BACHELOR OF TECHNOLOGY (C.B.C.S.) (2021-COURSE)

B. Tech. Sem - III Computer Science & Business Systems : WINTER- 2022

SUBJECT : COMPUTATIONAL STATISTICS

Day : Thursday | Time : 10:00 AM-01:00 PM

Date : 15-12-2022 | W-24145-2022 | Max. Marks : 60

N.B.

1) All questions are **COMPULSORY.**
2) Figures to the **RIGHT** indicate **FULL** marks.
3) Assume suitable data **WHEREVER** necessary.
4) Draw neat labeled diagrams **WHEREVER** necessary.

Q.1 What are the different data structures used in Python? Explain in detail with examples. Also list important functions associated with each data structure. **(10)**

OR

Q.1 Explain in detail the process of data wrangling. Also explain the significance of data wrangling in data analysis. **(10)**

Q.2 Explain the significance of time series in data analysis. Also demonstrate use of important time series functions. **(10)**

OR

Q.2 Demonstrate how to split the data using Groupby function on any column and find out mean, minimum and maximum value for that column. **(10)**

Q.3 Derive formula for Conditional Normal Distribution. Interpret the significance of conditional mean and variance. **(10)**

OR

Q.3 Derive the formula for Multivariate Normal Distribution formula where variables are dependent on each other. **(10)**

Q.4 Discuss the assumptions of Multivariate Regression Model. Also discuss various distance measures used in Multivariate Regression Model. **(10)**

OR

Q.4 Design the model of Multivariate Regression. Also discuss the assumptions and there significance. **(10)**

Q.5 Design the classic equation of Principal Component Analysis $|S-\lambda I|=0$. **(10)**

OR

Q.5 Design model for Factor Analysis. Also list assumptions of Factor Analysis. What is a factor loading matrix? **(10)**

Q.6 Explain clustering as a Process Model. Discuss different distance measures to find proximity in clusters. **(10)**

OR

Q.6 Write a note on k-means clustering demonstrating on an example. Also specify its advantages and disadvantages. **(10)**

*****