Program: B.Com (Economics and Analytics)	Semester: III
Course: Data Wrangling with Python AY:2024-25	Course Code:

Teaching Scheme				Evaluation Scheme	
Lecture (Hours per week)	Practical (Hours per week)	Tutorial (Hours per week)	Credit	Continuous Assessment (CA) (Marks - 30)	Semester End Examinations (SEE) (60 Marks - in Question Paper)
4 LABS	-	-	4		

Course Learning Objectives:
1. To understand basics of python programming to prepare for data manipulation and analysis processes
2. To learn and use python programming libraries like NumPy, Pandas, and more to efficiently manipulate and analyse datasets
3. To acquire and develop skills in cleaning, formatting, and wrangling datasets to keep them ready for analysis and decision making

Course Outcomes:
After completion of the course, learners would be able to:

CO1: Write efficient python code to solve problems and perform basic data analysis

CO2: Manipulate and analyse datasets using python libraries for performing advanced analytics tasks

CO3: Apply data formatting and cleaning to ensure data quality for further modeling

Pedagogy:
Hands-On Learning, Computer Laboratory Based Learning

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	Introduction to Python Programming	15
2	Data Manipulation and Analysis Techniques with Python	15
3	Data Cleaning and Formatting Techniques	15
4	Data Transformation, Pre-processing, and Wrangling	15
Total		**60**
Practicals		**-**

Unit	Topic	No. of Hours
Module 1	**Introduction to Python Programming** Introduction to Python, Significance in Problem Solving, Overview of Python for Analytics, Comments, Variables and Their Scope, Standard Data Types, Python Identifiers, Reserved Words, Operators, Statements and Expressions, Basic Input and Output, Keyboard Input, If Statement, If-Else Statement, If-Elif-Else Statement, Nested Conditional Statement, For Loop, While Loop, Break Statement, Continue Statement, Return Statement, Role of Functions, Function Definition, Parameters, and Arguments, Built-In Functions, Introduction of Strings, Lists, Tuples, and Dictionary	15
Module 2	**Data Manipulation and Analysis Techniques with Python** Introduction to NumPy Understanding Data Types in Python, The Basics of NumPy Arrays: 1D, Computation on NumPy Arrays, Aggregations: min, max, Basic Indexing and Slicing, Sorting, Transposing, and Swapping Axes of Arrays, Multidimensional Arrary, Array Manipulations Getting Started with Pandas Introduction to Pandas Data Structures: Series and DataFrames, Hierarchical Indexing, Handling Missing Data, Data Transformation, String Manipulation, Aggregating, Combining, Grouping and Joining Pandas DataFrame, Pivot Tables, Summarizing and Computing Descriptive Statistics, Exploratory Data Analysis	15
Module 3	**Data Cleaning and Formatting Techniques** Handling Missing Data Complete Case Analysis, Handling Missing Numerical Data, Mean or Median Imputation, End of Distribution, Arbitrary Value Imputation, Handling Missing Categorical Data, Frequency Category Imputation, Missing Category Imputation, Handling Mixed Values, Handling Date Data Type, Handling Time Data Type Handling Large and Imbalanced Datasets Large Datasets, Handling Techniques, Imbalanced Dataset, Down Sampling, Up Sampling, Handling Sparse Data	15
Module 4	**Data Transformation, Preprocessing, and Wrangling** Encoding Categorical Data One Hot Encoding, Label Encoding, Frequency Encoding, Ordinal Encoding, Mean Encoding Data Discretization Equal Width Discretization, Equal Frequency Discretization, Custom Discretization Outlier Treatment	15

| | Outlier Trimming, Outlier Capping using IQR, Outlier Capping using Mean and Standard Deviation, Outlier Capping with Quantiles, Outlier Capping using Custom Values

Feature Scaling
Standardization, Min-Max Normalization, Mean Normalization, Maximum Absolute Scaling, Median and Quantile Scaling | |

Reference Books:

[1]. E. Balaguruswamy, "Introduction to Computing and Problem Solving with Python", McGraw Hill Education, 2018

[2]. Martin, "Python : The Complete Reference", McGraw Hill Education, 2018

[3]. Wes McKinney, "Python for Data Analysis", 3rd Edition, O'Reilly Media Inc., August 2022

[4]. Avinash Navlani, Armando Fandango, Ivan Idris, "Python Data Analysis", 3rd Edition, Packt Publishing, February 2021

[5]. Scott McCoy, "Murach's Python for Data Analysis", Murach, August 2021

[6]. Jake VanerPlas, "Python Data Science Handbook", 2nd Edition, O'Reilly Media Inc., December 2022

[7]. Ayodele Oluleye, "Exploratory Data Analysis with Python Cookbook", 1st Edition, Packt Publishing, 2023

[8]. Ihab F. Ilyas, "Data Cleaning", Association for Computing Machinery, 2019

[9]. Dr. Tirhtjyoti Sarkar, Shubhadeep Roychowdhury, "Data Wrangling with Python", Packt Publisher, 2019

[10]. Jason Osborne, "Best Practices in Data Cleaning – A Complete Guide to Everything You Need to Do Before and After Collecting Your Data", 2012

EXAM PATTERN FOR THE COURSE

1. **Internal Class Test - 10 Marks Paper Pattern (Pen Paper Based)**

Question 1	Fill in the Blanks	4 Marks
Question 2	Descriptive Question	3 Marks
Question 3	Descriptive Question	3 Marks

2. **10 Marks Internal Assignments / Practical Study / Case Study / Mini Project**

3. **60 Marks Final Exam Paper Pattern**

a. Examination shall be conducted in machine test form in Batches (Max. Batch Size – 30 Students) in the computer laboratory

b. One external examiner must be present along with the internal examiner (subject faculty in-charge) for the conduct of examination

c. Question paper should have maximum number of distinct sets, kept faced down on table, from which student will pick up one question paper

d. Duration of the examination is 2.5 Hrs

e. Marks distribution is as follows:

1	Practical Question Two questions for 20 marks each = 2 * 20 = 40 and One question for 10 marks = 1 * 10 = 10	50 Marks
2	Viva Voce	05 Marks
3	Coursework Journal	05 Marks

Evaluation shall be done by the examiners, both internal and external, on machine in the computer laboratory

Students must prepare answer book during the examination with the code and output in it, which further must be printed