

MASTER OF COMPUTER APPLICATIONS (CBCS 2018 COURSE)
M.C.A. Sem-V : WINTER : 2021
SUBJECT: DATA SCIENCE

Day : Tuesday
Date : 11-01-2022

W-20589-2021

Time : 02:00 PM-05:00 PM
Max. Marks: 60

N.B.

- 1) Q.No. 4 from Section I is **COMPULSORY**.
 - 2) Answer **ANY TWO** questions from Q.1,2,3 in Section – I.
 - 3) Answer **ANY TWO** questions from Q.5,6,7 in Section – II.
 - 4) All questions carry **EQUAL** marks.
 - 5) Answer to both the sections should be written in **SAME** answer book.
 - 6) Draw labeled diagrams **WHEREVER** necessary.
-

SECTION – I

- Q.1** What is data analysis? What are its different roles? Explain role of simple statistical measures in it. (12)
- Q.2** Why data visualization is needed? List various visualization methods. Also explain any two visualization methods. (12)
- Q.3** What is Clustering? Explain Agglomerative Clustering in details. (12)
- Q.4** Write short notes on **ANY TWO** of the following: (12)
- a) Hypothesis testing
 - b) Cross validation
 - c) Data transformation
 - d) HIVE

SECTION – II

- Q.5** What is PCA? How it works? List its advantages and disadvantages. (12)
- Q.6** What is Linear regression? With suitable example elaborate its use. (12)
- Q.7** Consider standard Iris dataset. Describe each attribute in dataset. Write statements for followings. (12)
- i) Read data from data set.
 - ii) Display summary of data.
 - iii) Check correlation between attributes.
 - iv) Percentage of different classes in dataset.
 - v) With suitable chart, visualize the class distribution in dataset.
