I used Python, specifically the Pandas and mlxtend libraries. I checked some of the rules by hand afterwards and they seem accurate. For example, the first rule was Coals -> Burgers, 0.5, 1. i.e. Coals were bought 5 out of the 10 times, and burgers were always bought with them.

These are the steps I took:

1. Import libraries and CSV dataset.
2. Convert CSV into 7 columns, TID & one column for each item., Milk, Bun etc.
3. Turn each transaction into a binary representation of the items bought.
4. Produce list of itemsets with support > 0.25.
5. Derive rules with confidence >= 0.5 from the itemsets.
6. Select only rules where Coals was in the antecedents, as we only care about associations when Coals are bought.

Code:

Text

Description automatically generated

Output 43 rules (sorted by confidence):

Table

Description automatically generated

Questions:

1. I filtered the rules that were made so that I only included rules where Coals was in the antecedent. Is this the correct way to do it?
2. I tried to filter the itemsets before generating the rules to make the process less computationally effective but I kept getting errors. Is it advisable/possible to do this?
3. Did anybody else get 43 rules?