LJ University

University with a Difference

Diploma in Computer Science & Engineering

Course Code: 025130404

Data Science with Python

programme / Branch Name			Diploma in Computer Science & Engineering			
Course Name	Data Science with Python				**Course Code**	025130404
Course Type	HSSC	BSC	ESC	PCC	OEC	PEC

Legends: HSSC: Humanities and Social Sciences Courses BSC: Basic Science Courses
ESC: Engineering Science Courses PCC: program Core Courses
OEC: Open Elective Courses PEC: program Elective Courses

1. Teaching and Evaluation Scheme

Teaching Hours / Week / Credits				Evaluation Scheme			
L	**T**	**P**	**Total Credit**	**CCE**	**SEE (Th)**	**SEE (Pr)**	**TOTAL**
3	0	4	5	50	50	50	150

Legends:
L: Lectures T: Tutorial P: Practical
CCE: Continuous & Comprehensive Evaluation
SEE (Th): Semester End Evaluation (Theory)
SEE (Pr): Semester End Evaluation (Practical)

2. Prerequisites

- ✓ Basic knowledge of C programming
- ✓ Basic knowledge of Python Programming
- ✓ Basic knowledge of computers.

3. Rationale

- ✓ In today's data-driven world, data science has become a crucial field with widespread applications across industries and disciplines. For college students, learning data science is about acquiring technical skills and building a future-ready mindset. Students should embrace data science as part of their academic and personal development. The global demand for data science professionals is growing rapidly across industries, such as technology, healthcare, finance, and marketing. By learning data science, students position themselves as valuable candidates in a competitive job market. Data science is versatile and integrates with biology, economics, social sciences, engineering, and the arts. Students from any discipline can apply data science skills to enhance their understanding of domain-specific challenges and solve real-world problems.

4. Objectives

- ✓ The theory should be taught and the practical should be carried out in such a manner that students can acquire different learning outcomes in cognitive, psychomotor, and affective domains to demonstrate the following course outcomes.
  - Use data science to address practical problems in areas like healthcare, business, finance, education, and more.
  - Acquire the ability to interpret, process, and work with data effectively.
  - Understand different types of data (structured, unstructured, and semi-structured) and their relevance to real-world scenarios.

- Gain proficiency in essential tools, frameworks, and programming languages like Python, R, SQL, and Excel

5. Contents

Unit No.	Topics	Sub-Topics	Learning Outcomes	% Weightage	Hours
1	Introduction to Data Science	1.1. Introduction to Data Science 1.2. Fundamental of Data Science. 1.3. Types of Data 1.4. Mathematical Foundation for Data Science: Linear Algebra. 1.5. Analytical and numerical solutions of linear equations; 1.6. Tools for Data Science	• Basics and history of Data Science • Various types of Data • Concepts of statistics in solving problems arising in data science	15	6
2	Data Collection and Management	2.1. Introduction to Data Collection 2.2. Various types of Data Pre-Processing techniques 2.3. Necessary Python Libraries Pandas, Scipy, NumPy, Seaborn 2.4. First Program to Preprocess Dataset. 2.5. Data Storage and Management	• Knowledge of Pre-Processing Methods • Preprocess any Dataset using Python Libraries. • Knowledge of Data Storage and Management.	20	8
3	Data Analysis	3.1. Introduction to Data Analysis 3.2. Process of Data Analysis 3.3. Types Data Analysis 3.4. Tools for Data Analysis 3.5. Various Data Analytical Techniques: 1) Statical Analysis 2) Machine Learning 3) Data Visualization	• Why is Data Analysis Important? • Data Analysis Process • Types of Data Analysis	25	10

		3.6. Application of Data Analysis.	• Forecasting future outcomes based on historical data.		
4	Data Visualization	4.1. Introduction to Data Visualization 4.2. Categories of Data Visualization 1) Standard Reporting a. Bar Charts b. Line Charts c. Pie Charts 2) Multidimensional visualization a. Heat Maps b. Bubble Maps c. Tree Maps 4.3 Data visualization in Python using Matplotlib and Seaborn. 4.4 Importance of data visualization	• Visually presenting large, complex datasets to uncover patterns, correlations, and insights. • Determining correlations. The best way to determine the relationship and correlations between two variables is to compare them visually.	20	10
5	Data Wrangling	5.1. What is Data Wrangling? 5.2. Data Wrangling Process 5.3. Introduction to Scikit-Learn Library. 5.4. Data Wrangling formats. 5.5. Data Wrangling vs ETL 5.6. Introduction of Thread 5.7. Use Cases of Data Wrangling 5.8. Benefits of Data Wrangling	• Gain the ability to clean, transform, and organize raw data into a structured format ready for analysis. • Improve efficiency in handling large and complex datasets.	20	8
				Total Hours	**42**

6. List of Practicals / Exercises

The practical/exercises should be properly designed and implemented in an attempt to develop different types of skills so that students can acquire the competencies/programme outcomes. Following is the list of practical exercises for guidance.

Sr. No.	Practical / Exercises	Key Competency	Hours
1	Install Python or set up Jupyter Notebook and perform "Hello Word" program.	To be able to install Python	2
2	Install various packages NumPy, SciPy, and pandas, and import them into the Jupyter environment	To be able to install various Libraries	2
3	Write a simple program to perform arithmetic operations on the NumPy array.	To be able to perform operations on an array.	4
4	Write a program to load a CSV file into a pandas data frame.	To be able to load CSV file into the Data frame	2
5	Create a pandas data frame like a 2-dimensional array	To be able to use Data frame	2
6	Write a program to load a CSV file into a pandas data frame and perform head (), tail (), and info () on a data frame	To be able to Analyze data frames	2
7	Write a program to perform dataframe.replace(), dataframe.mask(),dataframe.resample(),dataframe.transform() method in pandas.	To understand various methods	2
8	Write a program to plot a bar graph of the subject and marks.	To be able to use matplotlib Library	2
9	Write a program to plot Distplot using the Seaborn Library.	To be able to plot various Distributions.	2
10	Write a program to plot data from CSV file using Pandas Library.	To be able to plot data from CSV.	2

Total Hours 22

7. Suggested Specification Table with Hours

Unit No.	Chapter Name	Teaching Hours	Distribution of Topics According to Bloom's Taxonomy					
			R %	U %	App %	C %	E %	An %
1	Introduction to Data Science	6	40	30	20	-	5	5
2	Data Collection and Management	8	30	30	20	10	5	5
3	Data Analysis	10	20	30	30	10	5	5
4	Data Visualization	10	20	20	20	10	15	15
5	Data Wrangling	8	20	20	30	10	10	10

Legends: R: Remembering U: Understanding
App: Applying C: Creating
E: Evaluating An: Analyzing

8. Textbooks

1) Python for data science for dummies 2nd Edition, John Paul Mueller, Luca Massaron, Wiley.
2) Pandas for everyone: Python Data Analysis, Daniel Y. Chen, Pearson.
3) Data Analytics, Anil Maheshwari, McGraw-Hill.

9. Reference Books

1) Data Science from Scratch: First Principles with Python, Joel Grus, SPD

10. Open Sources (Website, Video, Movie)

1) https://numpy.org/
2) https://matplotlib.org/
3) https://scikit-learn.org/stable/
4) https://pandas.pydata.org/
5) https://www.python.org/