

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER-V EXAMINATION – WINTER 2025****Subject Code:3150713****Date:17-11-2025****Subject Name:Python for Data Science****Time:10:30 AM TO 01:00 PM****Total Marks:70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

		MARKS
Q.1	(a) What are the advantages of python as a preferred programming language for data science?	03
	(b) Write a program that takes a list of words as input and returns the longest word in the list.	04
	(c) In EDA, why is it essential to analyze both covariance and correlation when exploring relationships between variables? Provide examples of scenarios where covariance might indicate a strong relationship while correlation might suggest otherwise.	07
Q.2	(a) Describe the difference between a list and a tuple in Python (Provide at-least three points).	03
	(b) Implement a function that takes a string as input and returns the count of each character in the string as a dictionary.	04
	(c) Explain Data Science Pipeline in detail.	07
	OR	
	(c) Write a note on python ecosystem for data science.	07
Q.3	(a) What are the advantages of Jupyter Notebook?	03
	(b) Explain how one-hot encoding is performed on a given text snippet, such as "The quick brown fox jumps over the lazy dog." What specific steps are involved in transforming this text into one-hot encoded vectors?	04
	(c) Describe the process of parsing an HTML document using the BeautifulSoup library in Python.	07
	OR	
Q.3	(a) Discuss the advantages of pandas library in python.	03
	(b) Explain the process of calculating the TF-IDF score for a term in a document with proper example.	04
	(c) Discuss different options to deal with missing data as part of data cleaning process.	07
Q.4	(a) Discuss importance of pie chart and box plot in data visualization.	03
	(b) What are magic functions? Briefly explain any two magic functions.	04
	(c) Write code to create an undirected graph of six nodes using networkx and matplotlib.	07
	OR	
Q.4	(a) How can we add labels, annotations and legends in our graphs using matplotlib?	03
	(b) Given the sample CSV file "trigonometry_data.csv" containing trigonometry data (Angle (degrees), Sine, Cosine, Tangent), how would	04

you use Pandas to read the file in DataFrame and print the angle in degrees and its corresponding cosine value from the DataFrame?
Content of “trigonometry_data.csv” file is as below.

Angle (degrees), Sine, Cosine, Tangent

0,0.0,1.0,0.0

30,0.5,0.866,0.577

45,0.707,0.707,1.0

60,0.866,0.5,1.732

90,1.0,0.0,undefined

- (c) What is the importance of histogram in data visualization? Write code to create histogram using matplotlib. **07**
- Q.5** (a) What role does the chi-square test play in data analysis? **03**
(b) Briefly discuss skewness and kurtosis. **04**
(c) Describe various methods for slicing rows and columns in a Pandas DataFrame, supported by relevant examples. **07**
- OR**
- Q.5** (a) What is z-score standardization and why it is useful in EDA? **03**
(b) Explain working of %timeit and %memit magic functions. **04**
(c) Elaborate on the concepts of sorting data row-wise, column-wise, and shuffling data within a DataFrame by providing appropriate examples. **07**
