

Association Rule Mining

Haden Loveridge, Alex Parson, Biagio Alessandrello

Loading txt File for Analysis:

The ‘arules’ package we used expects data in the form of ‘transactions’ to do this, we used ‘read.transactions’ from the arules package. This differs from the example in the playlists.R file where each record was a single pair of artists and we had to then create baskets.

Rules Parameters:

For our rules we chose to have the following parameters:

- **Support = 0.001**
 - We chose a level this low compared to the default because we were open to looking at very niche transactions that only occur in as few as 0.1% of all transactions.
- **Confidence = 0.05**
 - Again, we lowered the confidence compared to the default that allowed the subsequent rules to be true only 5% of the time.
- **maxlen = 6**
 - We did limit the max length of a rule because when there was no limit we saw few interesting results and few large lifts. In a shopping list, few recipes, events, or late night cravings would include groups of items greater than 6 unless they were independent of one another.

Highest & Lowest Lift:

From the two tables bellow we can see the top 5 and lowest 5 ‘Lift’ from our rules. In other words, the rule combination of ‘If a person buys 1) Bottled Beer, and 2) red/blush Win, then they are 35.7 times more likely to also purchase Liquor. That would make sense since if a person is buying both beer and wine they are buying liquor for parties and other celebrations.

On the low end, of the 5 lowest lifts, 4 of them also included Bottled Beer. When a person buys Bottled beer they are less than half as likely to also purchase yogurt, veggies, or whole milk. This scenario also makes sense if you consider that if you are buying whole milk and yogurt you are likely shopping for toddlers (who likely arent drinking a Miller Lite).

Table 1: Top 5 Rules by Lift

Antecedent	Consequent	Support	Confidence	Coverage	Lift	Count
{canned beer}	{yogurt}	0.0053889	0.0693717	0.0776817	0.4972820	53
{canned beer}	{root vegetables}	0.0040671	0.0523560	0.0776817	0.4803372	40
{UHT-milk}	{whole milk}	0.0039654	0.1185410	0.0334520	0.4639280	39

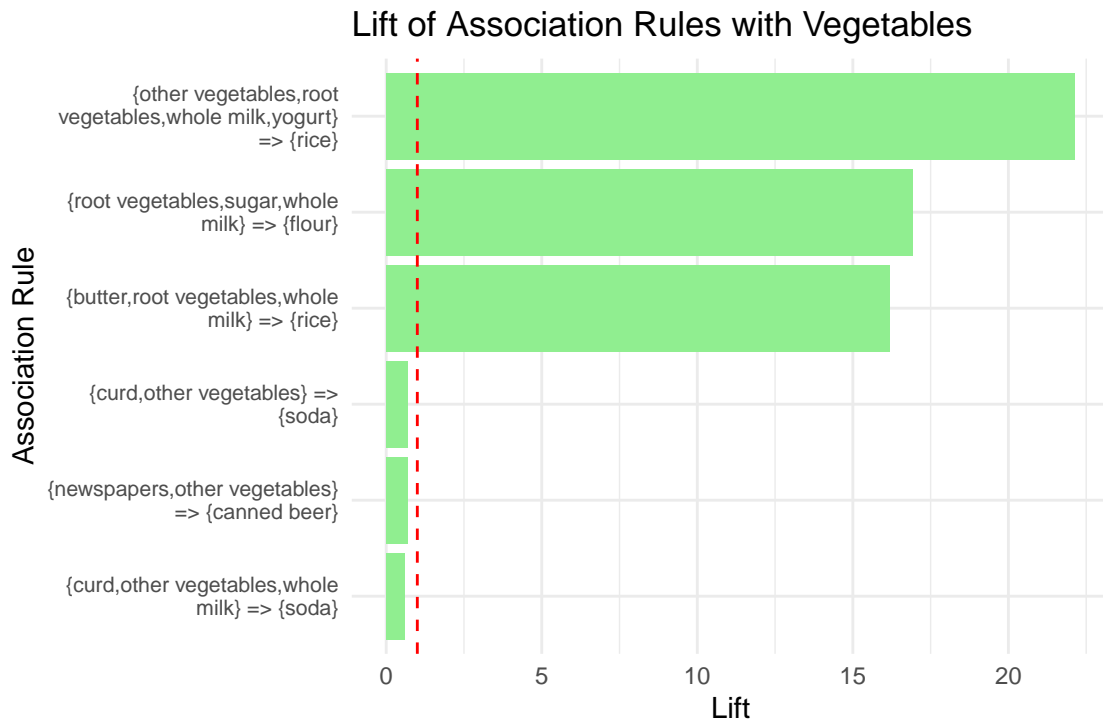
{canned beer}	{whole milk}	0.0088460	0.1138743	0.0776817	0.4456642	87
{canned beer,shopping bags}	{whole milk}	0.0012201	0.1071429	0.0113879	0.4193195	12

Table 2: Bottom 5 Rules by Lift

Antecedent	Consequent	Support	Confidence	Coverage	Lift	Count
{canned beer}	{yogurt}	0.0053889	0.0693717	0.0776817	0.4972820	53
{canned beer}	{root vegetables}	0.0040671	0.0523560	0.0776817	0.4803372	40
{UHT-milk}	{whole milk}	0.0039654	0.1185410	0.0334520	0.4639280	39
{canned beer}	{whole milk}	0.0088460	0.1138743	0.0776817	0.4456642	87
{canned beer,shopping bags}	{whole milk}	0.0012201	0.1071429	0.0113879	0.4193195	12

Visual of ‘Vegetable Rules’

The graph below shows the 3 highest and 3 lowest lift metrics for any rule that has vegetables as the antecedent.



Visual of Rules:

The final plot shows the relation ship of ‘Lift’ on the Y axis (increased/decreased likelihood of purchase given another item), ‘Support’ on the X axis (% of transactions containing item), and

‘Confidence’ as the darkness of the data point (ratio of both items compared to a given item).

Scatter plot for 37965 rules

