Seat No.: _____   Enrolment No._____

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
Q.1	**(a)**	Define business analytics and explain the use of it.	**03**
	(b)	Give the difference between descriptive analytics and predictive analytics.	**04**
	(c)	Explain the Framework for Data-Driven Decision Making process.	**07**
Q.2	**(a)**	Define continuous random variable and discrete random variable with example.	**03**
	(b)	Explain different categories of data.	**04**
	(c)	Explain Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of a Continuous Random Variable with suitable example.	**07**
Q.3	**(a)**	Below is the dataset of Pizza Price in given cities .Find Mean and	**03**

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$	

Median of both the cities.

	(b)	What is the need of Skewness and Kurtosis. Explain its types with example.	**04**
	(c)	How the Chi-square distribution is differ from student's t-distribution explain with example.	**07**
Q.4	**(a)**	Give the difference between Probabilistic Sampling and Non-Probability Sampling.	**03**
	(b)	Explain why Central Limit Theorem is called as a heart of the Data Science.	**04**

1

	(c)	Explain the importance of Sensitivity and specificity in logistic regression Classification Table.	**07**
Q.5	**(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**
Q.6	**(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**
Q.7	**(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**
Q.8	**(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**

****