GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Minor/Hons. Program

Level: UG

Branch: Minor/Hons. Data Science

Course / Subject Code : BE040AT011

Course / Subject Name : Introduction to Data Science

w. e. f. Academic Year:	2025-26
Semester:	4th
Category of the Course:	Core Courses

Prerequisite:	Basic understanding of computer fundamentals, statistics, and programming (preferably Python) along with familiarity with spreadsheets and data handling concepts.
Rationale:	This course provides a foundation in data science concepts, tools, and techniques. It emphasizes data collection, cleaning, analysis, visualization, and interpretation using Python and its libraries. Students will gain practical skills in data analytics, statistical reasoning, and visualization to support data-driven decision-making across various domains.

Course Outcome:

After Completion of the Course, Student will able to:

Sr. No.	CO statement	Marks % Weightage
CO-1	Understand the fundamentals of Data Science, including types of data, workflows, and roles of a data scientist.	10%
CO-2	Apply Python programming constructs and libraries such as NumPy and Pandas for data collection, manipulation, and analysis.	25%
CO-3	Utilize mathematical and statistical techniques including measures of central tendency, dispersion, and probability for data interpretation.	20%
CO-4	Perform exploratory data analysis, feature generation, and selection to derive insights and prepare datasets for analytics and modeling.	25%
CO-5	Develop visualizations and dashboards to communicate insights clearly and evaluate real-world applications of Data Science with consideration of ethical issues.	20%

GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Minor/Hons. Program

Level: UG

Branch: Minor/Hons. Data Science

Course / Subject Code : BE040AT011

Course / Subject Name : Introduction to Data Science

Teaching and Examination Scheme:

Teaching / Learning Scheme (in Hours per semester)					Total Credits = TH/30	Assessment Pattern and Marks					Total Marks
						Theory		Tutorial / Practical			
L	T	P	PBL*	TH		ESE (E)	PA / CA (M)	PA/ CA (I)	PBL (I)	ESE (V)	
45	00	30	45	120	4	70	0	0	30	50	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	**Introduction to Data Science:** • Definition, scope, importance of Data Science • Types of data: structured, semi-structured, unstructured • Data Science workflow: collection, cleaning, analysis, visualization, interpretation • Roles in Data Science: Analyst, Scientist, Engineer	05	11
2.	**Python for Data Science:** • Python Basics: variables, data types, operators, loops, conditionals, functions • Python Libraries: NumPy, Pandas, Matplotlib, Seaborn - Data handling: DataFrames, Series, reading/writing CSV, Excel, JSON • Jupyter Notebook & Google Colab overview	10	22
3.	**Mathematics for Data Science:** • Data Type (Qualitative, Quantitative) • Population and Sample • Measures of Central Tendency(Mean, Median, Mode) • Measures of Dispersion (Range, Variance, Standard Deviation) • Probability • Events (Independent, Dependent, Mutually Exclusive)	8	18
4.	**Data Analytics:** • Types of Analytics: Descriptive, Diagnostic, Predictive, Prescriptive • Exploratory Data Analysis (EDA)	10	22

	• Feature Generation: creating new features from existing data, encoding categorical variables • Feature Selection: correlation analysis, variance thresholding, univariate selection, recursive feature elimination		
5.	**Data Visualization:** • Importance of visualization • Charts: bar, line, histogram, scatter plot, heatmap • Advanced visualization: Seaborn, Matplotlib • Interactive visualization: Plotly, Dash	7	16
6.	**Data Science and Its Current Applications:** • Applications in finance, healthcare, marketing, e-commerce, social media • Case studies of data-driven decision-making • Ethics: privacy, bias, fairness • Trends: AI, ML, Big Data, IoT integration	5	11
	Total		**100**

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
7	**17**	**33**	**21**	**18**	**4**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

References/Suggested Learning Resources:
(a) Books:

1. **Python for Data Analysis (3rd Edition)** – Wes McKinney, O'Reilly Media
2. **Mathematics for Machine Learning** – Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Cambridge University Press
3. **Business Analytics – The Science of Data-Driven Decision Making** – U. Dinesh Kumar, Wiley India
4. **Introduction to Data Science** – Laura Igual, Santi Seguí, Springer
5. **Practical Statistics for Data Scientists (2nd Edition)** – Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media

(b) Open source software and website:

1. **Python (Anaconda Distribution)** – Open-source programming environment for Data Science and Machine Learning (https://www.anaconda.com/products/distribution)
2. **Jupyter Notebook / JupyterLab** – Interactive notebook for Python programming, data visualization, and analysis (https://jupyter.org/)
3. **Google Colab** – Cloud-based Python and Jupyter environment for Data Science and Machine Learning practice (https://colab.research.google.com/)
4. **Kaggle** – Open platform for datasets, competitions, and collaborative Data Science notebooks (https://www.kaggle.com/)

(c) Suggested MOOC Courses:

1. **Data Science Using Python** By Prof. Faisal Anwer, Prof. Mohammad Nadeem, (https://onlinecourses.swayam2.ac.in/ini25_cs04/preview)
2. **Basic Python Programming and its Application in Finance** By Dr. Chaya Bagrecha, Dr. Shalini R (https://onlinecourses.swayam2.ac.in/imb25_mg208/preview)

Course Practical List:

Sr. No	Practical	CO
	Practical Set-1: (Python Basics and Environment Setup)	**CO1,CO2**
1.1	Python Environment Setup & First Program: Install and configure Python using Anaconda, explore Jupyter Notebook and Google Colab. Write a program to display "Hello, World!", demonstrate variable declarations, type casting, and use of comments.	
1.2	Basic Input and Arithmetic Operations: Compute distance between two points, perform addition, subtraction, multiplication, division; convert temperature between Celsius and Fahrenheit.	
1.3	Number System and Variable Manipulation: Convert decimal to binary, octal, hexadecimal; swap values of two variables without using a third variable.	
1.4	Character and String Operations: Find ASCII value of a character, check whether a character is vowel/consonant, check if a string is a palindrome.	

Practical Set-2 (Conditional Statements and Loops)		**CO1,CO2**
2.1	Conditional Programs: Check whether a number is even/odd, check for leap year.	
2.2	Matrix Operations: Add two matrices, calculate sum/product, find maximum/minimum.	
2.3	Loops: Generate multiplication tables, sum of series, factorial of a number using loops.	
2.4	Pattern Printing: Print numeric and character patterns using nested loops.	
2.5	Random Operations: Select random item from a list, simulate dice roll, generate sequences with probabilities (1,3,5,7) using NumPy.	
Practical Set-3 (Data Handling using List and Tuple)		**CO2,CO3**
3.1	List and Tuple Operations: Shift list elements left by one position; insert/delete elements; sum, product, largest/smallest, remove duplicates, clone list, select random item; demonstrate count() and index() methods on tuples.	
3.2	Advanced List Operations: Shift elements left by one position; insert element at beginning/end; delete element by index; reverse a list.	
3.3	List Manipulations: Sum, product, largest/smallest, second largest/smallest, remove duplicates, clone, check empty, insert/delete at position, get unique values, select random item.	
3.4	Tuple Operations: Demonstrate count() and index() methods on tuples.	
Practical Set-4 (Data Handling using Set and Dictionary)		**CO2,CO3**
4.1	Demonstrate ten Python set methods such as union, intersection, difference, symmetric difference, add, remove, pop, clear, issubset, issuperset.	
4.2	Perform dictionary operations such as sorting by key and value, adding new key-value pairs, creating dictionaries using comprehension (keys 1–5 with square values), finding maximum and minimum values, removing duplicates, and merging multiple dictionaries while handling duplicate keys.	
Practical Set-5 (Functions & NumPy Array Operations)		**CO2,CO3**
5.1	Function Basics: Write functions for sum, product, and factorial; implement recursive functions.	
5.2	Array Operations: Element-wise arithmetic, find LCM/GCD of arrays, union/intersection of arrays.	
5.3	Array Statistics: Find mean, median, mode, standard deviation, min/max of arrays using NumPy.	
5.4	NumPy Arrays: Create 1D, 2D, 3D arrays; reshape, join, sort, and filter arrays.	
5.5	Copy vs View: Differentiate between copy and view in NumPy with examples.	
Practical Set-6 (Pandas for Data Handling)		**CO2,CO3**
6.1	Pandas Series and DataFrame: Create Series and DataFrames from lists, dictionaries, CSV.	
6.2	Data Import/Export: Load data from CSV; save DataFrames to CSV/Excel.	

6.3	Data Inspection: Use .head(), .tail(), .info(), .describe(), check nulls, remove duplicates.	
6.4	Data Cleaning: Replace invalid values, handle missing data, type conversions.	
6.5	Indexing and Selection: Filter rows/columns, select data using labels and positions, slicing techniques.	
Practical Set – 7 (Probability, Statistics, and Mathematics)		**CO3**
7.1	Compute and store qualitative and quantitative data values in a dataset.	
7.2	Calculate descriptive statistics including mean, median, mode, range, variance, and standard deviation for both population and sample data.	
7.3	Implement basic probability concepts such as combinations, permutations, independent/dependent events, and mutually exclusive events.	
Practical Set – 8 (Data Analytics & Feature Engineering)		**CO3,CO4**
8.1	Perform Exploratory Data Analysis (EDA): summary statistics, data types, missing values.	
8.2	Feature Generation: create new features from existing columns, encoding categorical variables.	
8.3	Feature Selection: correlation analysis, variance thresholding, univariate selection, recursive feature elimination.	
Practical Set – 9 (Data Visualization)		**CO4,CO5**
9.1	Create basic charts: bar, line, histogram, scatter plot using Matplotlib.	
9.2	Create heatmaps and advanced charts using Seaborn.	
Practical Set – 10 (Advance Data Visualization)		**CO4,CO5**
10.1	Create interactive visualizations using Plotly: scatter and bar charts.	
10.2	Customize plots: titles, labels, legends, color, size, style.	
10.3	Visualize real dataset features, perform trend analysis, and interpret results.	

Suggested Activities for Students:

Mini Project (45 Hours)

Mini Project Design Instructions (Develop a Python-based mini-project integrating concepts from all practical sets.)

Instructions:

1. **Project Definition:** Choose a suitable Python mini-project and clearly state its objective, scope, and expected outcomes.

2. **Dataset Selection:** Select an appropriate dataset relevant to your project objective.
3. **Implementation Platform:** Use Python with Jupyter Notebook or Google Colab for coding and analysis.
4. **Functions & NumPy:** Implement modular Python functions, and use NumPy for array creation, element-wise operations, and array statistics.
5. **Data Handling with Pandas:** Load, inspect, clean, and transform the dataset efficiently using Pandas.
6. **Statistical Analysis:** Apply descriptive statistics, probability calculations, and basic mathematical operations to analyze the dataset.
7. **Exploratory Data Analysis & Feature Engineering:** Generate new features, perform correlation analysis, and prepare the data for insight extraction.
8. **Data Visualization:** Create static and interactive charts using Matplotlib, Seaborn, and Plotly. Include interpretations of visualizations.
9. **Documentation:** Provide proper documentation with code comments, explanation of steps, and visualization interpretations.
10. **Submission:** Include complete code, dataset, project report
11. **Collaboration:** Work individually or in pairs.

Note: Apply concepts from all practical sets to ensure a comprehensive and integrated mini-project experience.

* * * * * * *