, Python for Data Analysis, Wes McKinney, O'Reilly Media, 2017

2 Data Science from Scratch: First Principles with Python, Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, 2019

3 Think Stats: Probability and Statistics for Programmers, Think Stats: Probability and Statistics for Programmers, Allen B. Downey, Green Tea Press, 2014

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	15.00	25.00	25.00	25.00	0.00

Instructional Method:

1 Practical

Supplementary Resources:

1 Kaggle – Online datasets and notebooks
2 Google Colab – Cloud-based Python coding
3 Jupyter Notebook – Interactive Python notebooks
4 Anaconda Navigator – Local IDE for data science
5 Seaborn & Matplotlib Docs – For visualization examples

DR. SUNIL LALCHAND BAJEJA
Digitally signed by (Name of HOD)

DR. RIJWAN KHAN
Digitally signed by (Name of Dean/ Principal)