UNIVERSIDAD DE GRANADA

PRÁCTICA 3

REGLAS DE ASOCIACIÓN CON KNIME

TRATAMIENTO INTELIGENTE DE DATOS

MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA

AUTOR

Pablo Valenzuela Álvarez (pvalenzuela@correo.ugr.es)

TABLA DE CONTENIDO

EJERCICIO 1. MARKET

Leemos la base de datos usando el nodo *CSV Reader* y obtenemos los resultados de la foto 1. Podemos ver que contiene valores 0 o 1 para las columnas correspondientes a un producto, también podemos deducir se corresponden a un ticket de compra y si el valor es 1, ese producto va en ese ticket.

ID	Bread	Honey	Bacon	Toothpaste	Banana	Apple	Hazelnut	Cheese	Meat	Carrot	Cucumber	Onion	Milk	Butter	ShavingFoam	Salt	Flour	HeavyCream	Egg	Olive	Shampoo	Sugar
Row0	1	0	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1
Row1	1	1	1	0	1	1	1	0	0	0	1	0	1	1	0	0	1	0	0	1	1	0
Row2	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	1
Row3	1	1	0	1	0	1	0	0	0	0	1	1	1	0	0	0	1	0	1	1	1	0
Row4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Row5	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
Row6	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0
Row7	0	0	1	1	1	0	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0
Row8	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0	1	1	1	0	0
Row9	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	0	1	0
Row10	0	0	1	1	0	1	1	0	0	1	0	0	0	1	0	0	1	1	1	1	0	0
Row11	0	1	1	1	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	0	0	1
Row12	1	1	1	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	1	1	1	1
Row13	1	0	1	1	1	0	1	1	0	1	0	0	0	0	1	1	0	1	1	1	0	1

Foto 1. Datos de market.csv.

En términos de base de datos, tiene una forma relacional (donde cada valor tiene un 0 o 1). Podemos aplicar diferentes algoritmos de extracción de reglas como **Apriori** y ver los resultados (No podemos aplicar **FPGrowth** porque necesita que los datos sean transaccionales).

```
1. Hazelnut=0 Carrot=0 Onion=0 Flour=0 80 ==> Toothpaste=0 75    <conf:(0.94)> lift:(1.52) lev:(0.06) [25] conv:(5.11)
2. Hazelnut=0 Onion=0 Flour=0 Shampoo=0 79 ==> Toothpaste=0 74    <conf:(0.94)> lift:(1.52) lev:(0.05) [25] conv:(5.05)
3. Bread=0 Cheese=0 Egg=0 Shampoo=0 Sugar=0 76 ==> Meat=0 71    <conf:(0.93)> lift:(1.53) lev:(0.05) [24] conv:(4.91)
4. Hazelnut=0 Onion=0 Flour=0 HeavyCream=0 80 ==> Toothpaste=0 74    <conf:(0.93)> lift:(1.5) lev:(0.05) [24] conv:(4.38)
5. Apple=0 Hazelnut=0 Onion=0 Flour=0 77 ==> Toothpaste=0 71    <conf:(0.92)> lift:(1.5) lev:(0.05) [23] conv:(4.22)
```

Foto 2. Resultados Apriori.

Como podemos ver, las reglas extraídas con mayor confianza son las de 0, es decir cuando no compra. Nos interesa justo lo contrario, por lo que hay que transformar los datos de relacional a transaccional. Para ello usamos los nodos de la foto 3 y obtenemos los resultados de la foto 4.

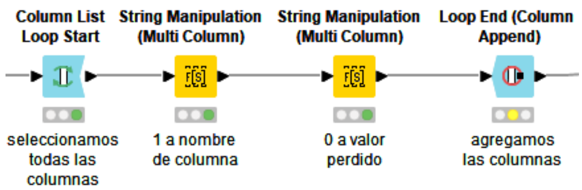

Foto3. Loop

ID	Bread	Honey	Bacon	Toothpaste	Banana	Apple	Hazelnut	Cheese	Meat	Carrot	Cucumber	Onion	Milk	Butter	ShavingFoam	Salt	Flour	HeavyCream	Egg	Olive	Shampoo	Sugar
Row0	Bread	?	Bacon	?	Banana	Apple	Hazelnut	?	?	Carrot	?	?	?	?	?	?	?	HeavyCream	Egg	?	?	Sugar
Row1	Bread	Honey	Bacon	?	Banana	Apple	Hazelnut	?	?	?	Cucumber	?	Milk	Butter	?	?	Flour	?	?	Olive	Shampoo	?
Row2	?	Honey	Bacon	Toothpaste	Banana	Apple	Hazelnut	Cheese	Meat	?	Cucumber	Onion	Milk	?	ShavingFoam	Salt	Flour	HeavyCream	Egg	?	?	Sugar
Row3	Bread	Honey	?	Toothpaste	?	Apple	?	?	?	?	Cucumber	Onion	Milk	?	?	?	Flour	?	Egg	Olive	Shampoo	?
Row4	?	Honey	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row5	?	Honey	?	Toothpaste	?	?	Hazelnut	?	?	?	?	Onion	?	?	ShavingFoam	?	?	?	?	?	?	Sugar
Row6	?	?	Bacon	?	Banana	Apple	?	?	?	Carrot	?	?	?	?	?	Salt	?	HeavyCream	?	?	?	?
Row7	?	?	Bacon	Toothpaste	Banana	?	Hazelnut	?	?	?	Cucumber	Onion	?	?	?	Salt	?	?	Egg	?	?	?
Row8	?	Honey	Bacon	?	Banana	Apple	Hazelnut	Cheese	Meat	Carrot	?	Onion	?	Butter	ShavingFoam	?	?	HeavyCream	Egg	Olive	?	?
Row9	?	?	?	?	?	?	?	?	?	?	?	Onion	Milk	?	?	Salt	Flour	HeavyCream	?	?	Shampoo	?
Row10	?	?	Bacon	Toothpaste	?	Apple	Hazelnut	?	?	Carrot	?	?	?	Butter	?	?	Flour	HeavyCream	Egg	Olive	?	?
Row11	?	Honey	Bacon	Toothpaste	?	?	?	?	Meat	?	?	Onion	?	Butter	?	?	Flour	HeavyCream	?	?	?	Sugar
Row12	Bread	Honey	Bacon	Toothpaste	Banana	Apple	?	?	Meat	Carrot	?	Onion	Milk	?	ShavingFoam	?	?	?	Egg	Olive	Shampoo	Sugar
Row13	Bread	?	Bacon	Toothpaste	Banana	?	Hazelnut	Cheese	?	Carrot	?	?	?	?	ShavingFoam	Salt	?	HeavyCream	Egg	Olive	?	Sugar
Row14	?	?	?	Toothpaste	Banana	?	?	?	?	Carrot	Cucumber	?	?	?	ShavingFoam	?	Flour	?	Egg	Olive	Shampoo	Sugar
Row15	?	?	Bacon	?	?	?	?	?	?	?	?	?	Milk	?	?	?	Flour	?	?	?	Shampoo	?
Row16	?	?	?	Toothpaste	?	?	?	?	?	?	Cucumber	?	?	?	?	?	Flour	?	?	?	?	?
Row17	?	?	?	?	?	?	?	?	?	?	?	Onion	?	?	ShavingFoam	Salt	?	?	?	Olive	?	?

Foto 4. Datos transformados

Aplicando los nodos *Create Collection Column* combinamos todas las celdas en una, obteniendo finalmente los datos de la foto 5.

ID	itemSets
Row0	[Bread, Bacon, Banana, Apple, Hazelnut, Carrot, HeavyCream, Egg, Sugar]
Row1	[Bread, Honey, Bacon, Banana, Apple, Hazelnut, Cucumber, Milk, Butter, Flour, Olive, Shampoo]
Row2	[Honey, Bacon, Toothpaste, Banana, Apple, Hazelnut, Cheese, Meat, Cucumber, Onion, Milk, ShavingFoam, Salt, Flour, HeavyCream, Egg, Sugar]
Row3	[Bread, Honey, Toothpaste, Apple, Cucumber, Onion, Milk, Flour, Egg, Olive, Shampoo]
Row4	[Honey]
Row5	[Honey, Toothpaste, Hazelnut, Onion, ShavingFoam, Sugar]
Row6	[Bacon, Banana, Apple, Carrot, Salt, HeavyCream]
Row7	[Bacon, Toothpaste, Banana, Hazelnut, Cucumber, Onion, Salt, Egg]
Row8	[Honey, Bacon, Banana, Apple, Hazelnut, Cheese, Meat, Carrot, Onion, Butter, ShavingFoam, HeavyCream, Egg, Olive]
Row9	[Onion, Milk, Salt, Flour, HeavyCream, Shampoo]
Row10	[Bacon, Toothpaste, Apple, Hazelnut, Carrot, Butter, Flour, HeavyCream, Egg, Olive]
Row11	[Honey, Bacon, Toothpaste, Meat, Onion, Butter, Flour, HeavyCream, Sugar]
Row12	[Bread, Honey, Bacon, Toothpaste, Banana, Apple, Meat, Carrot, Onion, Milk, ShavingFoam, Egg, Olive, Shampoo, Sugar]
Row13	[Bread, Bacon, Toothpaste, Banana, Hazelnut, Cheese, Carrot, ShavingFoam, Salt, HeavyCream, Egg, Olive, Sugar]
Row14	[Toothpaste, Banana, Carrot, Cucumber, ShavingFoam, Flour, Egg, Olive, Shampoo, Sugar]
Row15	[Bacon, Milk, Flour, Shampoo]
Row16	[Toothpaste, Cucumber, Flour]

Foto 5. Datos finales.

Para extraer los item sets más frecuentes podemos usar el nodo *Item Set Finder (Borgelt),* configurandolo para que use el algoritmo **Apriori,** un mínimo de 5 items y un soporte mínimo de 5, tenemos los resultados de la foto 6.

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row4	[Butter,ShavingFoam,Bacon,...	5	28	6.035
Row3	[Butter,ShavingFoam,Hazeln...	5	27	5.819
Row11	[Meat,Egg,Carrot,...]	5	26	5.603
Row0	[Butter,Onion,ShavingFoam,...	5	25	5.388
Row2	[Butter,Meat,ShavingFoam,...]	5	25	5.388
Row6	[Butter,Egg,Bacon,...]	5	25	5.388
Row7	[Onion,Meat,Carrot,...]	5	25	5.388
Row9	[Onion,ShavingFoam,Bacon,...]	5	25	5.388
Row10	[Meat,ShavingFoam,Honey,...]	5	25	5.388
Row13	[Meat,Carrot,Honey,...]	5	25	5.388
Row14	[ShavingFoam,Hazelnut,Baco...	5	25	5.388
Row1	[Butter,Onion,Bacon,...]	5	24	5.172
Row5	[Butter,Egg,Hazelnut,...]	5	24	5.172
Row8	[Onion,Meat,Honey,...]	5	24	5.172
Row12	[Meat,Egg,Honey,...]	5	24	5.172

Foto 6. ItemSets frecuentes.

Estos resultados significan para:

- ItemSetSupport: Frecuencia en la que nos encontramos el ItemSet en las demás filas o transacciones.
- RelativeItemSetSupport: Se refiere al soporte de un conjunto de ítems en relación con el soporte total de un subconjunto más grande del que forma parte.

Por último, necesitamos extraer reglas de estas transacciones. Para ello usando el nodo *Association Rule Learner* y la configuración de la foto 7, obtenemos los resultados de la foto 8.

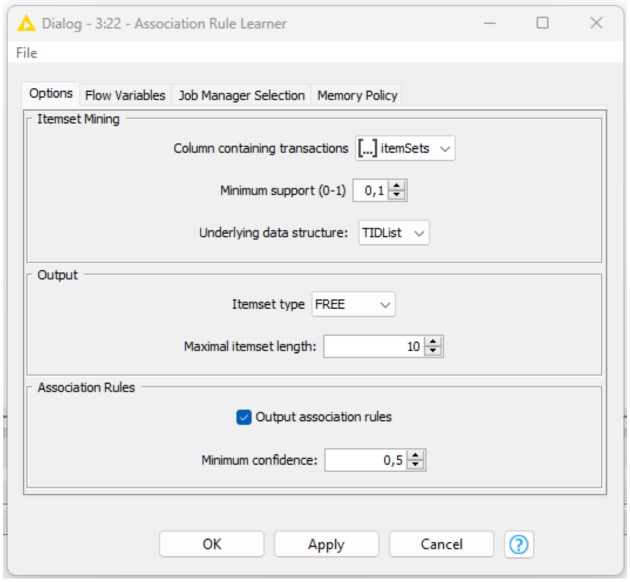

Foto 7. Configuración del nodo Association Rule Learner.

Row ID	D Support	D Confi...	D Lift	S Consequent	S implies	[...] Items
rule135	0.131	0.701	1.564	Banana	<---	[Onion,Bacon]
rule454	0.119	0.688	1.534	Banana	<---	[Olive,ShavingFoam]
rule122	0.121	0.675	1.565	Bacon	<---	[Butter,Banana]
rule264	0.127	0.67	1.555	Bacon	<---	[Butter,ShavingFoam]
rule259	0.138	0.667	1.502	Cheese	<---	[Butter,Bacon]
rule660	0.106	0.653	1.472	Cheese	<---	[Carrot,Butter]
rule729	0.106	0.653	1.472	Cheese	<---	[Egg,Butter]
rule26	0.121	0.651	1.511	Bacon	<---	[ShavingFoam,Bread]
rule794	0.116	0.651	1.465	Cheese	<---	[Onion,Butter]
rule221	0.131	0.649	1.462	Cheese	<---	[Egg,Bacon]
rule792	0.106	0.645	1.452	Cheese	<---	[Butter,Meat]
rule100	0.121	0.644	1.436	Banana	<---	[Apple,Bacon]
rule404	0.112	0.642	1.432	Banana	<---	[Egg,Onion]
rule520	0.108	0.641	1.525	Hazelnut	<---	[HeavyCream,Egg]
rule653	0.108	0.641	1.541	Honey	<---	[Carrot,Meat]
rule216	0.103	0.64	1.485	Bacon	<---	[Egg,Butter]
rule257	0.138	0.64	1.485	Bacon	<---	[Butter,Cheese]
rule359	0.119	0.64	1.427	Banana	<---	[Olive,Carrot]
rule445	0.114	0.639	1.438	Cheese	<---	[Butter,Banana]
rule429	0.11	0.637	1.422	Banana	<---	[Olive,Cucumber]
rule797	0.121	0.636	1.433	Cheese	<---	[Butter,ShavingFoam]

Foto 8. Reglas de asociación obtenidas.

Las reglas obtenidas que muestro son bastante normales, la mayoría implican que si comprar comida o vegetales también compras otro tipo de comida, fruta o vegetal. Por poner una excepción, podemos observa la regla 26 que asocia la compra de **pan** y **espuma** de afeitar con la de **bacon**.

EJERCICIO 2. CESTA DE LA COMPRA

Realizamos los pasos dictados en el documento de prácticas para leer y transformar los datos para que sean transacciones y podamos extraer item sets frecuentes y reglas.

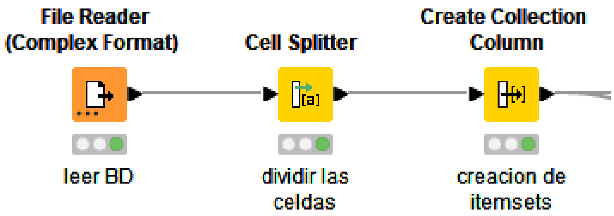

Foto 9. Pasos indicados en documento.

ID	ItemSets
Row0	[shrimp, almonds, avocado, vegetables mix, green grapes, whole weat flour, yams, cottage cheese, energy drink, tomato juice, low fat yogurt, green te
Row1	[burgers, meatballs, eggs]
Row2	[chutney]
Row3	[turkey, avocado]
Row4	[mineral water, milk, energy bar, whole wheat rice, green tea]
Row5	[low fat yogurt]
Row6	[whole wheat pasta, french fries]
Row7	[soup, light cream, shallot]
Row8	[frozen vegetables, spaghetti, green tea]
Row9	[french fries]
Row10	[eggs, pet food]
Row11	[cookies]
Row12	[turkey, burgers, mineral water, eggs, cooking oil]
Row13	[spaghetti, champagne, cookies]
Row14	[mineral water, salmon]
Row15	[mineral water]
Row16	[shrimp, chocolate, chicken, honey, oil, cooking oil, low fat yogurt]
Row17	[turkey, eggs]
Row18	[turkey, fresh tuna, tomatoes, spaghetti, mineral water, black tea, salmon, eggs, chicken, extra dark chocolate]

Foto 10. Transacciones obtenidas

Ahora usaremos los mismos nodos que en el ejercicio anterior para extraer item sets y reglas de asociación.

Para el primer caso, optaremos por buscar un mínimo de 5 items que suelan aparecer muchos en los tickets (Foto 11).

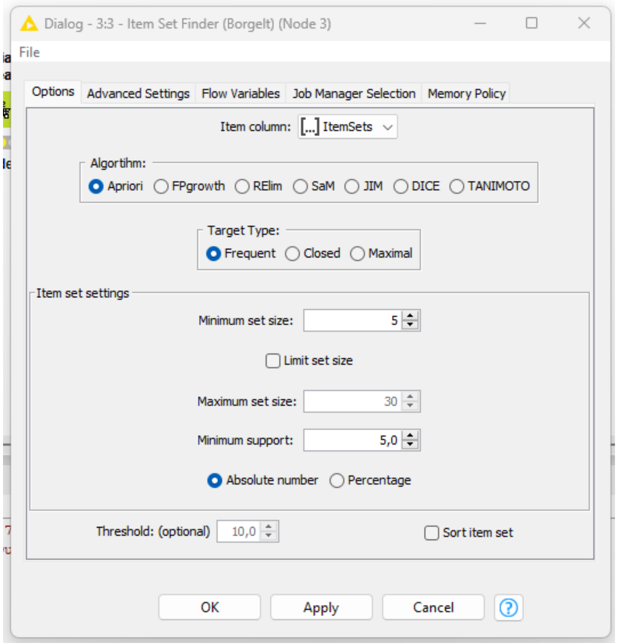

Foto 11. Configuración item sets.

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row483	[frozen vegetables,ground beef,milk,...]	5	15	0.2
Row503	[frozen vegetables,milk,chocolate,...]	5	15	0.2
Row486	[frozen vegetables,ground beef,milk,...]	5	14	0.187
Row393	[shrimp,frozen vegetables,ground beef,...]	5	13	0.173
Row402	[shrimp,frozen vegetables,chocolate,...]	5	13	0.173
Row491	[frozen vegetables,ground beef,chocolate,...]	5	13	0.173
Row504	[frozen vegetables,milk,eggs,...]	5	13	0.173
Row516	[ground beef,chocolate,eggs,...]	5	13	0.173
Row481	[frozen vegetables,ground beef,milk,...]	5	12	0.16
Row297	[tomatoes,frozen vegetables,milk,...]	5	11	0.147
Row348	[olive oil,frozen vegetables,milk,...]	5	11	0.147
Row353	[olive oil,frozen vegetables,chocolate,...]	5	11	0.147
Row400	[shrimp,frozen vegetables,milk,...]	5	11	0.147
Row485	[frozen vegetables,ground beef,milk,...]	5	11	0.147
Row490	[frozen vegetables,ground beef,chocolate,...]	5	11	0.147
Row334	[olive oil,pancakes,chocolate,...]	5	10	0.133
Row399	[shrimp,frozen vegetables,milk,...]	5	10	0.133
Row502	[frozen vegetables,milk,chocolate,...]	5	10	0.133
Row75	[herb & pepper,frozen vegetables,ground beef,...]	5	9	0.12
Row112	[soup,olive oil,frozen vegetables,...]	5	9	0.12

Foto 12. Resultados obtenidos.

Vistos los resultados de la foto 12, podemos sugerir al supermercado colocar los ítems correspondientes a los item sets en lugares cercanos.

Las reglas extraídas con la configuración de la foto 13, obtenemos los resultados de la foto 14.

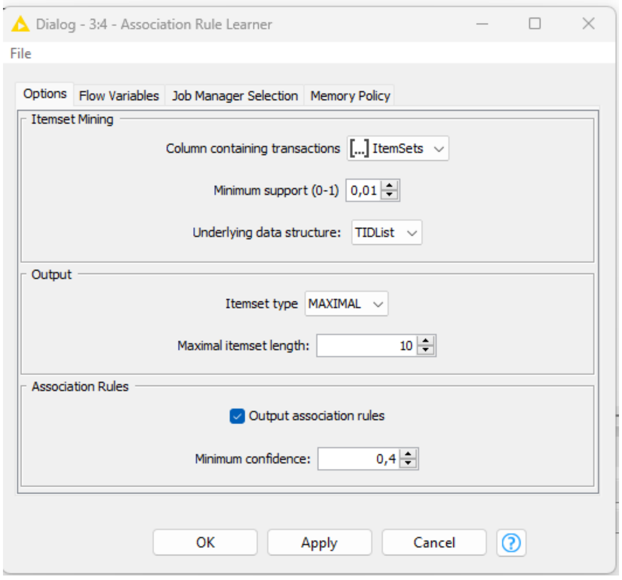

Foto 13. Configuración de reglas de asociación.

ID	Support	Confidence	Lift	Consequent	implies	Items
rule0	0.017064391414478068	0.4012539184952978	1.6833364891684723	mineral water	<---	[salmon]
rule1	0.010265297960271964	0.4476744186046512	1.8780793142916603	mineral water	<---	[spaghetti, olive oil]
rule2	0.027596320490601255	0.4190283400809717	1.757903567643942	mineral water	<---	[olive oil]
rule3	0.013064924676709772	0.42424242424242425	1.7797776421937497	mineral water	<---	[eggs, milk]
rule4	0.013464871350486601	0.40562248995983935	1.7016634771749188	mineral water	<---	[eggs, chocolate]
rule5	0.010131982402346354	0.5066666666666667	2.125563012677107	mineral water	<---	[eggs, ground beef]
rule6	0.011065191307825623	0.46892655367231634	1.967236062134253	mineral water	<---	[frozen vegetables, milk]
rule7	0.01573123583522197	0.44360902255639095	1.8610242048073204	mineral water	<---	[spaghetti, milk]
rule8	0.013998133582189041	0.43568464730290457	1.827779943746693	mineral water	<---	[milk, chocolate]
rule9	0.011065191307825623	0.503030303030303	2.110307775744017	mineral water	<---	[milk, ground beef]
rule10	0.023063591521130515	0.45646437994722955	1.9149548735929356	mineral water	<---	[soup]
rule11	0.011998400213304892	0.43062200956937796	1.8065412157605727	mineral water	<---	[spaghetti, frozen vegetables]
rule12	0.01586455139314758	0.40476190476190477	1.6980531586236283	mineral water	<---	[spaghetti, chocolate]
rule13	0.011465137981602452	0.455026455026455	1.9089225051193728	mineral water	<---	[spaghetti, pancakes]
rule14	0.017064391414478068	0.4353741496598639	1.8264773470909612	mineral water	<---	[spaghetti, ground beef]
rule15	0.017064391414478068	0.41693811074918563	2.3946805273580716	spaghetti	<---	[mineral water, ground beef]
rule16	0.010931875749900012	0.47398843930635837	1.988471634920019	mineral water	<---	[ground beef, chocolate]
rule17	0.040927876283162246	0.41655359565807326	1.7475215442008991	mineral water	<---	[ground beef]

Foto 14. Resultados obtenidos.

Analizándolos podemos observar que la gente, al hacer cualquier compra se lleva también agua mineral. Por lo que podemos aconsejar al supermercado mantener una reserva de este item alta y colocarlo de forma visible para el cliente.

Si bajamos la medida de confianza podemos obtener reglas como las 62, 17 y 90 en la foto 15, las cuales asocian la compra de agua, vino, aceite o carne con spaguettis. Algunos de estos ítems se pueden mezclar a la hora de elaborar ciertas comidas por lo que tiene sentido que se compren juntos.

Row ID	D Support	D Con...	D Lift	S Conseq...	S implies	[...] Items
rule29	0.01	0.507	2.126	mineral water	<---	[eggs,ground beef]
rule43	0.011	0.503	2.11	mineral water	<---	[milk,ground beef]
rule67	0.011	0.474	1.988	mineral water	<---	[ground beef,chocolate]
rule34	0.011	0.469	1.967	mineral water	<---	[frozen vegetables,milk]
rule51	0.023	0.456	1.915	mineral water	<---	[soup]
rule59	0.011	0.455	1.909	mineral water	<---	[spaghetti,pancakes]
rule16	0.01	0.448	1.878	mineral water	<---	[spaghetti,olive oil]
rule37	0.016	0.444	1.861	mineral water	<---	[spaghetti,milk]
rule40	0.014	0.436	1.828	mineral water	<---	[milk,chocolate]
rule61	0.017	0.435	1.826	mineral water	<---	[spaghetti,ground beef]
rule52	0.012	0.431	1.807	mineral water	<---	[spaghetti,frozen vegetables]
rule20	0.013	0.424	1.78	mineral water	<---	[eggs,milk]
rule18	0.028	0.419	1.758	mineral water	<---	[olive oil]
rule62	0.017	0.417	2.395	spaghetti	<---	[mineral water,ground beef]
rule78	0.041	0.417	1.748	mineral water	<---	[ground beef]
rule26	0.013	0.406	1.702	mineral water	<---	[eggs,chocolate]
rule56	0.016	0.405	1.698	mineral water	<---	[spaghetti,chocolate]
rule14	0.017	0.401	1.683	mineral water	<---	[salmon]
rule82	0.01	0.399	1.674	mineral water	<---	[cereals]
rule148	0.039	0.399	2.291	spaghetti	<---	[ground beef]
rule66	0.02	0.394	1.654	mineral water	<---	[cooking oil]
rule23	0.014	0.391	1.638	mineral water	<---	[eggs,spaghetti]
rule74	0.011	0.389	1.63	mineral water	<---	[red wine]
rule72	0.023	0.38	1.594	mineral water	<---	[chicken]
rule55	0.036	0.375	1.572	mineral water	<---	[frozen vegetables]
rule17	0.01	0.372	2.136	spaghetti	<---	[mineral water,olive oil]
rule46	0.048	0.37	1.554	mineral water	<---	[milk]
rule49	0.01	0.367	1.54	mineral water	<---	[spaghetti,french fries]
rule145	0.01	0.365	2.096	spaghetti	<---	[red wine]
rule73	0.024	0.357	1.497	mineral water	<---	[tomatoes]
rule75	0.034	0.355	1.489	mineral water	<---	[pancakes]
rule90	0.023	0.348	2	spaghetti	<---	[olive oil]
rule5	0.012	0.348	1.46	mineral water	<---	[avocado]
rule80	0.017	0.345	1.447	mineral water	<---	[herb & pepper]
rule48	0.02	0.344	1.443	mineral water	<---	[whole wheat rice]
rule64	0.06	0.343	1.439	mineral water	<---	[spaghetti]
rule119	0.011	0.34	1.954	spaghetti	<---	[milk,chocolate]

Foto 15. Resultados bajando la confianza.