COURSE TITLE	**DATA HANDLING & EDA USING PYTHON**
COURSE CODE	**05CA0406**
COURSE CREDITS	**2**

Objective:

1 To introduce students to core Python libraries such as NumPy, Pandas, and Matplotlib for data handling and visualization

2 To enable students to clean, transform, and explore datasets using statistical and visual techniques

3 To develop foundational skills in generating insights through exploratory data analysis (EDA) for machine learning readiness

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

1 Apply Python libraries (NumPy, Pandas, Matplotlib, Seaborn) to manipulate and visualize datasets

2 Analyze real-world datasets by identifying patterns, trends, and outliers using EDA techniques

3 Evaluate dataset quality by handling missing values, encoding variables, and preparing data for machine learning models

Pre-requisite of course: Basics of Python

Teaching and Examination Scheme

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

Contents : Unit	Topics	Contact Hours
Total Hours		

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Python Libraries for Data Handling & Visualization** Working with NumPy Arrays – Creation, Indexing, Broadcasting, Vectorized Operations & Basic Math with NumPy, Series and DataFrames with Pandas, Importing, Exporting, and Cleaning Datasets, Data Aggregation and GroupBy in Pandas, Filtering and Boolean Indexing in Pandas, Handling Missing and Duplicated Data, Matplotlib for Basic Plotting (Line, Bar, Pie), Seaborn for Statistical Visualizations (Boxplot, Histogram, KDE), Lab Task: Clean and visualize a CSV dataset (e.g., Titanic, Sales)	30

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

DR. R SRIDARAN
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
2	**Exploratory Data Analysis (EDA) – Fundamentals** Overview of EDA – Why and How Summary Statistics – mean, median, std, skew, kurtosis, Univariate Analysis – Histograms, Value Counts, Bivariate Analysis – Scatter Plots, Correlation Heatmaps, Outlier Detection – Boxplots and IQR Method, Missing Value Analysis & Treatment (Imputation), Encoding Categorical Variables – Label vs One-Hot, Feature Scaling – MinMax, StandardScaler, EDA Reporting Techniques & Tools, Lab Task: Perform EDA on a real-world dataset and summarize findings	30
Total Hours		**60**

Textbook :

1 Python for Data Analysis, Wes McKinney, O'Reilly Media, 2022

2 Data Wrangling with pandas, acqueline Kazil, Katharine Jarmul, O'Reilly Media, 2016

References:

1 Effective Pandas: Patterns for Data Manipulation, Effective Pandas: Patterns for Data Manipulation, Matt Harrison, Self-published, 2021

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

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**
20.00	30.00	25.00	15.00	10.00	0.00

Instructional Method:

1 Board Work

2 PPT

3 Demo

Supplementary Resources:

1 https://pandas.pydata.org/docs/

2 https://www.geeksforgeeks.org/exploratory-data-analysis-eda/

3 https://www.kaggle.com/learn/data-cleaning

4 https://realpython.com/learning-paths/pandas-data-science/

DR. SUNIL LALCHAND BAJEJA

Digitally signed by (Name of HOD)

DR. R SRIDARAN

Digitally signed by (Name of Dean/ Principal)

Supplementary Resources:

5 https://matplotlib.org/stable/tutorials/index.html