Data Mining Assignment

1. Given Minimum Support Threshold = 25%

Given Minimum Confidence Threshold = 50%

Support = (2000/5000) \*100 = 40%

Confidence (2000/3000) \*100 = 66.66%

Strong association rules satisfy both minimum support and minimum confidence. Hence this association rule is strong.

1. In the table given, children and homeowners have an association with them. Also, we can determine with the zip codes that people with similar zip codes live close by and are neighbors. But here since it’s a group of customers belonging to a particular area according to the zip code, we shall use clustering instead of association rules.
2. 1) In the above table, Eyebrow Pencil was purchased only in transaction #3.

The bag was purchased only in the 10th Transaction.

In Transaction 1, Bag and eyebrow pencils were the only items that were not purchased.

In Transaction 2, Nailpolish concealer and Bronzer were purchased.

2) The first row has Rule#2, the confidence is 60.19%,

Therefore by interpreting this value we can say that 60.19% times we say that

Brushes and Concealers appear along with Bronzer and Nail polish whenever

Bronzer and Nail polish is purchased.

Conf(A -> B) = Support(A, B)/Support(A)

For First row,

Conf% = (62/103)\*100

3)This rule is about the purchase of Brushes and concealer with Bronzer and Nail polish. The confidence % is 60.19 and lift ratio is 3.909. Hence we can say that while purchasing bronzer and nail polish, brushes and concealer are also purchased.

4) I would choose rule 1 as it has a High lift ratio of 3.909 and high confidence support of 80.52%