

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE- SEMESTER-V (NEW) EXAMINATION – WINTER 2020****Subject Code:3151608****Date:22/01/2021****Subject Name: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
<b>Q.1</b>	(a) Define business analytics and explain the use of it.	<b>03</b>
	(b) Give the difference between descriptive analytics and predictive analytics.	<b>04</b>
	(c) Explain the Framework for Data-Driven Decision Making process.	<b>07</b>
<b>Q.2</b>	(a) Define continuous random variable and discrete random variable with example.	<b>03</b>
	(b) Explain different categories of data.	<b>04</b>
	(c) Explain Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of a Continuous Random Variable with suitable example.	<b>07</b>
<b>Q.3</b>	(a) Below is the dataset of Pizza Price in given cities .Find Mean and Median of both the cities.	<b>03</b>

A	B	C
places	New Delhi	Lucknow
1	1\$	1\$
2	2\$	2\$
3	3\$	3\$
4	3\$	4\$
5	4\$	5\$
6	5\$	6\$
7	6\$	7\$
8	7\$	8\$
9	9\$	9\$
10	11\$	10\$
11	66\$	

<b>Q.3</b>	(b) What is the need of Skewness and Kurtosis. Explain its types with example.	<b>04</b>
	(c) How the Chi-square distribution is differ from student's t-distribution explain with example.	<b>07</b>
	<b>Q.4</b>	(a) Give the difference between Probabilistic Sampling and Non-Probability Sampling.
	(b) Explain why Central Limit Theorem is called as a heart of the Data Science.	<b>04</b>

	(c) Explain the importance of Sensitivity and specificity in logistic regression Classification Table.	07
<b>Q.5</b>	(a) Explain different types of Data Measurement scales.	03
	(b) How do you calculate maximum likelihood estimation?	04
	(c) Explain Simple Linear Regression model with example.	07
<b>Q.6</b>	(a) What is Outlier Analysis explain in detail.	03
	(b) Compare linear regression vs. Logistic regression.	04
	(c) Explain the Validation process of the Simple Linear Regression Model.	07
<b>Q.7</b>	(a) How to select Variable Selection in Logistic Regression.	03
	(b) Explain pros and cons of Decision Tree algorithm.	04
	(c) Explain Decision tree algorithm with suitable example.	07
<b>Q.8</b>	(a) Which classification algorithm is preferable when number of records are very large, random forest/ decision tree. Justify your answer.	03
	(b) Explain Chi-Square Automatic Interaction Detection (CHAID) in detail.	04
	(c) Explain Random forest algorithm with suitable example.	07

\*\*\*\*\*