

PSG COLLEGE OF ARTS & SCIENCE  
(AUTONOMOUS)

MSc DEGREE EXAMINATION MAY 2024  
(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 × 1 = 10)

Question No.	Question	K Level	CO
1	Which function is used to read data from an external CSV file in R? a) read.csv()    b) import_data()    c) load_file()    d) read_data()	K1	CO1
2	What symbol is used for a single-line comment in Python? a) //    b) --    c) #    d) /* */	K2	CO1
3	Which step involves encoding categorical variables into numerical format? a) Data Cleaning    b) Data Transformation c) Data Integration    d) Data Reduction	K1	CO2
4	Which Python library is commonly used for creating static data visualizations? a) Pandas    b) NumPy    c) Matplotlib    d) Scikit-Learn	K2	CO2
5	What is the purpose of the ggplot() function in ggplot2? a) To create a blank canvas for a plot b) To add data points to a plot c) To customize the appearance of the x-axis d) To save a plot to a file	K1	CO3
6	Which library in Python is commonly used to create violin plots? a) Pandas    b) Seaborn    c) Matplotlib    d) Plotly	K2	CO3
7	In correlation analysis, what does a positive correlation coefficient (r) indicate? a) A strong negative relationship between variables b) No relationship between variables c) A strong positive relationship between variables d) An inverse relationship between variables	K1	CO4
8	What is web analytics? a) Analyzing the content of web pages b) Measuring website traffic and user behavior c) Conducting sentiment analysis on social media posts d) Analyzing the performance of web servers	K2	CO4
9	What is deep learning? a) A type of machine learning that uses shallow neural networks b) A branch of artificial intelligence focused on decision trees c) A subset of machine learning that uses deep neural networks with multiple layers d) A technique for training models with labeled data	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

Cont...

**SECTION - B (35 Marks)**Answer **ALL** questions**ALL** questions carry **EQUAL** Marks

(5 × 7 = 35)

Question No.	Question	K Level	CO
11.a.	What are the types of analytics and give some real world applications where they are commonly used?	K2	CO1
	(OR)		
11.b.	Summarize on Data science process model.		
12.a.	Discuss on Functions in Python.	K4	CO2
	(OR)		
12.b.	Examine the Steps involved in EDA Using Python Programming.		
13.a.	Explain the box plot with quartiles.	K2	CO3
	(OR)		
13.b.	Explain how violin plot is used to display multiple nodes.		
14.a.	Compare Independent and Dependent Variables.	K3	CO4
	(OR)		
14.b.	Elucidate web and social media analytics.		
15.a.	Differentiate Machine Learning and Deep Learning.	K6	CO5
	(OR)		
15.b.	Examine the convolution Neural Networks in detail.		

**SECTION - C (30 Marks)**Answer **ANY THREE** questions**ALL** questions carry **EQUAL** Marks

(3 × 10 = 30)

Question No.	Question	K Level	CO
16	Describe the role of open-source tools and libraries, such as Python and R, in the practice of Data Science and their contributions to the field.	K5	CO1
17	Summarize the significance of understanding data preprocessing process.	K5	CO2
18	Explain the significance of Violin Plot, Swarm Plot, Rug Plot, Cat Plot, Pair Plot in detail.	K3	CO3
19	Explain the concept of sentiment analysis in the context of social media and discuss its applications, challenges, the impact it has on businesses and decision-making with a real time example.	K4	CO4
20	Discuss the future of AI and how might AI impact society and the job market in the coming decades, and what steps can be taken to ensure a positive outcome.	K6	CO5

Z-Z-Z

END