

(Time: 2 ½ Hours.)

[Total Marks: 75]

- N. B.:** (1) All questions are compulsory.
 (2) Make suitable assumptions wherever necessary and state the assumptions made.
 (3) Answers to the same question must be written together.
 (4) Numbers to the right indicate marks.
 (5) Draw neat labelled diagrams wherever necessary.
 (6) Use of non-programmable calculators is allowed.

1. Attempt any four of the following:

20

- What are the types of IDEs?
- Explain the types of exploratory data analysis.
- What makes Data Visualization Effective?
- Why is Data Analysis Important?
- Explain Data Cleaning process.
- Why Do We Need Data Collection?

2. Attempt any four of the following:

20

- Write a short note on Data Definition language(DDL).
- What is Distributed Database Storage? Explain with its types.
- Write in brief note on GitHub.
- Draw and Explain AWS in detail.
- Discuss any five operations in Relational Algebra.
- What is cloud computing? Explain with its types.

3. Attempt any four of the following:

20

- Explain the working of data aggregation.
- What are the different Regularization techniques?
- Describe the methods used for Cross-Validation.
- What are the common techniques of dimensionality reduction?
- Define Linear Model. Explain types of linear model.
- Demonstrate the working of k-means algorithm. State its Application.

4. Attempt any three of the following:

15

- What are the Different kinds of data?
- Describe the Data Analysis Process.
- Write a note on i. XQuery ii. XPath iii. JSON
- Write in brief about large scale data systems.
- What are Regression trees?
- Explain the Principal Components Analysis (PCA).
