Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER–V (NEW) EXAMINATION – WINTER 2020

Subject Code:3150713 Date:22/01/2021

Subject Name:Python for Data Science

Time:10:30 AM TO 12:30 PM Total Marks: 56

Instructions:

1. **Attempt any FOUR questions out of EIGHT questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

			MARKS
Q.1	**(a)**	Discuss the role of indentation in python.	**03**
	(b)	Explain range() function with suitable examples.	**04**
	(c)	Write a python program to find the factorial of a given number using recursion.	**07**
Q.2	**(a)**	Explain sampling in terms of data science?	**03**
	(b)	List and explain different coding styles supported by python.	**04**
	(c)	Discuss why python is a first choice for data scientists?	**07**
Q.3	**(a)**	Explain TF-IDF transformations.	**03**
	(b)	Explain categorical variables in detail.	**04**
	(c)	Write a python program to read the data from XML file using pandas library.	**07**
Q.4	**(a)**	Describe date time transformation using datetime module.	**03**
	(b)	Explain a bag of words model in detail.	**04**
	(c)	Explain imputation in detail with example.	**07**
Q.5	**(a)**	List the features of matplotlib.	**03**
	(b)	Write a python program to read data from a text file using pandas library.	**04**
	(c)	Explain time series plot with appropriate examples.	**07**
Q.6	**(a)**	List the type of plots that can be drawn using matplotlib.	**03**
	(b)	Write a python program to read data from CSV files using pandas.	**04**
	(c)	Explain pie chart plot with appropriate examples.	**07**
Q.7	**(a)**	List and explain interfaces of SciKit-learn.	**03**
	(b)	List the multiprocessing tasksthat can be done using SciKit-learn?	**04**
	(c)	Define the classification problem. How can it be solved using	**07**

SciKit-learn?

Q.8	**(a)**	Define EDA. List the tasks need to be carried out in EDA?	**03**
	(b)	How hash functions can be useful to solve data science problems?	**04**
	(c)	Define the regression problem. How can it be solved using SciKit-learn?	**07**

****

Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–V (NEW) EXAMINATION – WINTER 2021

Subject Code:3150713 Date:15/12/2021

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)**	What is the role of Python in Data science?	**03**
	(b)	Differentiate List and Tuple in Python	**04**
	(c)	Explain data science pipeline in details.	**07**
Q.2	**(a)**	What is HTML parsing?	**03**
	(b)	Write a python code to find factorial of number using function.	**04**
	(c)	Explain Dictionary in Python with example	**07**
		OR	
	(c)	Is String a mutable data type? Also explain the string operations length, indexing and slicing in detail with an appropriate example	**07**
Q.3	**(a)**	List and explain any three Magic function.	**03**
	(b)	Explain Slicing rows and columns with example.	**04**
	(c)	What do you mean by missing values? Explain the different ways to handle the missing value with example.	**07**
		OR	
Q.3	**(a)**	What is Categorical Variables? Explain it with example.	**03**
	(b)	How to read data from relational database? Briefly explain it.	**04**
	(c)	What is the use of following operations on Panda's Data Frames? Explain with a small example of each. 1. shape 2. tail() 3. describe()	**07**
Q.4	**(a)**	Explain hist() function with code.	**03**
	(b)	Write a program using Numpy to count number of "C" element wise in a given array.	**04**
	(c)	What do you understand by Data visualization? Discuss some Python's data visualization techniques.	**07**
		OR	
Q.4	**(a)**	Explain bar() function with code.	**03**
	(b)	What are the different ways to remove duplicate values from dataset?	**04**
	(c)	Write a simple python program that draws a line graph where x = [1,2,3,4] and y = [1,4,9,16] and gives both axis label as "X-axis"and "Y-axis".	**07**
Q.5	**(a)**	What is Scikit-learn?	**03**
	(b)	Explain Box plot with example.	**04**

	(c)	Write a Python programming to create a pie chart with a title of the popularity of programming Languages. Sample data: Programming languages: Java, Python, PHP, JavaScript, C#, C++ Popularity: 22.2, 17.6, 8.8, 8, 7.7, 6.7	**07**

OR

Q.5	**(a)**	Define covariance and correlation	**03**
	(b)	Explain scatterplots with example.	**04**
	(c)	What is Data Wrangling process? Define data exploratory data analysis? Why EDA is required in data analysis?	**07**

********All THE BEST*********

Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–V (NEW) EXAMINATION – SUMMER 2021

Subject Code:3150713 Date:07/09/2021

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)**	Differentiate the list and dictionary data types of python by their characteristics along with example in brief.	**03**
	(b)	What do you mean by slicing operation in string of python? Write an example of slicing to fetch first name and last name from full name of person and display it.	**04**
	(c)	Which are the basic activities we performed as a part of data science pipeline? Summarize and explain in brief.	**07**
Q.2	**(a)**	What is the core competencies needed to become a data scientist? Explain in brief.	**03**
	(b)	Compare and summarize four different coding styles supported by Python language.	**04**
	(c)	Summarize the characteristics of NumPy, Pandas, Scikit-Learn and matplotlib libraries along with their usage in brief.	**07**
		OR	
	(c)	What do you mean by prototyping? List the phases of prototyping and experimentation process and explain in brief.	**07**
Q.3	**(a)**	Compare the numpy and pandas on the basis of their characteristics and usage.	**03**
	(b)	For what purpose sampling is used. Demonstrate random sampling with example.	**04**
	(c)	What is the need of streaming the data? Explain data uploading and streaming data with example.	**07**
		OR	
Q.3	**(a)**	How XPath is useful for analysis of html data? Explain in brief.	**03**
	(b)	Define term n-gram. Explain the TF-IDF techniques.	**04**
	(c)	List the techniques to handle missing data. Explain various techniques with example.	**07**
Q.4	**(a)**	List various types of graph/chart available in the pyplot of matplotlib library for data visualization. Explain any two of them in brief.	**03**
	(b)	What kind data is analyzed with Bag of word model? Explain it with example.	**04**
	(c)	What do you mean by time series data? How can we plot it? Explain it with example to plot trend over time	**07**
		OR	

Q.4	**(a)**	Compare bar graph, box-plot and histogram with respect to their applicability in data visualization.	**03**
	(b)	Define stemming. Explain the concept of stemming with example.	**04**
	(c)	What is the use of scatter-plot in data visualization? Can we draw trendline in scatter-plot? Explain it with example.	**07**
Q.5	**(a)**	Define the term Data wrangling. Explain the steps needed to perform data wrangling.	**03**
	(b)	Why we need to perform Z-score standardization in EDA? Justify it with example.	**04**
	(c)	What is the use of hash function in EDA? Express various hashing trick along with example.	**07**

OR

Q.5	**(a)**	What do you mean by Exploratory Data Analysis (EDA)? How t-test is useful for EDA?	**03**
	(b)	What do you mean by covariance? What is the importance of covariance in data analysis? Explain it with example.	**04**
	(c)	List different way for defining descriptive statistics for Numeric Data. Explain them in brief.	**07**

****

Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–V(NEW) EXAMINATION – SUMMER 2022

Subject Code:3150713 Date:02/06/2022

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)**	List Advantages of Python.	**03**
	(b)	Differentiate Numpy and Pandas.	**04**
	(c)	Explain Exploratory Data Analysis (EDA).	**07**
Q.2	**(a)**	Explain String Slicing in python with Example.	**03**
	(b)	List and Explain different programming styles in python.	**04**
	(c)	Write a program to check whether the given number is prime or not.	**07**
		OR	
	(c)	Write a program to print Fibonacci series up to number given by user.	**07**
Q.3	**(a)**	Differentiate rand and randn function in Numpy.	**03**
	(b)	Explain DataFrame in Pandas with example.	**04**
	(c)	Write a program to print following patterns.	**07**

1)

```
*
* *
* * *
* * * *
```

2)

```
$ $ $ $
$ $ $
$ $
$
```

3)

```
# # # # #
 # # #
  #
 # # #
# # # # #
```

		OR	
Q.3	**(a)**	Explain Groupby function in pandas with example.	**03**
	(b)	Explain how to deal with missing data in Pandas.	**04**
	(c)	Explain Web Scrapping with Example using Beautiful Soup library.	**07**
Q.4	**(a)**	Explain Bag of Word model.	**03**
	(b)	Differentiate join and merge functions in pandas.	**04**
	(c)	Write a program which takes 2 digits, X,Y as input and generates a 2-dimensional array of size X * Y. The element value in the i-th row and j-th column of the array should be i*j.	**07**

OR

Q.4	**(a)**	Explain Hashing Trick in python with example.	**03**
	(b)	Write a brief note on NetworkX library.	**04**
	(c)	List and Explain different graphs in MatPlotLib.	**07**
Q.5	**(a)**	Explain Labels, Annotation and Legends in MatPlotLib.	**03**
	(b)	Differentiate Supervised and Unsupervised learning.	**04**
	(c)	Explain Regression with example.	**07**

OR

Q.5	**(a)**	Write a program to print Current date and time.	**03**
	(b)	Write a program to interchange the List elements on two positions entered by a user	**04**
	(c)	Explain Classification with example.	**07**

****