PVP23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY
(Autonomous)
KANURU, VIJAYAWADA-520007
DEPARTMENT OF CSE (Data Science)
II B. Tech – I Sem CSE (DATA SCIENCE)
Data Science Lab

Course Code	23DS3351	**Year**	II	**Semester**	I
Course Category	Professional Core	**Branch**	CSE (DATA SCIENCE)	**Course Type**	Practical
Credits	1.5	**L-T-P**	0-0-3	**Prerequisites**	Probability & Statistics
Continuous Internal Evaluation:	30	**Semester End Examination:**	70	**Total Marks**	100

Course Outcomes		
Upon successful completion of the course, the student will be able to		
CO1	Demonstrate experimental procedures through oral communication and submit comprehensive documentation reports.	**L2**
CO2	Apply the Data Science Lifecycle process to develop data science projects using Python libraries	**L3**
CO3	Analyze data science problems, and critically evaluate the performance and limitations of proposed solutions.	**L4**
CO4	Evaluate the performance of the Data Science Lifecycle process in a given data science project using relevant metrics across diverse datasets.	**L5**

Contribution of Course Outcomes towards achievement of Program Outcomes & Strength of correlations (3: Substantial, 2: Moderate, 1: Slight)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO**
CO1	2									2				
CO2	3													
CO3		3										2		
CO4				3								2		

Syllabus		
Exp. No.	**Contents**	**Mapped CO**
1	Apply NumPy library for creating, manipulating, and performing mathematical operations on arrays.	CO1 to CO4
2	Apply Pandas library for data manipulation, focusing on creating, accessing, and manipulating DataFrames.	CO1 to CO4
3	Implement data collection techniques using Python libraries	CO1 to CO4
4	Implement web scraping techniques to extract data from websites using Beautiful Soup and Scrapy.	CO1 to CO4
5	Perform data cleaning operations on a given dataset, including handling missing values, removing duplicates, and correcting inconsistencies using pandas and numpy libraries.	CO1 to CO4
6	Apply various data transformation techniques such as normalization, standardization, and encoding categorical variables.	CO1 to CO4
7	Conduct exploratory data analysis on a given dataset. Calculate and interpret descriptive statistics, and generate summary reports using pandas and scipy libraries.	CO1 to CO4
8	Analyze relationships between variables in a dataset using correlation analysis and statistical tests. Visualize correlations using heatmaps and scatter plots.	CO1 to CO4
9	Create various types of plots and charts to visualize data patterns and trends using matplotlib and seaborn libraries.	CO1 to CO4
10	**Capstone Project**: Implement a complete data science project pipeline, including data collection, preprocessing, exploratory data analysis, and visualization, focusing on a specific application area such as social media sentiment analysis or stock market prediction.	CO1 to CO4

Learning Resources
Text Books
1. Python for Data Analysis, Wes McKinney, Second Edition, 2017, O'Reilly Media
Reference Books
1. Data Science from Scratch, Joel Grus, Second Edition, 2019, O'Reilly Media 2. Python Programming, S Sridhar, J Indumathi, V M Hariharan, Second Edition, 2024, Pearson. 3. Introduction to Programming Using Python, Y. Daniel Liang, First Edition, 2021, Pearson.
E-Resources & other digital material
1. https://www.analyticsvidhya.com/blog/2020/04/the-ultimate-numpy-tutorial-for-data-science-beginners/ **2.** https://www.analyticsvidhya.com/blog/2021/07/data-science-with-pandas-2-minutes-guide-to-key-

concepts/
3. https://www.analyticsvidhya.com/blog/2020/04/how-to-read-common-file-formats-python/
4. https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing-python-scikit-learn/
5. https://www.analyticsvidhya.com/blog/2020/02/beginner-guide-matplotlib-data-visualization-exploration-python/6. https://www.nltk.org/book/ch01.html