### PSG COLLEGE OF ARTS & SCIENCE (AUTONOMOUS)

### **MSc DEGREE EXAMINATION DECEMBER 2023**

(First Semester)

### Branch - COMPUTER SCIENCE

### PYTHON AND R PROGRAMMING FOR DATA SCIENCE

Time: Three Hours

Maximum: 75 Marks

### SECTION-A (10 Marks)

Answer ALL questions

ALL questions carry EQUAL marks  $(10 \times 1 = 10)$ 

	ALL questions carry EQUAL marks (10 x 1 = 10)					
Question No.	Question	K Level	со			
1	Which of the following has a root in statistics?  a) Data science b) Python c) R d) Operating system	K1	CO1			
2	Which symbol is used for assignment in R?  a) = b) == c) <- d) =>	K2	CO1			
3	Which type of data visualization is best suited for showing the distribution of a single continuous variable?  a) Scatter plot b) Bar chart c) Histogram d) Pie chart	K1	CO2			
4	What does a kernel density plot represent?  a) The distribution of data points in a scatter plot b) The correlation between two variables c) The composition of a whole divided into parts d) The estimated probability density of a continuous variable	K2	CO2			
5	Which ggplot2 function is used to create a bar chart?  a) geom_line() b) geom_histogram() c) geom_point() d) geom_bar()	K1	CO3			
6	What is the primary purpose of the matplotlib library in Python?  a) Data preprocessing b) Statistical analysis c) Data visualization d) Machine learning	K2	CO3			
7	<ul> <li>What is multicollinearity in the context of regression analysis?</li> <li>a) A situation where there are multiple dependent variables</li> <li>b) A situation where two or more independent variables are highly correlated</li> <li>c) A situation where the target variable is missing</li> <li>d) A situation where outliers are present in the data</li> </ul>	K1	CO4			
8	What is a word cloud commonly used for in text analytics?  a) Analyzing numerical data b) Identifying the most frequently used words in a text c) Extracting numerical insights from text d) Generating pie charts	K2	CO4			
9	Which deep learning framework is known for its flexibility and is widely used in both research and industry?  a) TensorFlow b) PyTorch c)Keras d) Caffe	K1	CO5			
10	How can data analytics benefit agriculture?  a) By controlling the weather  b) By improving crop yield and quality c) By automating the planting process d) By creating genetically modified crops	K2	CO5			

## SECTION - B (35 Marks)

Answer ALL questions
ALL questions carry EQUAL Marks

 $(5\times7=35)$ 

Question					
No.	Question	K Level	СО		
11.a.	What are the ethical and privacy considerations associated with Data Science?				
(OR)					
11.b.	Explain the key differences between Python and R as programming languages for Data Science, highlighting their strength and common use cases.	K2	CO1		
12.a.	Examine the functions in R.				
	(OR)	K4	CO2		
12.b.	Explain the significance of understanding data preprocessing process.				
13.a.	Interpret on scatter plot, strip chart, stacked bar chart, and grouped bar chart using R.				
(OR)		K2	CO3		
13.b.	Illustrate how Violin plot is used to is used to display multiple nodes?				
14.a.	Assess the notion of factor analysis for an application using python.				
	(OR)				
14.b.	Explain word cloud and its uses in text analytics.				
15.a.	Compare and contrast Deep learning and Machine learning techniques.				
(OR)		K6	CO5		
15.b.	Elaborate the different applications of analytics in the field of agriculture.				

# SECTION -C (30 Marks)

Answer ANY THREE questions

ALL questions carry EQUAL Marks

 $(3 \times 10 = 30)$ 

Question No.	Question	K Level	СО
16	Explain the different types of libraries and its uses in Python.	K5	CO1
17	How important is data visualization techniques and explain its role in decision making in different domains?	K5	CO2
18	Analyze how to perform data visualization using Ggplots library in R? Justify any three Ggplot libraries with examples and also state its importance.	K4	CO3
19	Analyze the relationship between variables with an example.	K4	CO4
20	Generalize building blocks and role of agents in reinforcement training.	K6	CO5