

SPPU-TE-COMP-CONTENT – KSKA Git

Total No. of Questions : 4]

SEAT No. : [REDACTED]

PD96

[Total No. of Pages : 2

[6410]-417

T.E. (Computer Engineering) (Insem) DATA SCIENCE AND BIG DATA ANALYTICS (2019 Pattern) (Semester - II) (310251)

Time : 1 Hour]

[Max. Marks : 30

Instructions to the candidates:

- 1) Answer Q.1 or Q.2, Q.3 or Q.4.
- 2) Neat diagrams must be drawn wherever necessary.
- 3) Figures to the right indicate full marks.
- 4) Assume suitable data if necessary.
- 5) Use of Scientific Calculator is permitted.

- Q1) a) What is data science? Compare data science and information science. [5]
- b) What are the steps of data preprocessing? What steps may the change the dimensions of the data? [5]
- c) What is data discretization? What are the different techniques used for data discretization? [5]

OR

- Q2) a) Describe different ways of handling missing data with suitable example. [5]
- b) In urban planning applications, devise the problems that can be solved using data science. [5]
- c) List and explain data reduction techniques with example. [5]
- Q3) a) For different levels of measurements like nominal, ordinal, interval and ratio, write the measures (among mean, median and mode) that can be used? [5]
- b) The table below shows the weights of members (in kg) in a sports club. Calculate mean, mode and median of the distribution. [5]

Weights	40-49	50-59	60-69	70-79	80-89	90-99
Frequency	6	8	12	14	7	3

- c) Explain the steps of hypothesis testing with suitable example. [5]

OR

P.T.O.

SPPU-TE-COMP-CONTENT – KSKA Git

- Q4) a) What is normal distribution and skewed distribution? What is the relation between mean, mode and median values in both the cases? [5]
- b) List out measures of dispersion with their significance and mathematical formulae. [5]
- c) Imagine a company wants to test the claim that their batteries last more than 40 hours. Using a simple random sample of 15 batteries yielded a mean of 44.9 hours, with a standard deviation of 8.9 hours. Calculate t-value (one-sample t-test) Test this claim using a significance level of 0.05 and P value = 0.026. [5]

