



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)**

Level: Under Graduate

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

w. e. f. Academic Year:	2026-27
Semester:	5
Category of the Course:	Discipline-Specific Elective-1

Prerequisite:	Basic computer knowledge, elementary statistics, and basic programming concepts (preferably Python fundamentals).
Rationale:	• To introduce students to the fundamentals of Data Science and data-driven decision making. • To develop understanding of data collection, preprocessing, and analysis techniques. • To provide practical exposure to data handling using Python tools. • To enable students to interpret data through statistical analysis and visualization. • To build analytical thinking skills required for modern computing applications.

Course Outcome:

After Completion of the Course, the Student will able to:

No	Course Outcomes	RBT Level
01	Understand fundamental concepts, lifecycle, and applications of Data Science.	R,U
02	Perform data acquisition and preprocessing using data handling techniques.	U,A
03	Apply statistical methods for data analysis and interpretation.	A,N
04	Create meaningful visualizations for representing datasets.	A,C
05	Analyze real-world datasets to derive insights using modern tools.	N,E

**Revised Bloom's Taxonomy (RBT)*

Teaching and Examination Scheme:

Teaching Scheme (in Hours)			Total Credits L+T+ (PR/2)	Assessment Pattern and Marks				Total Marks
L	T	PR	C	Theory		Practical		
				ESE (E)	PA / CA (M)	PA(I)	ESE (V)	
4	0	2	5	70	30	20	30	150



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)**

Level: Under Graduate

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

Course Content:

Unit No.	Content	No. of Hours	% of Weightage
1.	Introduction to Data Science Introduction to Data Science, importance and scope of Data Science, Data Science versus Data Analytics and Business Intelligence, applications of Data Science, Data Science lifecycle. Types of data: structured, semi-structured and unstructured data, data sources and data collection methods, role of data scientist, challenges in data handling, ethical issues and data privacy.	10	18%
2.	Data Collection and Data Preprocessing Data acquisition techniques, working with CSV, Excel and JSON data formats, data cleaning concepts, handling missing values and inconsistent data, noise and outlier detection, data integration and transformation, data reduction techniques, normalization and scaling, introduction to NumPy, introduction to Pandas, Series and DataFrame, data loading, selection, filtering and basic data manipulation.	12	20%
3.	Data Analysis and Statistical Techniques Introduction to statistics for data science, descriptive statistics, data summarization techniques, measures of central tendency: mean, median and mode, measures of dispersion: range, variance and standard deviation, correlation and covariance, basic probability concepts, aggregation and grouping operations, interpretation of statistical results.	14	22%
4.	Data Visualization Techniques Importance of data visualization, principles of visual representation, selection of appropriate visualization methods, graphical techniques including line chart, bar chart, histogram, pie chart, scatter plot and box plot, introduction to Matplotlib and Seaborn libraries, plot customization, comparison and presentation of analytical results using visualization.	12	20%
5.	Exploratory Data Analysis using Python Introduction to exploratory data analysis, dataset inspection and understanding structure, data filtering and sorting, data transformation using Pandas operations, identification of trends and patterns, combining statistical analysis with visualization, case study-based data analysis, interpretation of results and preparation of analytical summaries.	12	20%
	Total	60	100



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)**

Level: Under Graduate

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks (in %)					
R Level	U Level	A Level	N Level	E Level	C Level
12	24	28	18	10	8

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 — Wes McKinney, O'Reilly Publication.
(All syllabus topics including Pandas, NumPy, preprocessing, statistics, and visualization are covered in this book.)
2. Data Science for Business — Foster Provost & Tom Fawcett, O'Reilly
3. Practical Statistics for Data Scientists — Peter Bruce & Andrew Bruce, O'Reilly

(b) Open-source software and website:

1. Development Environment & IDEs
 - Anaconda Distribution – <https://www.anaconda.com>
 - Jupyter Notebook – <https://jupyter.org>
 - VS Code – <https://code.visualstudio.com>
2. Programming Language
 - Python – <https://www.python.org>
3. Data Analysis Libraries
 - NumPy – <https://numpy.org>
 - Pandas – <https://pandas.pydata.org>
 - Matplotlib – <https://matplotlib.org>
 - Seaborn – <https://seaborn.pydata.org>
4. Dataset Sources
 - Kaggle – <https://www.kaggle.com/datasets>
 - UCI Repository – <https://archive.ics.uci.edu>



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)
Level: Under Graduate**

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

Suggested Project List:

1. Installation and Python Environment Setup

- Install Anaconda or Python environment and create a Jupyter Notebook.
- Write a Python program to demonstrate basic data structures including lists, tuples, dictionaries, and basic input/output operations used in data analysis.

2. Data Loading and Inspection

Write a Python program to load a dataset from CSV and Excel files using Pandas. Perform the following tasks:

- Load dataset using Pandas
- Display first and last records
- Check dataset dimensions
- Display column information and data types
- Generate statistical summary of data

(Use any open dataset from Kaggle or UCI Repository.)

3. Handling Missing Values

Write a Python program to identify and handle missing values in a dataset. Perform the following tasks:

- Detect missing values
- Replace missing values using mean/median/mode
- Remove rows or columns with missing data
- Compare dataset before and after cleaning

(Use Titanic dataset: <https://www.openml.org/d/42438>)

4. Data Filtering and Selection

Write a Python program using Pandas to perform data selection and filtering operations. Perform:

- Column selection
- Row filtering using conditions
- Sorting dataset values
- Selecting subsets of data



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)
Level: Under Graduate**

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

5. Data Transformation and Normalization

Write a Python program to perform data transformation operations. Perform the following tasks:

- Create new calculated columns
- Apply normalization and scaling techniques
- Rename and modify columns
- Convert categorical values where required

6. Statistical Analysis using Python

Write a Python program to compute statistical measures for a dataset including:

- Mean, Median, Mode
- Variance and Standard Deviation
- Minimum and Maximum values
- Correlation between variables

(Use Iris Dataset: <https://archive.ics.uci.edu/dataset/53/iris>)

7. Data Aggregation and Grouping

Write a Python program to perform aggregation operations using Pandas. Perform:

- Group data using groupby()
- Calculate aggregated statistics
- Compare grouped results
- Generate summarized dataset report

8. Data Visualization using Matplotlib

Write a Python program to visualize dataset attributes using Matplotlib. Create:

- Line chart
- Bar chart
- Histogram
- Pie chart

Interpret the visual results.



GUJARAT TECHNOLOGICAL UNIVERSITY

**Program Name: Bachelor of Computer Applications /
Bachelor of Computer Applications (Honours/Honours with Research) /
Master of Computer Applications (Integrated)
Level: Under Graduate**

Course / Subject Code: BC05001021

Course / Subject Name: Introduction to Data Science

9. Data Visualization using Seaborn

Write a Python program to visualize relationships between variables using Seaborn library. Create:

- Scatter plot
- Box plot
- Distribution plot
- Heatmap for correlation analysis

10. Exploratory Data Analysis (EDA)

Write a Python program to perform exploratory data analysis on a dataset. Perform:

- Data inspection and cleaning
- Statistical summary generation
- Visualization of important features
- Identification of patterns and trends

(Use any real-world dataset from Kaggle.)

CO- PO Mapping:

Semester 5	Course Name: Introduction to Data Science										
	POs										
Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11
CO1	3	2	-	-	2	2	2	-	-	1	-
CO2	3	3	2	2	3	1	1	-	-	-	-
CO3	3	3	-	3	2	-	1	-	-	-	-
CO4	2	2	3	2	3	-	-	-	2	-	-
CO5	3	3	2	3	3	1	2	-	2	1	1

Legend: '3' for high, '2' for medium, '1' for low and '-' for no correlation of each CO with PO.

* * * * *