

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 labeled diagrams wherever necessary.

(6) Use of Non-programmable calculators is allowed.

1. Attempt any two of the following:

12

- Explain Schema-on-Write and Schema-on-Read.
- Explain the following utilities: organize utilities, report utilities, maintenance utilities, backup utilities, Checks Data Integrity Utilities and History Cleanup Utilities.
- Explain the basics of operational management layer.
- What is MoSCoW? Explain.

2. Attempt any two of the following:

12

- What are processing-stream definitions? How are they managed? Explain.
- Explain data lakes and data swamps.
- Explain the following shipping terms: Seller, Carrier, Port, Ship, Terminal, Named Place and Buyer.
- How can the following be achieved? Explain with example.
  - Loading a table from .csv file
  - Display the datatype of column
  - Verification of data field name
  - Adding unique identifier for each row
  - Generate histogram across every column
  - Generate minimum and maximum values in each column

3. Attempt any two of the following:

12

- What are the four things that can be done with the errors found in data? Explain.
- What are the different ways of treating missing values in data using pandas package? Explain with example.
- "Graph theory is always a useful tool to use when relationships between business entities require analyzing." Explain with example.
- How can housekeeping be performed in data science to maximize processing capacity? Explain with example.

4. Attempt any two of the following:

12

- What are the different typical reference satellites? Explain.
- Explain the international classification of vehicles with examples.
- Explain the techniques for outlier detection and treatments.
- Explain hypothesis testing, t-test and chi-square test with respect to data science.

5. Attempt any two of the following:

12

- What are decision trees? How are they used in data science? Explain.
- What is tensor flow? Explain in detail.
- Explain the report superstep.
- Explain scatter matrix graph with example.