(2 ½ Hours) [Total Marks: 75]

N.B. 1) All questions are compulsory.
2) Figures to the right indicate marks.
3) Illustrations, in-depth answers and diagrams will be appreciated.
4) Mixing of sub-questions is not allowed.

Q. 1 Attempt ANY FOUR from the following: **(20M)**

(a) Describe various domain specific applications of Data Science
(b) Differentiate between Data Science and Artificial Intelligence.
(c) Explain difficulties in handling unstructured data and explain how difficulties can be resolved.
(d) What is data cleaning? Discuss the steps involved in handling missing values and outliers.
(e) What is data wrangling? Describe data wrangling techniques.
(f) Why Sci-kit Learn is popular among data scientists, explain.

Q. 2 Attempt ANY FOUR from the following: **(20M)**

(a) What is EDA? Describe any four data visualization techniques applicable for EDA.
(b) How hypothesis testing can be done with the help of t-tests and chi-square tests? Give proper examples.
(c) Justify the comparison between supervised and unsupervised learning.
(d) Describe stepwise regression with its role in model selection.
(e) Discuss various techniques available for evaluating model performance.
(f) Write a note on Artificial Neural Network.

Q. 3 Attempt ANY FOUR from the following: **(20M)**

(a) Describe evaluating model for imbalanced data sets.
(b) What is the role of story telling in data science, explain.
(c) Discuss various data management activities used in ensuring data usability and quality.
(d) Explain data pipeline and the stages involved in ETL.
(e) Write a note on Seaborn and Tableau visualization tools.
(f) Describe various data privacy and security considerations in data management practices.

Q. 4 Attempt ANY FIVE from the following: **(15M)**

(a) Write advantages of using NumPy Library for data science in Python.
(b) Define Bias variance tradeoff in terms of Data Science.
(c) What is AUC? How it can be used?
(d) What is dummification? Give example.
(e) Describe application of simple linear regression in predictive modelling.
(f) Write python program to draw line chart. Assume your own data.

****