GANPAT UNIVERSITY
U. V. PATEL COLLEGE OF ENGINEERING
B. TECH- VI (COMPUTER ENGINEERING/INFORMATION TECHNOLOGY)
FIRST INTERNAL EXAMINATION – MARCH 2023
2CEIT6PE4: DATA SCIENCE AND MACHINE LEARNING

Time: 1 Hour   **Total: 20 Marks**

Instructions:
1) Figures to the right indicate full marks.
2) Make necessary assumptions wherever required.
3) Assume suitable data, if required.
4) The text just below marks indicates the Course Outcomes Nos, (CO) followed by Bloom's taxonomy level of the question, i.e., R: Remember, U: Understand, A: Apply, N: Analyse, E: Evaluate, C: Create

Q.1   Answer the following:

[4]

1) The high school principal wants to conduct a survey on student satisfaction for the entire school. You contacted your classmates about their opinion and then presented them to the principal. Would you say this was population or sample data? What is the value you presented called?

2R

2) You are trying to estimate the average valuation of start-ups in the USA. Imagine you are able to visit 200 start-ups in Silicon Valley in a random manner. Identify a possible problem with your study and give a reason for it.

2U

Q.2   _Calculating and analyzing the Body Mass Index values.:_ Assume that you are given the dataset _bmi.csv_, which contains the following headers: _gender, height (in m), and weight (in kg)_. A sample value of the dataset is shown in the following table.

[4]

Gender	Height (in m)	Weight (In kg)
Male	1.74	96
Male	1.89	87
Female	1.85	110
Female	1.95	104
Male	1.49	61
Male	1.89	104

[P.T.O]

Perform the following steps for the above described dataset (Each carries 1 Mark):

1) Write a code to extract the HeightArray and WeightArray as a numpy array of heights and weights from the *bmi.csv* file. (Also mention the statements to import different required libraries) 2A

2) Write a code to calculate the BMIIndex array, which contains the BMI values for each height and weight value in HeightArray and WeightArray, respectively. 2N

   Note: The BMI value can be calculated as follows: w/h^2, where w is the weight in kg, and h is the height in m.

3) Write a code to plot a bar plot and histogram plot for HeightArray, WeightArray, and BMIIndex array. (Also mention the statements to import different required libraries) 2N

4) Now that the BMI is calculated, one can categorize the person based on the index as follows: 2N
   0 - Extremely Weak, 1 - Weak, 2 - Normal, 3 - Overweight, 4 - Obesity,
   5 - Extreme Obesity

Given the above criteria, write a code to count the number of extremely weak persons, weak persons, normal persons, overweight persons, obsessed persons and extremely obsessed persons. Also, write a code to count the number of extremely obsessed males and extremely weak females in the dataset.

Q.3 Classify the Binomial and Poison distribution. On the basis of your differentiation in which case do you think instead of applying a t-test, if we apply a z-test, also be considered true? Justify. [4] 1N

Q.4 The amount spent per month by a segment of credit card users of a bank has a mean value of 12000 and standard deviations of 2000. Calculate the proportion of customers who are spending between 8000 and 16000. [4] 1E

Q.5 Two factories A and B produce heaters for car seats. A customer received a defective car seat heater and the manager at factory B would like to know if it came from her factory. Use the table below to determine the probability that the heater came from factory B. [4] 4A

FACTORY	%OF PRODUCTION	PROBABILITY OF DEFECTIVE ERROR
A	0.55	0.020 P(D|A)
B	0.45	0.014 P(D|B)

--------END OF PAPER-------