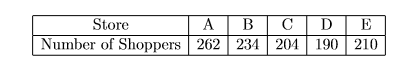
1. For the following group of data: 200, 400, 800, 1000, 2000, 2200, normalize them with min = 0 and max = 100.

2. For the above group of data, partition them into two bins by each of the following methods: (1) equal-width partitioning, (2) equal-frequency partitioning.

3. What are the pre-processing method required for the below data set and how you will do?

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Age | Salary | Purchased |
| France | 44 | 72000 | No |
| Spain | 27 | 48000 | Yes |
| Germany | 30 | 54000 | No |
| Spain | 38 | 61000 | No |
| Germany | 40 |  | Yes |
| France | 35 | 58000 | Yes |
| Spain |  | 52000 | No |
| France | 48 | 79000 | Yes |
| Germany | 50 | 83000 | No |
| France | 37 | 67000 | Yes |

4. A department store, A, has four competitors: B,C,D, and E. Store A hires a consultant to determine if the percentage of shoppers who prefer each of the ﬁve stores is the same. A survey of 1100 randomly selected shoppers is conducted, and the results about which one of the stores shoppers prefer are below. Is there enough evidence using a signiﬁcance level α = 0.05 to conclude that the proportions are really the same?



5. Discuss (shortly) whether or not each of the following activities is a data mining task.

(a) Dividing the customers of a company according to their profitability.

(b) Predicting the outcomes of tossing a (fair) pair of dice.