Total No. of Questions : 5]

SEAT No. :

PD-1574

[Total No. of Pages : 2

[6468]-54
T.Y. B.Sc.
COMPUTER SCIENCE
CS - 354 : Foundations of Data Science
(2019 Revised) (CBCS) (Semester - V)

Time : 2 Hours] *[Max. Marks : 35*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Neat diagrams must be drawn wherever necessary.***

Q1) **Attempt any EIGHT of the following :** **[8 × 1 = 8]**

a) What do you mean by primary data?

b) What do you mean by Data Quality?

c) Define outlier.

d) Define interquartile range.

e) What do you mean by missing values?

f) What are uses of Zip files.

g) What do you mean by XML files data format?

h) Define data discretization.

i) What is tag cloud?

j) What is visual encoding?

Q2) **Attempt any FOUR of the following :** **[4 × 2 = 8]**

a) Explain applications of Data Science.

b) Explain null and alternate hypothesis.

c) What do you mean by Noisy data? Explain any two causes of noisy data.

d) What do you mean by data Visualization? Give example of any two data visualization libraries in Python.

e) List 3V's of data science.

P.T.O.

Q3) **Attempt any TWO of the following :** **[2 × 4 = 8]**

a) What do you mean by data Transformation? Explain any three strutegies for data transformation.

b) What is mean, mode, median and range for the following list of values:

24, 29, 24, 25, 24, 27, 25, 32, 24

c) Explain any four data Visualization tools?

Q4) **Attempt any TWO of the following :** **[2 × 4 = 8]**

a) Differentiate between structured and unstructured data.

b) What do you mean by Data attributes? Explain types of attributes with example.

c) How to visualize geospatial data? Explain in detail.

Q5) **Attempt any ONE of the following :** **[1 × 3 = 3]**

a) What do you mean by Data reduction? Explain data cube aggregation method.

b) Calculate the Standard Deviation for the below data:

Interval	0-10	10-20	20-30	30-40	40-50
Frequency	30	12	5	11	7