DWM Module 2 Question Bank

Q. 1) Suppose a group of sales price records has been sorted as follows

6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232

Partition them into 3- bins by equal frequency partitioning method. Perform data smoothing by
a) Bin mean (b) Bin median (c) Bin boundaries.

Q. 2) For the given attribute Price values in dollars :

16 , 16 , 180 , 4 , 12 , 24 , 26 , 28

Apply following Binning technique for smoothing the noise
a) Bin mean (b) Bin median (c) Bin boundaries.

Q. 3) For the given set of data points :

11 , 13 , 13, 16 , 20, 20, 20, 30, 40, 45 ,75

a) Find mean, median and mode
b) Show a box plot of the data , clearly indicating 5 point summary

Q. 4) Consider the following data points :

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30,33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

(a) What is the *mean* of the data? What is the *median*?
(b) What is the *mode* of the data? Comment on the data's modality (i.e.,unimodal, bimodal, trimodal, etc.).
(c) What is the *midrange* of the data?
(d) Can you find (roughly) the first quartile (*Q*1) and the third quartile (*Q*3) of the data?
(e) Give the *five-number summary* of the data.
(f) Show a *boxplot* of the data.

Q.5) For the given attribute marks values : **35, 45, 50, 55, 60, 65, 75**
Compute mean*, median, mode*. Also compute the *five-number summary* of above data.

Q. 6) In real world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.

Q.7) Write short notes on Data Pre-processing.

Q.8) Explain major issues in Data Mining and what are applications of data mining.

Q. 9) Explain Data-Reduction technique.

Q.10) Describe the steps involved in Data Mining when viewed as a process of knowledge discovery.