

(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)

- Discuss in detail different Data Sources.
- Describe in detail Data Transformation.
- Discuss the difference between Data Science & Business Intelligence.
- Define and discuss Data warehousing and Data mining.
- Explain any five libraries for data science.
- Discuss in detail Data wrangling techniques.

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

- Define Descriptive Statistics. Explain Mean, Mode, Median and Standard Deviation in detail.
- Discuss Ensemble learning in detail.
- Define Hypothesis Testing. Discuss Null Hypothesis and Alternate Hypothesis.
- Explain Data Visualization Techniques in detail
- Discuss in detail Support Vector Machine (SVM)
- Explain in detail Dimensionality Reduction.

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

- Discuss visualization tools in detail.
- Discuss Data Management Activities in detail.
- Define Accuracy, Precision, Recall, F1 Score and Confusion Matrix
- Explain Data Quality along with it's importance.
- Define Data Security. Discuss types of data security.
- Discuss Data Privacy.

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

- Discuss the difference between Data Science and Machine Learning.
- Define Underfitting of data and Overfitting of data.
- Discuss ETL (Extraction, Transformation and Loading)
- Explain Data Preprocessing.
- Discuss Bias, Variance Trade-off
- Discuss the difference between Data governance and Data quality.
