SARVAJANIK UNIVERSITY
Sarvajanik College of Engineering and Technology
Bachelor of Technology

Year: B. Tech III (Semester VI)

Subject Name: Data Analysis and Visualization **Subject Code:** BTIT13601
Type of course: Professional Core Course
Prerequisite (if any): Programming for Problem Solving

Rationale: Data science helps in managing, analysing and understanding trends in data leading to design the strategy for better profitability and results. Data visualization sees the pattern in data and also sees the pattern when data is not part of pattern.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments / projects / open book tests / closed book tests) CA2: Sincerity in attending classes / class tests / timely submissions of assignments / self-learning attitude / solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work / Quality of work submitted / Active participation in lab sessions / viva on practical skills learned in course.

Contents:

Sr. No.	Contents	Total Hrs
1.	**Introduction to Data Science:** Component of Data Science, Data Science Process, Data Science Roles, Difference between Data Science and Business Intelligence, Applications of Data Science in various fields, Challenges of Data Science Technology	04
2.	**Exploratory Data Analytics** : Exploratory Data Analysis (EDA), Types of Exploratory Data Analysis, Exploratory Data Analysis Tools (Python, R), Box Plots, Pivot Table, Heat Map, Histograms, Line Graphs	08
3.	**Model Development and Evaluation**: Linear Regression (Simple and Multiple Regression) , Overfitting and underfitting, Regularization techniques, Logistic Regression, Confusion matrix, Precision, Recall, F-Score, ROC curve, Introducing clustering basics, Recognizing the Difference between Clustering and Classification, Making sense of data with nearest neighbour analysis, classifying data with K- nearest neighbour algorithms, Solving Real-world problems	16
4.	**Getting Started with Pandas:** Arrays and vectorized computation, Introduction to pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics. **Data Loading, Storage and File Formats :** Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, Interacting with Web APIs, Interacting with Databases	07

	Data Cleaning and Preparation : Handling Missing Data, Data Transformation, String Manipulation	
5.	**Data Wrangling:** Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting. **Data Visualization matplotlib:** Basics of matplotlib, plotting with pandas and seaborn, other python visualization tools	06
6.	**Data Aggregation and Group operations:** Group by Mechanics, Data aggregation, General split-apply-combine, Pivot tables and cross tabulation	04

Suggested Specification table with Marks (Theory): (For B. Tech only)

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
10	**20**	**15**	**5**	**-**	**-**

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (Revised Bloom's Taxonomy)

Reference Books:

Sr No	**Title of book /article**	**Author(s)**	**Publisher and details like ISBN**
1	Doing Data Science: Straight Talk from the Frontline	Cathy O'Neil	O'Reilly Media
2	The Art of Data Science	Roger Peng	Lulu.com
3	Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython	McKinney, W.	O'Reilly Media

Note: Students should refer to the latest editions of books

Course Outcomes (CO):

Sr. No.	**CO statements**	**Marks % weightage**
CO-1	Understand the basic concept of data science.	10%
CO-2	Learn various type and tools of exploratory data analysis.	20%
CO-3	Explore how to develop model and evaluate model.	40%
CO-4	Use data analysis tools in the pandas library.	15%
CO-5	Create informative visualization and summarize data sets.	15%

List of Open learning website:

- https://onlinecourses.nptel.ac.in/noc21_cs69/course

Suggested List of Experiments:

1. Data Analysis practicals will be based on discussion in the classroom.
2. Perform experiments using NumPy.
3. Explore Pandas Data Structures.
4. Write a program for Data Loading, Storage and File Formats.
5. Perform experiments based on Interacting with Web APIs.
6. Explore Data Wrangling.
7. Perform Data Visualization using matplotlib.
8. Perform Practical based on Data Aggregation.
9. Develop Regression model based on sample data.
10. Implement k-nearest neighbour algorithm.
11. Demonstrate classification models for sample data set.