

TILAK MAHARASHTRA VIDYAPEETH, PUNE
MASTER OF COMPUTER APPLICATIONS
(Specialization in Artificial Intelligence & Machine Learning)
EXAMINATION : JANUARY - 2025
SEMESTER - III

Sub: Exploratory Data Analysis (MCAI 23-302)

Date : 08/01/2025

Total Marks :60

Time: 10.00 am to 12.30 pm

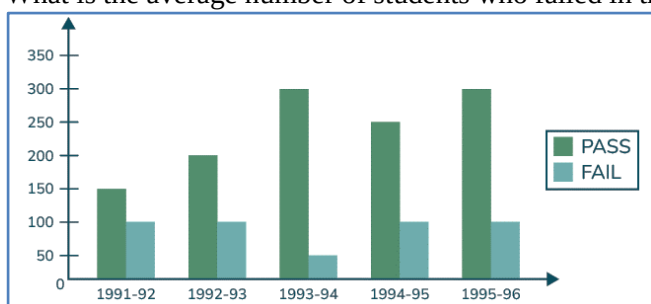
Instruction:

1. All questions are compulsory unless and otherwise stated.
2. Bold figures to the right of every question are the maximum marks for that question.
3. Candidates are advised to attempt questions in order.
4. Answers written illegibly are likely to be marked zero.
5. Use of scientific calculators, Log tables, Mollier Charts is allowed.
6. Draw neat and labelled diagram wherever necessary.

Q.1. Answer the following in 2-3lines. (Any 5)

(10)

1. Define the term standard deviation.
2. Define geometric mean.
3. Observe the bar diagram and answer the question. The first bar out of two adjacent bar states passed students and the next bar states failed students.
 What is the average number of students who failed in the school in the last 5 years?



4. For EDA, what are the sequence and focus of the intermediate steps?
5. What is the purpose of the Histogram in the EDA 4 plots?
6. Explain how EDA can be useful in areas such as Healthcare, Fraud Detection.
7. What do you mean by Kendall's Tau Coefficients?

Q. 2. Answer the following in short. (Any 4)

(20)

1. Write a short note on the role of graphics in EDA.
2. What are the ways to classify the data based on the observation?
3. Differentiate outliers with high-leverage data.
4. Write a short note on ANOVA.
5. Write a short note on the importance of scatter plots in the context of EDA.
6. The calculated uncertainties for commonly used statistics become meaningless if the randomness assumption does not hold. Justify.

Q. 3. Answer the following in detail. (Any 3)

(30)

1. Explain how to deal with non-numerical data in missing places.
2. In a health survey of school children, it is found that the mean hemoglobin level of 55 boys is 10.2 per 100ml with a SD of 2.1. Can we consider this group as taken from a population with a mean of 11.0 g/100ml? Explain the answer.
3. The mean height of the plant population obtained from a random sample of size 100 is 64 inches. The SD of the distribution of height of the plant population is known to be 3 inches. Test the mean height of the population is 67 inches. Solve and explain the example.
4. State and explain four EDA techniques used for routine testing of underlying assumptions.
5. State and explain the interpretation of quartile and percentile values with examples and diagrams.
