



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Subject Code: BE05000231

Subject Name: Python for Data Science

w. e. f. Academic Year:	2024-25
Semester:	5
Category of the Course:	Professional Elective Course - 1

Prerequisite:	Programming concepts, Statistical and numerical methods
Rationale:	The data collected by organization needs insights to take the decisions, for predictions as well as for finding hidden patterns inside the data. Python is an appropriate language supporting all the features and libraries to perform data science activates. This subject covers the overview of the python with emphasis on various python data structures and various libraries like Pandas, NumPy, Matplotlib for performing various data science function including data preparation, cleaning, exploratory analysis and visualization

Course Outcomes:

Sr. No.	CO statement	Marks% weightage
CO-1	Understand and learn basic language features of python language	20
CO-2	To learn the fundamentals of some of the widely used python packages and modules	20
CO-3	Understand data lifecycle and study Probabilistic and Inferential Statistics for data science	20
CO-4	To prepare the data for design applications through various data preprocessing operations	20
CO-5	Apply various data visualization techniques on data	20

Teaching and Examination Scheme:

Teaching / Learning Scheme (in Hours per semester)					Total Credits	Assessment Pattern and Marks					Total Marks
L	T	P	PBL	Total no of hours per semester		Theory		Tutorial / Practical			
						ESE (E)	PA / CA (M)	PA/CA (I)	PBL (I)	ESE (V)	
45	0	30	15	90	3	70	30	20	30	50	200

* *Problem-Based Learning (PBL) aims to accommodate learning beyond syllabus as per clause 9.4 of NBA manual.*



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Subject Code: BE05000231

Subject Name: Python for Data Science

Content:

Sr. No.	Content	Total Hrs
1	<p>Introduction, Data Types and Operators:</p> <p>Features & History of Python, Installation and working with Python, Variables and data types in python, perform computations and create logical statements using Python's operators: Arithmetic, Assignment, Comparison, Logical, Membership, Identity, Bitwise operators, Array, list, tuple, set, dictionary and string operations.</p> <p>Python Decision making, Loops & Functions:</p> <p>Conditional statements using If statement, if ...else statement, elif statement and Boolean expressions, while loop, for loop, Nested Loop, Break, Continue, Pass, defining custom functions, Organizing Python codes using functions, Create and reference variables using the appropriate scope, Modules - create own and use basic inbuilt modules like random, statistics, math.</p>	10
2	<p>Python File Handling and exception handling:</p> <p>An introduction to file I/O, use text files, use CSV files, use binary files, Exception handling in python. Python MySQL basic operations</p>	8
3	<p>Data Science and Descriptive Statistics:</p> <p>Understanding data, its type and data science, Considering the emergence and fundamentals of data science, Data science life cycle, Data analysis (Univariate, bivariate and Multivariate), Data analytics, Data measurement scale, Data descriptive statistics (Measures of central tendency, dispersion/variation, measure of location, Shape and symmetry), Understanding Python's role in data science.</p>	6
4	<p>Exploring Python libraries:</p> <p>Exploring statistics of data using stats package from SciPy, Numerical computing and working with multidimensional array through NumPy, Data manipulation through Pandas, Basic functionalities of machine learning using Scikit-learn, Extracting components from HTML document through beautiful soup.</p>	6
5	<p>Probabilistic and Inferential Statistics:</p> <p>Random variable, basics of probability, Probability mass function, Probability density function, Cumulative distribution function, Discrete probability distribution (Binomial and Poisson), Continuous probability distributions (Normal and Exponential). Sampling and its various techniques, Estimation, point estimation and interval estimation, Standard normal distribution, Central limit theorem, Hypothesis testing through one sample test.</p>	5
6	<p>Data preparation:</p> <p>Data loading, Data cleaning, dealing with missing data, removing duplicates, Slicing and Dicing, Filtering, and selecting data, Concatenating and transforming, adding new cases and variables, removing data, Sorting and shuffling, Aggregating data, Handling outliers, Data wrangling, Data Normalization</p>	5



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Subject Code: BE05000231

Subject Name: Python for Data Science

7	Data Visualization: Visualizing data through figure, subplot and its properties, graphs, plots in matplotlib, basics of advanced visualizations with Seaborn	5
TOTAL		45

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

Distribution of Theory Marks in %					
R Level	U Level	A Level	N Level	E Level	C Level
30	20	25	25	00	00

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

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

The syllabus of *Python for data science* directly contributes to,

SDG 8	Decent work and economic growth
--------------	--

Reference Books:

1. Python for data science for dummies, John Paul Mueller, Luca Massaron, and Wiley
2. Pandas for everyone: Python Data Analysis, Daniel Y. Chen, Pearson
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, Arno D.B. Meysman, et al., Manning
4. Applied Data Science with Python and Jupyter: Use powerful industry-standard tools to unlock new, actionable insights from your data, Packt
5. Data Analytics, Anil Maheshwari, McGrawHill

List of Experiments:

1. At least 20 practical which includes all units of syllabus
2. To Prepare Minor Project (Group of 3 students per project).

List of Open-Source Software

1. Python
2. Python libraries

List of learning website:

1. www.python.org
2. www.learnpython.org



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering

Level: UG

Subject Code: BE05000231

Subject Name: Python for Data Science

3. www.hackr.io/tutorials/learn-python
4. www.sololearn.com/learning/1073
5. www.nptel.iitm.ac.in

List of suggested activities for Problem-based Learning (PBL):

Sl. No	PBL category	Name of the activity	No. of hours	Evaluation Criteria
1	Case Study Analysis / Seminar	Seminar / Presentation	Duration for study and preparation=10h Report writing=3h Presentation=2h Total=15h	Topic can be selected technical content beyond syllabus
2	Assignment / Technical Writing / Research Writing	Poster/chart/power point preparation on technical topics	Duration = 15h	Based on poster/chart preparation and presentation skills
3	Case Study Analysis / Seminar	Real world case studies-based learning	Duration of data collection/study = 5h Report preparation = 10h Total = 15h	Based on in-depth study, technical depth, data collected, fact finding, etc.
4	Self-Learning	Self-learning on-line course	Minimum duration of the course should be 15h.	Examination based assessment at the end of course. Based on the certificate produced.
5	Micro Project	Application/Software development (Mini Project)	Duration = 15 h	Depending on the complexity of the Application/Software
6	Video Based Learning	Technical Video based learning related to the subject	Duration of video = 5h Report preparation & Presentation = 10h Total = 15h	Report /presentation based on the video learning outcomes.

Note:

1. The hours allocated to specific activities should be proportionate to the total no. of PBL hours and marks.
