(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)   Discuss in detail different Data Sources.
(b)   Describe in detail Data Transformation.
(c)   Discuss the difference between Data Science & Business Intelligence.
(d)   Define and discuss Data warehousing and Data mining.
(e)   Explain any five libraries for data science.
(f)   Discuss in detail Data wrangling techniques.

Q. 2   Attempt ANY FOUR from the following: (20M)
(a)   Define Descriptive Statistics. Explain Mean, Mode, Median and Standard
        Deviation in detail.
(b)   Discuss Ensemble learning in detail.
(c)   Define Hypothesis Testing. Discuss Null Hypothesis and Alternate
        Hypothesis.
(d)   Explain Data Visualization Techniques in detail
(e)   Discuss in detail Support Vector Machine (SVM)
(f)   Explain in detail Dimensionality Reduction.

Q. 3   Attempt ANY FOUR from the following: (20M)
(a)   Discuss visualization tools in detail.
(b)   Discuss Data Management Activities in detail.
(c)   Define Accuracy, Precision, Recall, F1 Score and Confusion Matrix
(d)   Explain Data Quality along with it's importance.
(e)   Define Data Security. Discuss types of data security.
(f)   Discuss Data Privacy.

Q. 4   Attempt ANY FIVE from the following: (15M)
(a)   Discuss the difference between Data Science and Machine Learning.
(b)   Define Underfitting of data and Overfitting of data.
(c)   Discuss ETL (Extraction, Transformation and Loading)
(d)   Explain Data Preprocessing.
(e)   Discuss Bias, Variance Trade-off
(f)   Discuss the difference between Data governance and Data quality.

****