PRN:		Course Code:	**BCA302**

000708

Third Year **B.C.A.** Semester End Examination Winter 2024

Data Science

Day and Date: Sat 30 NOV 2024 Max Marks: **100**

Time: 02:00PM To 05:00PM Duration: **03** hrs

Instructions: 1. All questions are compulsory.
2. Draw Diagrams whenever necessary.
3. Calculator is allowed.

			Marks	Bloom's Level	CO
Q.1		**Attempt the following**			
	a)	Discuss Data Science and its Life Cycle	08	2	1
	b)	Explain integrated development Environment (IDE)	08	2	1
		OR			
	b)	Elaborate exploratory data analysis and data visualization.	08	2	1
Q.2		**Attempt the following**			
	a)	Explain issues of data integration with suitable examples	08	2	1
	b)	Describe various data sources and methods of data collection.	08	2	1
		OR			
	b)	Explain Data cleaning, extraction and data transformation with python commands.	08	2	1
Q.3		**Attempt any Two from the following**			
	a)	Describe schema and schema-based query language.	08	2	2
	b)	Analyze data curation with its benefits.	08	4	3
	c)	Differentiate Structured, Unstructured and Semi Structured data with a suitable example.	08	2	2
Q.4		**Attempt any Two from the following**			
	a)	Analyze over fitting, under fitting and good fitting with its diagram.	08	4	4
	b)	Elaborate statistical Modelling techniques in data analysis.	08	2	4
	c)	Explain cross validation in detail with its types.	08	2	3
Q.5		**Attempt any Two of the following**			
	a)	Distinguish Linear regression and logistic regression.	09	3	4
	b)	Identify time series analysis with types.	09	3	4
	c)	Make use of information on ages and cholesterol levels for a random sample of 10 men.	09	4	4

Backlog

Age	58	69	43	39	63	52	47	31	74	36	
Cholesterol level	189	235	193	177	154	191	213	165	198	181	

a. Taking age as an independent variable and cholesterol level as a dependent variable, compute SSxx, SSyy, and SSxy.
b. Find the regression of cholesterol level on age.

Q.6 **Attempt any Two of the following**

a)	Describe Clustering and its types.	09	3	4
b)	Describe hierarchical clustering for the one-dimensional data and how to plot dendrogram to visualize it.	09	4	4
c)	Illustrate K-means clustering algorithm with an example.	09	2	4

Backlog