Answer the following questions based on the data sets shown in Figure 1. Note that each data set contains 1000 items and 10000 transactions. Dark cells indicate the presence of items and white cells indicate the absence of items. We will apply the Apriori algorithm to extract frequent item sets with *minsup* = 10% (i.e, itemsets must be contain at least 1000 transactions).

(a). Assume that the minimum support threshold is equal to 10%. How many closed frequent itemsets will be discovered from data set 1?

(b). Which data set will produce frequent itemsets with high support?

(c). Which data set will produce the most number of closed frequent itemsets?

Items Items



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | 2000 |  |  |  | 2000 |
|  |  |  | 4000 |  |  |  | 4000 |
| Transactions |  |  |  | Transactions |  |  |  |
|  |  |  | 6000 |  |  |  | 6000 |
|  |  |  | 8000 |  |  |  | 8000 |
| 200 | 400 | 600 | 800 | 200 | 400 | 600 | 800 |

(a) Synthetic data set 1. (b) Synthetic data set 2.

Figure 1: Data sets