GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–V (NEW) EXAMINATION – SUMMER 2024

Subject Code:3150713   **Date:16-05-2024**
Subject Name:Python for Data Science
Time:02:30 PM TO 05:00 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)**	Briefly discuss major coding styles supported by python programming language.	**03**
	(b)	What will be the output of the below python code?	**04**

```
import numpy as np
a = np.array([[1, 2.5, 3],[4, 5, 6.5]])
print(a.shape)
print(a.ndim)
print(a.itemsize)
print(a.nbytes)
```

	(c)	Write a python program to take two strings and position value (starting from zero) as input from the user. Insert the second string in the first string after the given position. (Example: String-1 = "Gujarat University", String-2="Technological ", position = 7, Output="Gujarat Technological University")	**07**
Q.2	**(a)**	Discuss characteristics of Set and Dictionary data structures in python.	**03**
	(b)	Write a short note on TF-IDF transformations.	**04**
	(c)	Explain different stages of the data science pipeline.	**07**

OR

	(c)	Discuss major libraries available with python language with its application in the field of Data Science.	**07**
Q.3	**(a)**	Explain different data structures available in pandas library.	**03**
	(b)	Below is the content of a csv file named "data.csv".	**04**

```
Name,Salary
abc,10000
xyz,20000
pqr,40000
```

		Write a python program to read data from given csv file and print each pair of names and salaries in the format "Name -> Salary". (Example: abc -> 10000 and so on)	
	(c)	Why it is important to find missing values from a dataset? Discuss different approaches to handle missing values with appropriate examples.	**07**

OR

Q.3	**(a)**	Write a short note on categorical variables.	**03**
	(b)	Consider below dataframe in python	**04**

A   B   C

0	7	1	4
1	2	5	8
2	3	6	9

     i.   Write a code to sort this dataframe by column A.

     ii.   Write a code to sort this dataframe by row 0 (first row).

(c) Explain the process of Word Tokenization, Stemming and Stop word removal using nltk library with suitable example code.   **07**

Q.4 **(a)** Discuss how to create pie chart using matplotlib library with suitable example.   **03**

(b) Demonstrate how to remove a row and a column from a dataframe with appropriate example code.   **04**

(c) Data of marks for five students in three subjects is as given below.   **07**

Maths = [100, 82, 76, 51, 88]

Physics = [92, 89, 81, 83, 80]

Chemistry = [71, 90, 75, 88, 45]

Write a python program to create boxplot from the above data. What type of important observation about data can be derived from the boxplot?

OR

Q.4 **(a)** Explain how to add markers and change line color and line style in a line graph using matplotlib library.   **03**

(b) An adjacency matrix for a graph with for nodes named 0, 1, 2, 3 is given below.   **04**

	0	1	2	3
0	0	1	1	0
1	0	0	0	0
2	0	1	0	1
3	1	0	0	0

Write a python program to draw the undirected graph from above matrix using networkx library.

(c) Below is the data about average temperature (in °C) at a place for 15 days.   **07**

temp = [30, 34, 35, 32, 38, 26, 29, 45, 42, 32, 40, 33, 36, 34, 36]

Write a python program to create histogram with 4 bins from the above data. What type of important observation about data can be derived from the histogram?

Q.5 **(a)** What is the use of %timeit and %%timeit magic functions?   **03**

(b) Discuss importance of covariance and correlation in EDA.   **04**

(c) What is regression and classification problems? Explain any one with suitable example.   **07**

OR

Q.5 **(a)** Explain different interfaces from Scikit-learn library.   **03**

(b) Explain the importance of chi-square test in EDA.   **04**

(c) Write a note on different descriptive statistics measures for numeric data.   **07**

****