[6328]-54

T.Y.B.Sc.

COMPUTER SCIENCE

CS-354 : Foundations of Data Science

(Revised 2019 Pattern) (Semester-V)

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

Instructions to the candidates:

    1)   *All questions are necessary.*

    2)   *Figures to the right side indicate full marks.*

    3)   *Neat diagrams must be drawn wherever necessary.*

Q1) Attempt any Eight of the following.   **[8×1=8]**

    a)   Define Data Science.

    b)   List any two Social media data sources.

    c)   What is Open Data Source?

    d)   What is outlier?

    e)   Define variance.

    f)   What is nominal attribute?

    g)   Define Data wrangling.

    h)   What is one hot encoding?

    i)   Define data visualization.

    j)   List any two data visualization tools.

Q2) Attempt any four of the following.   **[4×2=8]**

    a)   What do you understand by structured and unstructured data?

    b)   Write any four applications of data science.

    c)   What is inferential statistics?

P.T.O.

d) Define data discretization.

e) What is visual encoding?

Q3) Attempt any two of the following. [2×4=8]

a) Explain any one outlier detection method.

b) Write and explain data visualization libraries in Python.

c) What is data cleaning? Explain any two data cleaning methods.

Q4) Attempt any two of the following [2×4=8]

a) Explain 3 V's of data science with diagram.

b) Elaborate data cube aggregation in detail.

c) Describe Exploratory Data Analysis (EDA) in detail.

Q5) Attempt any one of the following [1×3=3]

a) Find the mean and median for the following list of values.

23, 18, 14, 21, 16, 28

b) Write a short note on feature extraction.

❀❀❀❀