Seat No.: _____   Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY
ME – SEMESTER – I (New)– EXAMINATION – WINTER-2019

Subject Code: 3710219   **Date: 07-01-2020**

Subject Name: 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.**

Q.1	(a)	What is data science? Why did a need for data science emerged? Briefly describe the use of data science in business, decision making and predictive analytics.	7
	(b)	What are various data types used in data science? Explain them with examples. List the statistical methods applicable for those data types.	7
Q.2	(a)	Which measures are used to analyze the central tendency of data?	3
	(b)	Calculate population standard deviation and sample standard deviation for the given ages of 8 students with following ages: 20, 21, 21, 20, 19, 17, 17, and 19.	4
	(c)	What is the significance of data science in agriculture and e-commerce? Explain any two applications in both domains.	7

OR

	(c)	Describe the life cycle of data science process.	7
Q.3	(a)	Write down a program in python to implement SVM.	7
	(b)	Mention various tools available for data science toolkit.	3
	(c)	How the missing values in data set can be compensated? Explain.	4

OR

Q.3	(a)	Differentiate between supervised and unsupervised learning models. Name any three machine learning algorithms. Explain Naïve Bayes.	7
	(b)	What is SVM? How to identify the right hyper-plane? Explain with different scenarios.	7
Q.4	(a)	What does Central Limit Theorem (CLT) state?	3
	(b)	Explain its significance with respect to data distribution using example.	4
	(c)	i) Explain the purpose of following commands:	7

  Dataset.head()
  Dataset.head(n)
  Dataset.describe()
  Dataset.info()

  ii) Explain the use of following python libraries:
  NumPy
  Pandas

OR

Q.4 (a) Exemplify the different visual encoding variables? 3

(b) Consider the following data: 4

Item Types	Quantity
Books	10
Note-Books	9
Pencils	8
Pens	9

i) Which are the possible choices of retinal variables for item types and quantity?
ii) Depict data given in the above table applying retinal visual encoding.

(c) What are different types of data visualization? Explain three out of them. 7

Q.5 (a) How to select a particular plotting library out of Matplotlib, Seaborn, Bokeh? 2

(b) How a HoverTool is applied for providing the interactivity with the data using Bokeh. 5

(c) Why data cleaning or data cleansing is performed on the data? Explain the process. 7

OR

Q.5 (a) What are the issues in classification and prediction? Explain the classification process. 7

(b) For what purpose Pearson's correlation measure is used? Mention its formula. 3

(c) Calculate the person correlation for the given values of x and y. 4

	x	y
1	40	52000
2	24	33000
3	33	15000
4	37	26000
5	64	29000
6	45	32000
7	26	68000
8	57	45000
9	18	80000
10	25	41000

***