CS 345: Data Mining Names:_____

In-Class Activity 2 _____

_____

_____

1. Use the following data (in increasing order) for the attribute age to answer the following questions: 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. 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.

b. How might you determine outliers in the data?

c. What other methods are there for data smoothing?

2. Using the data for age given in Question 1, answer the following:
    a. Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0].

    b. Use *z*-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.

    c. Use normalization by decimal scaling to transform the value 35 for age.

3. Using the data for age given in Question 1, plot an equal-width histogram of width 10.

4. Discuss issues to consider during data integration.