

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2024

B.Tech-V Semester (CSE/IT)

COURSE CODE(CREDITS): 20B1WCI531 (2)

MAX. MARKS: 15

COURSE NAME: Foundation for Data Science and Visualization

COURSE INSTRUCTORS: Dr. Ravindara Bhatt, Dr. Rakesh Kanji

MAX. TIME: 1 Hour

1. [CO1] [2 MARKS] List the steps of the data science process.
2. [CO1] [2 MARKS] In real-world data, tuples with missing values for some attributes are a common occurrence. List various methods for handling this problem
3. [CO2] [2 MARKS] Given a pandas DataFrame df containing information about students, extract the names of students who have scored below the mean both in Math and Science.

Name	Math	Science
John	85	92
Mary	78	88
Alex	92	95
Sara	65	75

4. [CO5] [1 MARKS] For the following variables, give the name and indicate the type of the variable (nominal, ordinal, interval, and ratio). a) Education of people as high school, under-graduate, post-graduate b) number of student absent for a data science class.
5. [CO5] [1+ 1 = 2 MARKS] Explain which distribution seems most appropriate for the following phenomenon: Binomial, Normal, Poisson, or Power Law?
 - a) The number of leaves on a fully grown tree.
 - b) The population of cities.
6. [CO5] [2 MARKS] The mileage which car owners get with a certain kind of radial tire is a random variable having exponential distribution with mean 40000 km. Find the probabilities that one of these tires will last atmost 30000 km.
7. Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the following result.[CO5] [2 + 2 = 4 MARKS]

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- (a) Calculate the mean, median of age.
- (b) Draw the boxplot for age.