

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code: U1802

M.E. DEGREE EXAMINATION, DEC 2025

First Semester

Computer Science and Engineering (Data Science)

21PDS102 – PRINCIPLE OF DATA SCIENCE

(Regulations 2021)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART - A (5 x 20 = 100 Marks)

1. (a) Demonstrate how data science can be used in healthcare to improve diagnosis accuracy. CO1-U (20)
Or
(b) Write a detailed note on the lifecycle of a data science project. CO1-U (20)
2. (a) Explain the workflow of a simple machine learning project using Scikit-learn. CO1-U (20)
Or
(b) Explain how Scikit-learn is used for classification and regression problems. CO1-U (20)
3. (a) Using a dataset of your choice, create visualizations (such as line charts, scatter plots, or box plots) to show how data visualization helps in detecting patterns, trends, and anomalies. Explain what each visualization reveals about the data. CO2-App (20)
Or
(b) Implement Python code to drop the column 'Address' from df without affecting the original Data Frame. CO2-App (20)
4. (a) Using a sample dataset, calculate and interpret a 95% confidence interval for the mean using Python. Explain how this interval helps in understanding the uncertainty of your estimate. CO2-App (20)
Or
(b) Using a sample dataset, demonstrate how sampling, estimation, and hypothesis testing are connected. Select a random sample, estimate the population mean, and perform a hypothesis test to check if the sample mean differs significantly from a known population mean. CO2-App (20)

5. (a) Describe in detail the complete machine learning modeling process from data collection to prediction. CO1-U (20)

Or

(b) Explain cross-validation and its role in improving model performance. CO1-U (20)