Tutorial 4

(Data Science)

-----------------------------------------------------------------------------------------------------------------------------------------

1. Suppose that the data for analysis include the attribute age. The age values for the data tuples
2. are (in increasing order): 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.
3. Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.
4. How might you determine outliers in the data?
5. What other methods are there for data smoothing?
6. Use the two methods below to normalize the following group of data:

200, 300, 400, 600, 1000.

1. min-max normalization by setting min = 0 and max = 1
2. z-score normalization
3. Suppose a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, and 215. Partition them into three bins by each of the following methods.
4. equal-frequency (equidepth) partitioning
5. equal-width partitioning
6. clustering