

(2 Hours)

[Total Marks: 60]

- 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

Q1. Attempt the two of the following.

12 Marks

- Explain data vault
- What are the different ways to deal with errors?
- Explain the Person section of TPOLE.
- Explain Logistic Regression.

Q2. Attempt the two of the following.

12 Marks

- Explain Python, Pandas, Numpy.
- Explain Yoke solution
- Explain the principles of data analysis.
- Explain the shipping terms: i) Seller ii) Carrier iii) Port iv) Ship

Q3. Attempt the two of the following.

12 Marks

- Explain in detail the audit layer
- Write a python program to drop the columns where any of the elements is missing values.
- What is an event? Explain explicit and implicit events.
- Explain Clustering Techniques.

Q4. Attempt the two of the following.

12 Marks

- Explain Apache Cassandra and Apache Spark.
- Explain the five fundamental data science process steps.
- Write a Python program to pick the content for Bill Boards from the given data.
- Explain the transform Superstep.

Q5. Attempt the two of the following.

12 Marks

- Explain Data Lake.
- Give an overview of the Drum-buffer-rope Method.
- Explain importance of Graphics and Picture in report.
- Explain Person-to-Time Sun Model.
