# Assignment no.1
# Intelligent data Analysis (IDA)

B Tech(h) Data Science

**Date: 7th October,23** *Date of submission: 17th October 23*

1. Discuss whether each of the following activities is a data mining task, explain your answer:
   I. Dividing the customers of a company according to their gender.
   II. Dividing the customers of a company according to their profitability.
   III. Computing the total sales of a company.
   IV. Sorting a student database based on student identification numbers.
   V. Predicting the outcomes of tossing a (fair) pair of dice.
   VI. Predicting the future stock price of a company using historical records.
   VII. Monitoring the heart rate of a patient for abnormalities.
   VIII. Monitoring seismic waves for earthquake activities.
   IX. Extracting the frequencies of a sound wave.

2. Suppose that you are employed as a data mining consultant for an Internet search engine company. Describe how data mining can help the company by giving specific examples of how techniques, such as clustering, classification, association rule mining,

3. In a survey of 10 households, the number of children was found to be
   4, 1, 5, 4, 3, 7, 2, 3, 4, 1
   (a) State the mode.
   (b) Calculate.
      (i)the mean number of children per household
      (ii)the median number of children per household.
   (c) A researcher says: "The mode seems to be the best average to represent the data in this survey." Give ONE reason to support this statement.

4. Find the mean, median, mode and range of each set of numbers below.
   (a) 3, 4, 7, 3, 5, 2, 6, 10
   (b) 8, 10, 12, 14, 7, 16, 5, 7, 9, 11
   (c) 17, 18, 16, 17, 17, 14, 22, 15, 16, 17, 14, 12
   (d) 108, 99, 112, 111, 108
   (e) 64, 66, 65, 61, 67, 61, 57
   (f) 21, 30, 22, 16, 24, 28, 16, 17

5. A football team keep records of the number of goals it scores per match during a season. The list is shown opposite. Find the mean number of goals per match.

| No. of Goals | Frequency |
|---|---|
| 0 | 8 |
| 1 | 10 |
| 2 | 12 |
| 3 | 3 |
| 4 | 5 |
| 5 | 2 |

6. Explain why data pre-processing is important for data mining?
7. Explain different visualization technique with proper example (take sample data and visualize them in the system and stick the printout in your assignment book).
8. Explain data objects and attributes. Write down the matrix of different attributes and Transformations that represents attribute level.
9. Explain sampling with example.
10. Explain the curse of dimensionality.
11. What features creation? what is the need of feature creation?
12. Explain different types of discretization?
13. What is the reason for variable transformation?
14. Explain different types of data transformation?
15. For the following group of data:

    200, 400, 800, 1000, 2000, 2200

    normalize them with min=0 and max=100.