Roll No. | | | | | | | | | | |

BCADS First Sessional Tests 2023-24
Fourth Semester
BCADS1402: Data Science

Time: 60 Minutes **Max. Marks: 30**

SECTION-A

Q. No. 1. Attempt *All parts* of the following: **[5*1 = 5]**

a) Name of data analytics methodologies that are commonly used.
b) Explain temporal data and Spatial Data with an example.
c) Explain the "Group by" function.
d) Write down the formula of Binomial and Poisson Distribution.
e) How to read "CSV" in Pandas?

SECTION-B

Q.No.2. Attempt *any Two* of the following: **[7.5*2 = 15]**

a) Discuss the steps involved in the data science lifecycle, highlighting the importance of each step. Provide examples where applicable.
b) Explain the various Probability sampling methods.
c) Define the Central Limit Theorem and briefly explain its significance in statistics and data analysis.
d) Outline the advantages of using NumPy and provide an example illustrating the functionality of NumPy's random module.

SECTION-C

Q. No. 3. Attempt any *One of* the following: **[10*1 = 10]**

a) Discuss the functionality and significance of the subplot() function in Matplotlib. Additionally, provide a code example demonstrating how to create a figure with a (2,2) subplot layout using Matplotlib.

b) A teacher claims that the mean score of students in his class is greater than 82 with a standard deviation of 20. If a sample of 81 students was selected with a mean score of 90 then check if there is enough evidence to support this claim at a 0.05 significance level. Take Critical value 1.65

c) *Scenario*: "Predicting customer churn for a telecommunications company." Using the CRISP-DM framework, outline the steps you would take to address the issue of customer churn.