(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) Define Data Science. Discuss in detail applications of Data Science.
(b) Explain Data Warehousing (DW) and Data Mining (DM) in detail.
(c) Explain in detail different Data Sources.
(d) Describe in detail Data Transformation.
(e) Define Data Wrangling. Discuss Data Wrangling Techniques.
(f) Discuss about Feature Engineering and Time Series Data.

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

(a) Explain Data Visualization techniques in detail.
(b) Define Descriptive Statistics. Explain Mean, Median, Mode and Standard Deviation in detail.
(c) Explain in detail Classification and Regression analysis.
(d) Define bias, variance and discuss about bias-variance tradeoff.
(e) Discuss different techniques for evaluating model performance.
(f) Explain Ensemble Learning with Bagging and Boosting.

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

(a) Explain storytelling in analysis in detail.
(b) Discuss visualization tools in detail
(c) List and Discuss Data Management Activities.
(d) Elaborate the concept of Data Governance.
(e) Illustrate Extraction, Transformation and Load (ETL) in detail.
(f) Give the importance of Data Quality.

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

(a) Differentiate between structured and unstructured data.
(b) Explain Hyperparameter Tuning.
(c) Define Accuracy and Explain in brief.
(d) Discuss any three libraries of Data Science.
(e) Differentiate between underfitting and overfitting.
(f) Define Precision, Recall and F1-Score.

****