Q. P. Code: 27561

(3 Hours) Max. Marks**: 80**

N.B.: (1) Question No. **1** is **compulsory**.
(2) Attempt any **Three** questions out of remaining **Five** questions.
(3) **Figures** to the **right** indicate **full** marks.
(4) Assume suitable data if **necessary**.

Qu-1		**Attempt the following.**	**20**
	a)	Explain Simple Linear Regression.	
	b)	Exploratory Data Analysis (EDA)? How Does Exploratory Data Analysis Differ from Summary Analysis?	
	c)	Explain Data Visualization basics.	
	d)	Explain Model-Based Clustering in short.	
Qu-2	**a)**	Explain data science process in detail with the help of diagram.	**10**
	b)	One of the great strengths of R is the user's ability to add functions. In fact, many of the functions in R are actually functions of functions. Give the Syntax for writing Functions in R and write a User Defined Functions (UDF) to compute the Factorial of given number.	**10**
Qu-3	**a)**	Give a detailed description of K-Nearest Neighbor (KNN) Algorithm and state clearly i) When do we use KNN algorithm? ii) How do we choose the factor K?	**10**
	b)	Explain Text analysis steps with a suitable text analysis example.	**10**
Qu-4	**a)**	Explain Logistic Regression in detail.	**10**
	b)	Explain sentiment analysis with suitable example.	**10**
Qu-5	**a)**	Explain the Global Innovation Network and Analysis Case Study with following: i) Business Problem Framed ii) Initial Hypotheses iii) Data iv) Model Planning Analytic Technique v) Result and Key Findings	**20**
Qu-6		**Write short note on**	**20**
	a)	Data science vs Bl	
	b)	Support vector machine	
	c)	Interactive dashboards	
	d)	TF and TFIDF	

****